

University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Doctoral Dissertations

Dissertations and Theses

Fall November 2014

A Defense of Russellian Descriptivism

Brandt H. van der Gaast
University of Massachusetts - Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

 Part of the [Philosophy of Language Commons](#), and the [Philosophy of Mind Commons](#)

Recommended Citation

van der Gaast, Brandt H., "A Defense of Russellian Descriptivism" (2014). *Doctoral Dissertations*. 243.
<https://doi.org/10.7275/kryh-fn53> https://scholarworks.umass.edu/dissertations_2/243

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

A Defense of Russellian Descriptivism

A Dissertation Presented

by

BRANDT H. VAN DER GAAST

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2014

Department of Philosophy

© Copyright by Brandt H. Van der Gaast 2014

All Rights Reserved

A Defense of Russellian Descriptivism

A Dissertation Presented

By

BRANDT H. VAN DER GAAST

Approved as to style and content by

Jonathan Schaffer, chair

Phillip Bricker, member

Hilary Kornblith, member

Angelika Kratzer, member

Joseph Levine, department chair
Department of Philosophy

ACKNOWLEDGMENTS

Thanks to everyone in the philosophy department at UMass Amherst for teaching great classes and being involved in an all-around excellent department. Thanks to my advisor and my committee members for their help and encouragement.

Thanks to Stephan Torre and Uri Leibowitz for lots of philosophical discussion throughout the years.

Thanks to Michael McDermott for inspiration and an early look at some of his articles.

ABSTRACT

A DEFENSE OF RUSSELLIAN DESCRIPTIVISM

SEPTEMBER 2014

BRANDT H. VAN DER GAAST, B.A. FREE UNIVERSITY AMSTERDAM

M.A. UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Jonathan Schaffer

In this dissertation, I defend a Russellian form of descriptivism. The main supporting argument invokes a relation between meaning and thought. I argue that the meanings of sentences are the thoughts people use them to express. This is part of a Gricean outlook on meaning according to which psychological intentionality is prior to, and determinative of, linguistic intentionality.

The right approach to thought, I argue in Chapter 1, is a type of functionalism on which thoughts have narrow contents. On this view, the attitude ascriptions of a regimented psychology capture what people really believe and desire. These attitude ascriptions have content clauses that are what David Lewis calls ‘modified Ramsey sentences.’

I then conclude that, since the meanings of sentences are the narrow contents of the thoughts speakers use them to express, the meanings of sentences can also be

represented with such descriptive sentences. I extend the view so that it applies to individual words. The resulting view is a form of descriptivism.

Referring, I claim in Chapter 2, is the expression of a *de re* attitude. I argue that the non-psychological, *de re* individuation of thoughts captures only contingent features of these thoughts. Furthermore, whether a thought counts as *de re* depends on the attributor's context. These two characteristics carry over to reference. The referential properties of speech acts and expressions are merely contingent features. Furthermore, whether a speech act or expression counts as referring depends upon the attributor's context.

In Chapter 3, I apply this version of descriptivism to indexicals, demonstratives and names. Indexicals turn out to have non-descriptive, context-insensitive, semantically determined meanings. Demonstratives have descriptive, context-sensitive, pragmatically determined meanings. Names, finally, have descriptive, context-insensitive, semantically determined meanings.

In the final chapter, I address Putnam's model-theoretic argument, the most formidable obstacle to the form descriptivism outlined here. I criticize Lewis's 'magnetist' solution that invokes primitive naturalness because it is committed to the existence of incorrigible error about the external world. I suggest an empiricist approach on which psychological intentionality, and so ultimately linguistic intentionality as well, is anchored in experience.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGMENTS.....	iv
ABSTRACT.....	v
INTRODUCTION.....	1
CHAPTER	
1: MEANING	17
1.1 An Argument Relating Belief and Meaning	17
1.2 The Second Premise: Functionalism	26
1.3 Conclusion: Descriptivism	43
2: REFERENCE.....	55
2.1 An Argument Relating <i>De Re</i> Belief and Reference	55
2.2 The Second Premise: <i>De Re</i> Belief	60
2.3 Conclusion: A View On Reference	69
3: INDEXICALS, DEMONSTRATIVES, AND NAMES.....	76
3.1 Taxonomical Issues	76
3.2 Indexicals.....	82
3.3 Demonstratives.....	89
3.4 Names	98
4: THE MODEL-THEORETIC ARGUMENT	112
4.1 The Argument	112
4.2 Structural Naturalness and Future Theorizing.....	116
4.3 Non-Structural Naturalness and Its Problems	125
4.4 An Appeal to Experience.....	138
LITERATURE.....	146

INTRODUCTION

In this dissertation, I contend that descriptivism is the right approach to meaning and reference. Due to the work of Saul Kripke, Hilary Putnam, and others, the theory has gone out of fashion. But I will argue that descriptivism is correct in the fundamentals, that it can be updated and that it remains an attractive view. Even a critic like Kripke cannot help but praise its “marvellous internal coherence” (1972: 5). Another one-time critic, David Kaplan, says that it represents “the Golden Age of Pure Semantics” (1978: 385). The move away from descriptivism has gone too far, I maintain. It is time for the pendulum to swing back.

In this introduction, I sketch the general outlook. I also address some of the starting points from which the arguments depart. Not all of these starting points can be defended in detail. So the reader may judge, on the basis of the picture that arises, whether these assumptions are any good, or perhaps whether something went wrong along the way. In a nutshell, the two main starting points are the following: a functionalist conception of thought, and a view on how meaning relates to thought. Much of this dissertation is an effort to make these starting points look attractive and to spell out their descriptivist implications.

The first starting point is a certain conception of thought. In Section 1.2, I formulate and defend a type of *functionalism*. This type of functionalism resembles David Lewis’s, except for a few tweaks. On the functionalist approach to intentionality, ‘belief’ and ‘desire’ are theoretical terms that are implicitly defined by folk-psychology.

We use this folk-theory in our day-to-day predicting and explaining of people. Folk-psychology posits certain internal states of people and consists of generalizations relating the environment and behavior to these internal states. Why is it needed to consider thought when examining the concept of meaning? Because they are related in very important ways, which brings us to the second starting point.

This second point of departure concerns the relation between meaning and thought. I defend the common-sensical view that the meaning of a sentence is the thought speakers use it to express. On this picture of language, the intentionality of thought is more fundamental than, and determinative of, the intentionality of language. This means that it must be possible to understand mental content independently of linguistic intentionality. In Section 1.2, I spell out this view in more detail and discuss some of the challenges and complications that arise.

The natural view that mental content is prior to, and determinative of, linguistic intentionality has been defended by such philosophers as Paul Grice and David Lewis. On the other side there are people like Wilfrid Sellars and Michael Dummett, who see the intentionality of language as prior to the intentionality of thought. Stephen Schiffer has also defended the Gricean approach. He writes that “the connection between semantics and psychology is such that, without such a theory, there can be no hope of an adequate theory of reference” (1978: 175).

The central argument from Chapter 1 is that these two starting points naturally lead to a certain theory of meaning, viz. descriptivism. To provide a preview: In Section 1.2, I sketch a functionalist account on which thoughts are internal states of people that

are causally related to environment and behavior. They are individuated in terms of their 'cognitive significance' and they are narrow, i.e. they supervene on subjects' intrinsic states. I argue that these narrow contents are best captured by certain descriptive sentences (so-called 'modified Ramsey sentences'). Since the meanings of sentences are the thoughts speakers use these sentences to express, it follows that such meanings are best captured by such descriptive sentences as well.

Along the way, I address some pressing questions. What exactly *is* meaning? What does it mean to try to understand meaning in terms of thought? How does our notion of meaning differ from competing notions of meaning, for instance the one employed in so-called 'truth-conditional semantics'? If the meanings of words can be captured in descriptions, does it follow that all words are definable? That there are lots of analytic sentences (that are not just syntactic tautologies)? Considering these questions will prove useful in locating the theory of descriptivism and also sheds light on the strengths and weaknesses of the theory.

Descriptivism, I maintain, is not only plausible in its own right, but also part of an attractive package of views. This package of views sees the relevant phenomena as stratified into a number of domains, with facts from certain domains made true by facts from more basic domains. Facts about linguistic meaning depend (largely, globally) on facts about mental content; facts about mental content depend (largely, globally) on facts about causal relations to environment and behavior. I maintain that once these relations of dependence are properly appreciated, many of the puzzles about content

and meaning lose their bite.¹ (An important yardstick by which to measure philosophical theories is their puzzle-solving ability.)

Whereas Chapter 1 deals with meaning, Chapter 2 is concerned with reference. Whereas meaning, pre-theoretically, is a relation between language and thought, reference is a relation between language and the world. Reference is closely related to truth, since the truth of sentences depends on what its parts refer to. But the truth of sentences also depends upon whether the thoughts they are used to express (i.e. their meanings) are true. How do these two claims fit together? If the truth of a sentence depends on both on what its parts refer to *and* on whether the thought people use it to express is true, then what is the relation between reference and thought?

I submit that linguistic expressions *refer* to entities just in case they are used by speakers to express thoughts that are *de re* with respect to these entities. This means that a proper investigation of reference requires careful study of *de re* thought. In seeing things this way, our approach agrees with Schiffer, who writes, “the basis of a theory of reference.... is a theory of *de re* propositional attitudes” (1978: 171). Only by getting clear on what exactly *de re* thought is can we find out what it takes to use a linguistic item to refer to something, what it takes to engage in the linguistic act of referring.

In Chapter 2, I argue that the notion of *de re* belief is best regarded as an alternative, non-psychological way of individuating belief tokens. On this way of typing

¹ These puzzles include puzzles with such characters as Kripke’s Pierre, Quine’s Ralph, Putnam’s Oscar, Perry’s shopper, etc.

beliefs, subjects can be said to stand in relations to ‘singular propositions’ or ‘wide contents.’ Such *de re* beliefs about environmental objects or kinds are not narrow states of subjects. Since belief is a psychological notion, and since *de re* typing is non-psychological, the *de re* characterization of a belief does not capture its essential nature. That is to say, that a belief has a certain wide content is a *contingent* property of that belief. Since I explain reference in terms of *de re* belief, this contingency claim carries over to the linguistic realm. We hold that the referential properties of expressions are contingent and extra-semantic properties of those expressions.

In that same chapter, I argue that *de re* belief is context-sensitive. That is to say, whether or not a subject counts as entertaining a belief that is *de re* with respect to some entity depends upon the attributor’s context. In the seminar room, we can even choose to be extremely permissive in counting people as having *de re* beliefs. On this ‘liberalist’ view, subjects can have *de re* beliefs about any object or property they are in a position to have beliefs about, *period*. Since I explain reference in terms of *de re* beliefs, this context sensitivity-claim carries over to the linguistic realm. Like *de re* belief, reference is context-sensitive. In the seminar room, we can be so liberal about reference that even speakers who use definite descriptions *attributively* count as referring.

An advantage of the collection of views just sketched is that it can respect the common-sensical insight that knowledge of what a sentence means differs sharply from knowledge whether that sentence is true.² Knowing the meaning of a sentence does not

² It is sometimes suggested that Wittgenstein’s remark in 2.0211 of the *Tractatus* concerns this issue.

require knowledge of the contingent, non-linguistic world. If the contingent, non-linguistic world were different in certain ways, certain true sentences would become false, and certain false ones would become true. But *English would remain the language that it is*, because the relation between thought and language would remain the same. That is, the meaning-relation that holds between sentences and the thoughts that are their meanings would remain the same. Knowledge of meaning is linguistic, knowledge of reference is not.³

Some philosophers theorize about reference by starting with *modality* instead of *de re* belief. Such theorists may take the notion of *rigid designation* as a point of departure. ‘Rigid designator’ is Kripke’s well-known term for expressions that single out the same thing across possibilities. Kripke famously argued that a proper name like ‘Aristotle’ picks out the same individual across possibilities, whereas a definite description like ‘The last great philosopher of antiquity’ does not. Names are rigid designators, descriptions are not. The former refer, the latter do not. Is it a good idea to try to understand meaning and reference by beginning with the modal notion of rigid designation?

No, it is not. We do not adopt the approach of trying to understand meaning in terms of modality. The main reason is that following this path is commits one to a certain view about the relative explanatory priority of the two notions. If meaning is to be explained in terms of modality, then presumably one cannot analyze modality in

³ This is not to say that linguistic knowledge does not include knowledge that instances of the schema “R’ refers to R’ are guaranteed to be true. See Chapter 3 for more discussion.

terms of meaning. That would be circular. But there is a long-standing and still respectable tradition of trying to understand modality in terms of meaning. Closing off this avenue of explanation ahead of time would be to limit one's theoretical options. That is why in this dissertation there will be very little discussion of rigid designation.

Now, a satisfactory account of modality in terms of meaning is a long way off, of course. Nobody has yet formulated a plausible explanation of one in terms of the other. This does not mean it cannot be brought off, however. One could, for instance, adopt a type of Lewisian counterpart theory to reduce *de re* modality to *de dicto* modality. And then *de dicto* modality could, in turn, perhaps be accounted for in terms of analyticity. (However *that* is to be understood. But see Section 1.3 for more on the subject.) In order to keep this explanatory trajectory available as an option, I will not rely on modality in theorizing about meaning.

A number of philosophers defend what they call 'two-dimensionalist semantics'.⁴ On this approach, sentences are said to have two intensions: a primary and a secondary intension. Roughly speaking, the primary intension of an expression is how its extension varies given different *actual* states of the world, whereas the secondary intension of an expression is how its extension varies given different *counterfactual* states of the world. Adopting the lingo of two-dimensionalism, the notion of meaning that we are concerned with here is that of primary intension. (The reason for not addressing secondary intensions is, again, that doing so would require relying on modal notions.)

⁴ Chalmers 2006.

In Chapter 3, I apply descriptivism to three different kinds of expression: indexicals, demonstratives and names. Because we understand meaning in terms of thought, our account of these expressions differs from the standard treatment they receive in truth-conditional semantics. Indexicals, especially, work differently on the descriptivist approach. To give a little preview: Indexicals have non-descriptive, semantically determined, context-*insensitive* meanings. Demonstratives receive a different treatment. They have descriptive, pragmatically determined, context-sensitive meanings. Names, finally, have descriptive, *semantically* determined, context-insensitive meanings. Along the way, I say a thing or two about the semantics-pragmatic distinction.

What is new in this dissertation? What original contributions does it contain? Versions of the views espoused here can be found in work by Brian Loar, David Lewis, Frank Jackson and others. Nevertheless, I try to cover new ground and to not re-invent the wheel. An important question I address is where to *situate* descriptivism. Exactly what is it a theory *of*? How does it conceive of the relation between the mental and the linguistic? Where does *de re* belief come in? Is descriptivism a plausible view for the meaning of *all* non-logical terms? It seems to me that a proper appreciation of descriptivism, whether it be positive or negative, can only be made once the theory is properly situated.

In addition to doing some of this groundwork, I also formulate a new argument. This argument is a criticism of a view called 'reference magnetism.' Lewis invokes reference magnetism in responding to Putnam's model-theoretic argument. I take the

model-theoretic argument to be the most formidable obstacle for certain forms of descriptivism, so I do think a descriptivist must present a satisfactory reply. I argue that reference magnetism is not a plausible solution, however. The view is committed to the existence of incorrigible error about the external world. Chapter 4 covers Putnam's model-theoretic argument, the magnetist response, and a sketch of a better solution.

Feelings on the fate of descriptivism run high. Many philosophers appear to think that the theory is definitively refuted. Devitt and Sterelny write, "[D]escription theories are wrong not merely in details but in fundamentals. The whole programme is mistaken" (1987: 59). Other philosophers, however, think that the criticisms leave the theory untouched. Jackson goes even so far as to say "There is... no way that an appeal to intuitions about possible cases can refute the description theory" (1998: 213).

In the end, the proof of the pudding is in the eating. Whether descriptivism should be endorsed or not depends upon how well it performs certain explanatory tasks. In order to judge how well it performs these tasks, we must consider: How many questions does the theory answer? How plausible are its answers to these questions? Especially with respect to the first issue, I think descriptivism has the competition beat. When it comes to the second question, I maintain that the descriptivist picture of belief, meaning and reference is very appealing. But in the end, of course, the reader must judge for himself.⁵

⁵ Whenever I talk of arbitrary individuals ('reader' or 'subject'), I will use male pronouns for better readability.

Historical: Russell and Frege on Reference

Bertrand Russell drew a distinction between two kinds of singular terms: proper names and definite descriptions. These two kinds of terms apply to—or stand for, or are about—individuals in different ways. Consider, for instance, the sentences ‘Socrates is mortal’ and ‘The most famous philosopher is mortal’. Russell writes, “‘Socrates is mortal’ ... express[es] a fact of which Socrates himself is a constituent: there is a certain object, namely Socrates, which does have the property of mortality, and this object is a constituent of the complex fact which we assert when we say “Socrates is mortal.”” (Russell and Whitehead 1927: 54). Because Socrates is a constituent of what is said, the sentence expresses a *singular proposition*.

Not so with definite descriptions. In the sentence ‘The most famous philosopher is mortal’, the description ‘The most famous philosopher’ does not contribute an individual to the proposition. About such sentences, Russell writes, “the grammatical subject is not a proper name, i.e. not a name directly representing some object. Thus in all such cases, the proposition must be capable of being so analyzed that what was the grammatical subject should have disappeared” (ibid.). According to his theory of descriptions, the sentence ‘The F is G’ expresses a *general proposition* about the co-instantiation of certain properties.

The main argument Russell advanced for his view concerns meaningfulness. According to Russell, a sentence like ‘The king of France is bald’ is clearly meaningful. As a result, ‘the king of France’ cannot contribute the king of France to the proposition

expressed, since there is no such person.⁶ Speakers can understand sentences containing descriptions that are empty—either contingently (as in ‘the king of France’) or necessarily (as in ‘the round square’). Since a speaker can understand such a sentence (i.e. grasp its meaning) without being acquainted with the entity that the singular term applies to, this entity is not a constituent of the proposition expressed (i.e. the sentence’s meaning).

On this picture, proper names differ fundamentally from definite descriptions. Proper names are ‘bare tags’ that contribute individuals to the proposition expressed; they do not “assign a property to an object, but merely and solely name it” (1910: 224). The name ‘Scott’ means Scott; it provides no information about the individual in question. This name does not describe the individual as having certain properties; it is the *sentence* that does that. That ‘Scott’ refers to Scott is a semantically relevant fact: ‘Scott’ contributes Scott to the meaning of the sentence in which it occurs. For such expressions, there is no gap between meaning and reference.

On the Russellian picture, logically proper names are such that their referents are guaranteed to exist. Furthermore, given Russell’s stringent acquaintance requirements, it is not possible for a rational subject to entertain the thought expressed by ‘a is F’ but not the thought expressed by ‘b is F’ when ‘a’ and ‘b’ are co-referential logically proper names. In such a case, the thought expressed by ‘a is F’ is the thought

⁶ Russell 1905: 46. Russell wrote that this sentence “has a *meaning* provided ‘the king of England’ has a meaning...” and that the sentence, “is not nonsense, since it is plainly false” (1905: 46). Strawson 1950 contests this latter claim. However, one can agree with Russell that it is meaningful without deriving it from the intuition that it is false.

expressed by 'b is F.' And it is of course impossible to both entertain and simultaneously not entertain one and the same thought.

Similar considerations apply to predicates that express universals. Russell writes that, "we have also... what may be called awareness of *universals*. [...] And the universal yellow is the predicate in such judgments as "this is yellow," where "this" is a particular sense-datum" (1910: 212). Predicates that express such universals work in a way that is similar to logically proper names: they are responsible for certain propositional contributions, and these contributions are guaranteed to exist and are wholly transparent to the subjects who entertain thoughts involving them.

Russell held the view that, "*Every proposition which we can understand must be composed wholly of constituents with which we are acquainted*" (Russell 1910: 219, italics original). Given this strict epistemic requirement, Russell came to the view that ordinary names, such as names for people and places, are not logically proper names after all. Only names for particulars with which we can be directly acquainted are logically proper names: the indexical 'I', and 'this' and 'that' as names for one's current sense data. About ordinary proper names, Russell writes that "[they] are usually really descriptions. That is to say, the thought in the mind of a person using a proper name correctly can generally only be expressed explicitly if we replace the proper name by a description" (1910: 216).

While Gottlob Frege is also considered a descriptivist by many, his view on meaning and reference is fundamentally different from Russell's. On Frege's view, both proper names and definite descriptions are singular terms. Singular terms have a

meaning (Sinn) and a *reference* (Bedeutung), and a singular term's meaning never coincides with its reference. On Frege's two-tiered system, the meaning of a singular term is the 'mode of presentation' of the referred-to object, whereas the reference is what the user of the singular term is talking about.⁷ For most singular terms in a language, there is unique sense; and for most of these senses there is a unique referent.

Frege writes,

If we say, 'The Evening Star is a planet with a shorter period of revolution than the Earth,' the thought we express is other than in the sentence, 'The Morning Star is a planet with a shorter period of revolution than the Earth'; for someone who does not know that the Morning Star is the Evening Star might regard one as true and the other as false. And yet the *Bedeutung* of both sentences must be the same; for it is just a matter of interchange of the words 'Evening Star' and 'Morning Star,' which have the same *Bedeutung*, i.e. are proper names of the same heavenly body.⁸

Frege held that concrete objects cannot be constituents of thoughts. In correspondence, he disagreed with Russell on whether the Mont Blanc can be the constituent of a thought. Russell insisted that, 'in spite of all its snowfields,' the Mont Blanc can be part of a thought. Frege disagreed. Now, Russell later came to believe that such things as mountains cannot be constituents of thoughts, but he nevertheless allowed particulars and universals to be components of thoughts.

Frege's view that all reference is mediated allows him to regard a large class of

⁷ Frege 1948: 214.

⁸ Frege 1891: 138.

expressions as referring expressions. For Russell, referring terms are the ones that contribute what they stand for to the proposition expressed. As we just mentioned, they only include 'I' and 'this' and 'that' as names for one's current sense-data. Frege, on the other hand, counts proper names for people and places as referring terms.⁹ For Frege, complex singular terms, such as 'the teacher of Alexander and pupil of Plato,' refer as well. Even complex function expressions, such as, 'is a positive whole number less than ten,' refer. This latter expression does not refer to an object, but it does have a *Bedeutung*.

Can these remarks about Russell and Frege be useful in locating a concept of reference? One option would be to take the Russellian approach. On this approach, a referring expression is an expression that contributes what it stands for to the proposition expressed. Since propositions are the kinds of things grasped by speakers, referring expressions stand for things that are guaranteed to exist and that speakers can be acquainted with. The problem with taking this approach is that, proper names are

⁹ Often-raised questions about Frege's view are the following: Are the senses of proper names expressible in language? Do the senses of proper names differ among speakers of a language? For instance, in a footnote, Frege wrote, "In the case of an actual proper name such as 'Aristotle' opinions as to the sense may differ... So long as the referent remains the same, such variations of sense may be tolerated" (1948: 210). But elsewhere, Frege emphasizes that senses must be *shared* (unlike 'subjective conceptions'). "The sense of a name is grasped by everyone who is sufficiently familiar with the language to which it belongs" (1948: 210). Frege also considers two speakers who associate different senses with the name 'Gustav Lauben.' He writes that these speakers "do not speak the same language, since, although they do in fact refer to the same man with this name, they do not know that they do so" (1956: 297). How can these remarks be squared? One suggestion is to distinguish names that competent speakers of a language must know from names that such speakers do not need to know. Proper names of people and places appear to be the latter kind of name. (See Chapter 3 for more discussion.) Another suggestion is to allow a certain amount of variance in meaning across conversational contexts, but not *within* conversational contexts. As long as within contexts, meanings to not vary between speakers too much, this allows or the transmission of mankind's "common store of thoughts" (1948: 212).

unlikely to count as referring expressions. In fact, it will turn out that no public language expression meets these requirements, not even the indexical 'I.' (See Section 3.2.)

Or we could adopt a Fregean approach on which ordinary names for people and places count as referring terms. If we adopt this view, reference is not as direct or unmediated as it is on the Russellian view. Speakers will not be acquainted with the referents of many terms. However, too many expressions will count as referring. On the Fregean approach, there is the risk of losing the interesting distinction between different kinds of singular terms. Complex singular terms, such as definite descriptions, will count as referring expressions as well.

Most philosophers since Russell and Frege claim that Russell's approach is too restrictive, while Frege's approach is too permissive. Among the referring expressions, they want to allow more than just the ones Russell includes, but not as many as Frege includes. Kripke, for instance, counts proper names of people and places as referring expressions, but not definite descriptions. He distinguishes the two by relying on the concept of rigid designation. Since I want to steer clear of *de re* modality, I will press forward by looking further at the notions that Russell and Frege employ in thinking about meaning: thought and belief.

It will turn out that Russell's restrictive notion of reference and Frege's permissive notion of reference are actually preferable to one that draws the line somewhere in the middle. In the chapters that follow, I will suggest a picture on which we can either follow Russell and allow reference to very few kinds of things, or follow Frege and allow reference to many different kinds of things. I argue that, to an extent,

this is a matter of context. In some, it is appropriate to talk like Russell; in others, to talk like Frege.

However, if we ignore the effects of context, and focus our attention on the real nature of certain representational acts, ultimately it is Russell who will be vindicated (or so I maintain). On the view favored here, the meanings of sentences can be captured in descriptive statements that ultimately *do* turn out to contain simple names. So, the view even incorporates a notion that resembles Russell's notion of a logically proper name. It is the work of the next four chapters to attempt to convince you, the audience, that such a Russellian view deserves to be taken seriously.

CHAPTER 1

MEANING

1.1 An Argument Relating Belief and Meaning

In this chapter, I will propose a certain view on belief and meaning and provide arguments in its support. As indicated in the introduction, I argue that meaning should be understood in terms of belief. Belief, in turn, should be understood on the model of functionalism. Together, these two theses provide support for a descriptivist approach to meaning. The goal of this chapter is to sketch the argumentative strategy that ultimately provides the strongest evidence for descriptivism. I will first lay out the central argument, and then spend one section on each of the argument's parts. The current section deals with Premise 1, Section 1.2 with Premise 2, Section 1.3 with the conclusion.

The central argument is the following:

An Argument Relating Belief and Meaning

Premise 1: The meaning of a sentence is the belief speakers use the sentence to express.

Premise 2: The correct theory of belief is a version of functionalism, on which beliefs have narrow contents. Narrow contents, furthermore, can be represented with descriptive sentences that contain no names of environmental

entities and kinds.

Conclusion: Therefore, the meaning of a sentence is the narrow content of the belief speakers use it to express. Such meanings can be represented with descriptive sentences that contain no names of environmental entities and kinds.

In the current section, I will discuss the argument and provide support for the first premise. Meaning, it is argued, is a relation between language and thought. Section 1.2 discusses Premise 1 and is concerned with the nature of belief.¹⁰ I formulate a type of functionalism, largely similar to Lewis's but diverging in a few spots. Finally, in Section 1.3, I spell out the conclusion of the argument. If the meanings of sentences are beliefs, and if functionalism is the right approach to belief, then what repercussions does this have for the notion of meaning?

In the next chapter, I will consider a second argument that concerns *reference*. Whereas meaning is explained in terms of belief, I propose to understand reference in terms of *de re* belief. *De re* beliefs are relations between subjects and singular propositions involving ordinary particulars and properties. If reference is to be understood in terms of *de re* belief, the nature of *de re* belief can teach us things about the nature of reference. The characterization of beliefs as being *de re* with regards to

¹⁰ What about other propositional attitudes besides belief? Would they not also be relevant in the explanation of meaning? In explaining the meanings of declarative sentences, the propositional attitude that is most relevant is that of belief or thought.

environmental entities and kinds, I argue, captures only contingent features of these beliefs. Furthermore, it is contextually sensitive. It follows, I argue, that the referential properties of expressions and speech acts are contingent and context-sensitive as well. These issues will return in Chapter 2.

Let me first start with the notoriously slippery term ‘meaning.’ In an article from 1980, Lewis provides a list of 30 (!) terms that have been used synonymously with ‘meaning’ (1980b). Intension, connotation, truth-condition, semantic value... Philosophers have used these technical terms in many different and incompatible ways. In this dissertation, I also provide an account of meaning. But what is the use of crowding the field with yet another notion of meaning? Does this not run the risk of obscuring the phenomena even more, instead of clearing things up?

I take it that a technical term is worth keeping around insofar as it belongs to a theory that does useful explanatory work. So if these 30 theories all do valuable explanatory work (that does not overlap too much), we have reason to not throw them out. However, these theories do not merely employ technical notions, they use notions that its proponents say captures *meaning*. At this point, a conflict arises. Which technical notion comes closest to playing the role of our non-technical, common-sense concept of meaning? One would expect some of these notions to do a better job at capturing this everyday concept than others.

Consider, for instance, philosophers who work in the tradition of ‘truth-conditional semantics.’ Within this discipline, it is customary to assign to expressions so-called ‘semantic values.’ Practitioners of this type of linguistics sometimes call these

semantic values ‘meanings.’ On the view favored in this dissertation, these semantic values are not meanings. Rather, they are referents. Now, the philosophers who call semantic values ‘meanings’ may, elsewhere in their view, employ a notion resembling the one that *we* call ‘meaning.’ But they will use a different label for this notion and probably relegate it to ‘meta-semantics.’

This is not purely a terminological debate. The question is: Which of these notions comes closest to capturing the common-sense concept of meaning? I claim that our notion of meaning *outperforms* many others in this respect. It outperforms, for instance, the notion of semantic value. In our sense of ‘meaning,’ competent speakers of the language know the meaning of expressions in that language. In our sense, the meanings of sentences determine the conditions under which they are true. And in our sense, the meanings of sentences are the beliefs speakers use the sentence to express. Not all proponents of the different notions of meaning can easily agree with these statements.

Let me make a second preliminary remark on the notion of meaning. The notion of meaning is central to many philosophical concerns. Where one pitches one’s tent when it comes to meaning is relevant to one’s stance in many debates in philosophy. Meaning, initially, appears to be a notion from the philosophy of language only, but it turns out that it is relevant to issues in epistemology, metaphysics, and philosophical methodology. In Chapter 4, for instance, it will turn out that questions of meaning ultimately lead to metaphysical and epistemological debates about structuralism and empiricism. Arguably, there is no philosophical notion more central than that of

meaning.

Consider now Premise 1, which reads: *The meaning of a sentence is the belief speakers use the sentence to express.* Now, certain expressions in this statement can be interpreted in such a way that it comes out vacuously true. If 'the belief expressed by' is stipulated to mean the same as 'the meaning of,' then the statement comes out true, but is not interesting. So, the premise must be read in such a way that it is substantive. At the same time, the first premise ought to be acceptable to philosophers with different takes on the surrounding issues. It would be a problem if Premise 1 preached to the choir and was only acceptable to people already signed up to the descriptivist program.

If we reflect on the purposes for which speakers use language, Premise 1 seems very plausible. Speakers employ language for the communication of thoughts to audiences, among other things. Such communication requires the existence of conventions that associate linguistic items with representational contents. Considering this important goal of human communication, these conventions can be regarded as relations between linguistic expressions and mental contents. What else could they be? Once such conventions are in place, speakers can transmit to others information about their surroundings. And these speakers are carriers of such information because they are *believers*.

When a speaker is a competent user of a language, he knows which sentences in that language express which beliefs. When such a language user has certain beliefs, he will be inclined to *assent* to the sentences in the language that express those beliefs.

And when such a speaker hears others assent to sentences in the language that express certain beliefs, he will attribute those beliefs to those speakers. If he considers the speaker a reliable source, he may even come to endorse these beliefs himself.¹¹

Knowledge of these sentence-thought relations amounts, it appears, to knowledge of meaning.

These remarks on the concept of meaning strike me as *locating* the subject of investigation; as such, they are non-negotiable and not up for debate. They capture features that are essential to meaning; without this relation to belief, the concept in question would not *be* meaning. As suggested earlier, this is not to say that there are no other notions in the neighborhood of this one that can do useful explanatory work. However, it is our view that these other notions are not as deserving of the label ‘meaning,’ because they do not come as close to playing the role of the common-sense notion of meaning.

At the same time, these remarks on the concept of meaning do not significantly restrict the range of viable theories. Philosophers who have diametrically opposing views on thought can nevertheless agree on this fundamental connection between meaning and thought. These philosophers will disagree on what meaning is, because they disagree on what thought is. But they agree on the meaning-thought relation. For instance, one could adopt a form of externalism about both thought and meaning, and agree with the preceding claim that the meanings of sentences are the thoughts that

¹¹ This claim will be qualified somewhat below when it comes to indexical belief. See Section 3.2 for more discussion.

speakers use them to express.

Let me now discuss a potential criticism. This criticism proceeds as follows: Understanding meaning in terms of thought is problematic, because meanings are public and stable over time, while the thoughts that speakers associate with sentences are not. It would indeed be a problem for the theory if linguistic meanings differed wildly across people and across time. In order to assuage this worry, let me mention four considerations. Some of these will return in later chapters. The picture of meaning that emerges from these chapters is one on which meanings are relatively stable over time and also shared between competent speakers of the language. But let me try to nip some of these worries in the bud right now.

First off, belief is not as individualistic as some suspect. Our favored account of belief, to be explained in some detail below, is a type of functionalism on which beliefs have narrow contents. Narrow contents supervene on people's internal states; and these internal states 'stop at the skin.' At the same time, however, the narrow contents of these internal states depend upon the causal roles of these states in a *population of believers*. The narrow content of Jimmie's internal state depends on the causal role of that internal state not just in Jimmie, but also in his fellow thinkers. So there is a sense in which narrow content is not narrow. More on this issue in the next section.

Second, some expressions are indeed such that speakers associate them with different beliefs. For instance, names of people and places ('Aristotle,' 'Paris'). However, one need not be familiar with these expressions in order to count as a competent user of English. In that sense, they are not really part of English. Or any other language, for

that matter. A theory of meaning should in the first place apply to those expressions that one must know in order to count as a competent speaker of the language. Once we have an account of their meaning, we can move on to other, less-central expressions. Sections 3.1 and 3.4 return to the issue of proper names.

Thirdly, in some cases the variance in the beliefs that speakers associate with expressions results from misunderstanding. Some people simply are not fully competent speakers of the language. To take Tyler Burge's familiar example, the person who believes that 'arthritis' is the name of a disease that one can have in one's thigh does not grasp the meaning of that term. Indeed, it is a virtue of the current account that it can draw a line between those understandings of a term that count as grasping its meaning and those that do not. In the chapters that follow, I will argue that many words are defined by 'theories' that are believed by speakers. Speakers with an insufficient grasp do not understand the meaning of the term in question.

Fourth, the meaning of certain terms does indeed change over time. The term 'mass' is sometimes provided as an example. In Newton's time, the term meant something different from what it means now. In Chapter 3, I will return to the question of meaning change. I will consider the question of how terms enter a language and how subsequent use of that term can affect its meaning. Changes in the meaning of a term can result in a change in reference, but this need not be the case. In the chapters up ahead, some of these issues relating to meaning variance and meaning change will return.

To now return to the main thread, the proposed claim here is that, if meanings

of sentences are anything at all, they are first and foremost thoughts. The claim is *not* that, for every sentence uttered by some speaker on some occasion, the sentence's meaning on that occasion is exactly identical to the thought the speaker intends to convey on that occasion. That would be a naïve claim. Since Grice's work on conversation, it is a commonplace that the thoughts that speakers convey often go beyond what their words mean. In order to be efficient communicators, people use all kinds of shortcuts. In order to maximize the effort-to-result ratio when they speak, they seldom make every single thing explicit.

But that does not conflict with our understanding of meanings as thoughts. Indeed, it is Grice's own 1957 landmark article 'Meaning' where he proposes to understand linguistic intentionality as derivative from psychological intentionality. And in another classic article, Lewis's 1975 'Languages and Language,' Lewis approaches things from much the same perspective. Understanding what meaning is begins with noticing that, "There are regularities whereby the production of sounds or marks depends upon various aspects of the state of mind of the producer. There are regularities whereby various aspects of responses to sounds or marks depend upon the sounds or marks to which one is responding" (Lewis 1975).

Meaning, then, appears to be a relation between linguistic items (e.g. sentences and expressions) and mental items (e.g. thoughts and concepts). The extension of 'meaning,' on this picture, is a relation, i.e., a class of ordered pairs. An important question is: Why does 'meaning' have the extension that it has? Why does meaning relate the things that it relates? Unsurprisingly, it turns out to require a complex story to

explain in detail exactly why ‘meaning’ has the extension that it has. I take Grice’s ‘Meaning’ and Lewis’s ‘Languages and Language’ to address this question. They are concerned with the question: What relation has to obtain between a thought and a sentence in order for the former to be the meaning of the latter?

Whatever the answer to this question, it does not tell us exactly what meanings *are*. It tells us why the facts about the meaning-relation are the way they are—but it does not tell us the nature of one of the *relata* of this meaning-relation. This dissertation follows the outlook of Grice and Lewis, but addresses the question about the nature of meanings. Its goal is to understand what exactly meanings are. As is apparent by now, the strategy is to understand meaning by understanding thought. In the next section, we will proceed by starting with the most promising theory of thought (*viz.* functionalism) and see where this takes us.

1.2 The Second Premise: Functionalism

Functionalism is an account of propositional attitudes. According to the view, ‘belief’ and ‘desire’ are theoretical terms from a folk-psychological theory. This theory posits certain internal states of subjects and consists of various *ceteris paribus* generalizations. We use these generalizations about belief and desire and other internal states to predict the effects of the environment upon people, and to explain the causes of the behavior of people. We also use these generalizations to explain and predict the internal states of people on the basis of other of their internal states.

These folk-psychological principles implicitly define the terms ‘belief’ and ‘desire.’ They capture what is *constitutive* of belief and desire. Furthermore, they are couched in causal terms. For instance, the fact that a subject’s internal state is (or would be) caused by looking at something red counts in favor of interpreting that state as being a belief that there is something red in front of one. And the fact that an internal state causes (or would cause) the subject to walk counts in favor of interpreting that state as a desire to walk.

These constitutive principles also govern the interrelations between these content-bearing internal states. If a belief that P and a belief that if P then Q cause a third state, then this counts in favor of interpreting that state as a belief that Q. And if a desire that Q and a belief that if P then Q cause a third state, then this counts in favor of interpreting that state as a desire that P. On this approach to interpretation, believers are guaranteed to draw mostly rational inferences, have mostly true beliefs, and act mostly rationally. On Lewis’s brand of functionalism, for instance, “constraints of rationality are *constitutive* of content” (1994: 321; also 1986: 36).

Two mental states that stand in different enough causal relations to other mental states, to behavior, or to environment, thereby differ in content. The thought expressed by ‘The Morning Star is bright,’ for instance, differs from the one expressed by ‘The Evening Star is bright’ (even though the Morning Star = the Evening Star = Venus). The two thoughts play different roles in the psychological economy of the believer, give rise to different behaviors, and are caused by different environmental circumstances (for instance, by looking at Venus in the morning or by looking at Venus

at night).

Digging a bit deeper, consider this Frege-case: A subject perceives a red ball and forms a belief that he would express with 'This is red.' Via a special mirror, the subject is presented with a second image of the ball that makes it seem green. Unaware that his two perceptions are of the same object, the subject forms a second belief that he would express with 'That is green.' Intuitively, the rational subject is justified in holding these two beliefs. For the functionalist, then, the demonstrative element in the two beliefs (expressed by 'this' and 'that') makes a different contribution to the belief's content. If it did not, the epistemically blameless subject would hold contradictory beliefs. This principle about the individuation of beliefs is sometimes called 'Frege's constraint'.¹²

At this point, two objections may be heard. The first one is that 'green' does not entail 'not red' and that therefore the beliefs are not contradictory. Someone could argue that 'not red' is not part of the definition of the term 'green' and that therefore 'green' does not entail 'not red.' This objection can be dealt with by pointing out that entailments do not have to be grounded in definitions. 'Green' entails 'colored,' but 'green' is not definable in terms of 'colored' and a remainder. Indeed, it is doubtful that

¹² This case involves two simultaneous beliefs to sidestep questions about whether beliefs about *past* circumstances are really justified. And the case involves positive beliefs that things have properties, rather than *negative* beliefs that things lack properties, to sidestep questions about whether the latter are really justified.

The subject from our example may also believe something he would express with 'This is not that.' Perhaps he infers this belief from his previous two beliefs. Frege's constraint applies here too. If we took the perceptual demonstratives as contributing their referent to the proposition that captures the content of the belief, the rational subject would be endorsing a contradiction. We can imagine the subject to be a logician who is well aware of the law of self-identity and who is inclined to reject any proposition that violates this law.

color terms—and many other terms—can be explicitly defined at all.¹³ (This issue of definability will return several times in the chapters that follow.)¹⁴

The second objection is the following. Is the belief that the ball is green really a justified belief? After all, the belief is false! What is more, the process by which the belief is acquired is far from reliable. The subject is attempting to ascertain the color of an object by looking through a color-distorting screen, which is not a process likely to result in true beliefs. This objection can be dealt with by insisting on an *internalist* notion of justification. This kind of justification is closely related to norms of rationality and epistemic blamelessness. It is reasonable to suppose that in this sense of ‘justified,’ the subject who believes the ball to be green has a justified belief (even though it happens to be false).

That rationality is a constraint on assigning content to mental states is not to say that on this theory people do not have irrational, false or even contradictory beliefs;

¹³ See Chapter 4 for more discussion of the meaning of color terms such as ‘green’ and ‘grue.’

¹⁴ Quine (1956: 182) considers a view on which our subject would be able to think, without contradiction, both the *singular* proposition that the ball is red and the singular proposition that the ball is green. Only entertaining the singular proposition that it is *both* red and green would be contradictory and therefore irrational. But our subject may be a logician who is well aware of certain principles of *conjunction*, for instance the principle that if object x has property F and object y has property G and x is identical to y, then object x has both properties F and G. Our rational subject will be inclined to believe any instances of this principle, so Quine’s attempt to save this subject from being irrational does not work.

Kaplan (1968: 234) also criticizes a certain aspect of this view of Quine’s. Changing Quine’s spy-example to our red ball-example, Kaplan makes the following point: If believing of the ball that it is red is not in contradiction with believing of the ball that it is non-red, neither should *we attributors* be contradicting ourselves when we say: S believes the ball to be red and it is false that S believes the ball to be red. Kaplan says that this latter statement (‘It is false that S believes the ball to be red’) follows from: S believes the ball to be non-red. He writes, “it is natural to claim that [the former] is a consequence of [the latter]” (1968: 234). However, committed direct-reference theorists nowadays resist this implication. So I think that my complaint about Quine’s attempt to preserve rationality is more to the point than Kaplan’s.

they *do* have such beliefs. Rather, it is to say that when content is assigned to internal states that play causal roles in believers, constraints of rationality narrow down the number of eligible interpretations. The most eligible interpretation may still attribute certain irrational, false and contradictory beliefs and desires to subjects. Constraints of rationality dictate that the amount of such propositional attitudes ought to be minimized.

On functionalism, 'belief' and 'desire' are theoretical terms that earn their keep from their use in a psychological theory. Does this mean that the belief attributions that speakers of English produce in their day-to-day explaining and predicting of other people are the best guides to psychological content? Does this mean that the best approach to understanding propositional attitudes is to study the semantics of belief attributions in English? No! Furthermore, it is important to see *why* this would not be the correct way to try to understand mental content. Let me explain.

First off, we use *English* when we engage in the folk-psychological explanation of people's actions on the basis of their internal states, and the prediction of people's internal states on the basis of environmental impact. But English is not a language specifically designed for this purpose, nor is any other existing language. The belief and desire attributions that speakers of English produce, therefore, are not necessarily reliable guides to the real nature of these beliefs and desires. They only imperfectly capture what people really believe and desire.¹⁵

¹⁵ See Loar's important 1988 article on the difference between psychological content and social content.

We can make progress on the issue of belief without becoming entangled in the tricky issue of the semantics for belief ascriptions in English. Remember our Gricean allegiance: Our goal is to explain linguistic intentionality in term of psychological intentionality. This means that the order of investigation should be: *first, belief; then, language*. To start with language would be to reverse the order of investigation. To study the content of thoughts by considering the meaning of words would be to use a derivative, complex phenomenon in order to explain a more basic, and simpler (but still very complicated) phenomenon.

Belief attributions are, after all, linguistic expressions that are about beliefs. They are the expression of *beliefs about beliefs*. So the order in which to proceed is (from simpler to more complex): first belief, then language, then belief about belief, and *only then* language about belief (belief attribution). Starting at the most complex end, as opposed to at the simpler end, would be methodologically unsound. Lewis agrees that starting with the analysis of belief ascriptions in English is far from ideal. “[I]t seems to me unfortunate that the study of the objects of belief has become entangled with the semantic analysis of the attributions of belief” (1979: 154).

Let me now raise and answer briefly three important questions that arise for the view suggested above. In the paragraphs below, I will motivate these answers in more detail. The first question is: Are mental contents narrow? That is, do they supervene on the thinker’s intrinsic, internal state? Answer: Yes, they are narrow. Second question: Can mental contents always be captured in the thinker’s language? Answer: No, it is too strong an assumption to suppose that they always can. Third question: Can we know the

content of a thinker's thoughts before knowing the meanings of his words? Answer: Yes, this is part of the fundamental assumption that mental content is prior to linguistic meaning.

Let us begin with the first question: Is mental content narrow? A moment ago, I claimed that belief ascriptions in English only imperfectly capture mental content, because English is not a language specially developed for the purpose of psychological explanation. What people really believe and desire depends upon what a proper *regimentation* of folk-psychology says they believe and desire. Now, this more rigorous psychology will consist of *ceteris paribus* generalizations relating internal states to environmental impact and to behavioral output. The question is: How will these regimented principles describe environmental inputs and behavioral outputs?

They will describe the environmental causes *narrowly*, as events at the subject's sensory surface. And they will describe behavior narrowly as well, as outputs going in the direction of the subject's environment. Why? Because of the *constitutive rationality* referred to earlier. The rationality of subjects' internal attitudes can be *maximized* if these principles are couched in terms of inputs and outputs, since these are the things *by means of which* we recognize and act on our environment. The means by which we recognize and act on our environment are in general such that we do not take contradictory or irrational attitudes towards them.

Suppose, for the sake of *reductio*, that this psychology's generalizations were not narrow, that they made reference to wide environmental circumstances. Interpreters would then attribute attitudes to subjects on the basis of generalizations as, 'If a subject

sees a red ball, he will believe there to be a red ball.’ Such an interpreter would not be able to understand the Frege-case from a few paragraphs back. According to the interpreter, the subject would be seeing the same thing twice (viz. the red ball), and believe there to be a red ball. He would not be able to make sense of the subject’s ‘mistaken distinctness’-beliefs. The subject is behaving as if there are *two* balls, but this would be *irrational* on such a wide psychology.

The same applies to ‘reversed Frege-cases,’ where a subject mistakenly identifies two distinct things, because they present him with the same appearance. For instance, suppose I encounter two qualitatively identical twins at different times, and believe them to be one and the same person. An interpreter using a wide psychology with generalizations like, ‘If a person sees X, he will believe there to be X,’ would attribute to me two different beliefs about two different persons. Such an interpreter would not be able to make sense of my ‘mistaken identity’ beliefs. I am behaving as if there is *one* person, but this would be *irrational* on a wide psychology.

So in order to interpret the behavior that results from such errors as rational, a wide psychology is useless. In order to make sense of this behavior, we need a narrow psychology. We need a psychology the generalizations of which do not make reference to environmental circumstances. What is more, given that our psychology should be one that covers successful as well as unsuccessful cognition, we need a narrow psychology *across the board*. (If error did not exist, we could use a wide psychology to attribute attitudes. But such a psychology would be useless in the actual world where, as we know all too well, these mistakes are common.)

Now for a qualification of this claim. As we suggested in the previous section, there is also a sense in which this content is not narrow. The content of an internal state depends on the causal role of that state in a *population* of believers. The hypothetical radical interpreter whose job it is to assign content to people's internal states will not only look at a state's causal role in Jones' psychological economy, but also at its causal role in Jones' fellow believers. What is more, he will even take into account its causal role in *future* members of Jones' population. (This latter issue will return in Chapter 4.)

An internal state (e.g. pain) may have a non-standard effect in one person (e.g. cause a desire that it continue), but if this internal state plays the pain-role in enough other members of the population (and causes a desire that it *stop*, among other things), it nevertheless is pain.¹⁶ The same applies to belief. A belief (e.g. the belief that there is something red) may have a non-standard cause in one person (e.g. be caused by seeing grass), but if this internal state plays the believing-that-there-is-something-red-role in enough other members of the population (and is caused by seeing fire trucks, among other things), it nevertheless is the belief that there is something red.

So on this form of functionalism, then, beliefs have narrow contents. The generalizations and attitude ascriptions from the regimented psychology will contain no terms referring to objects or properties in the thinker's environment. What will the content ascriptions from such a regimented psychology look like then? They will look very different from the belief ascriptions that we speakers of English make in our day-to-

¹⁶ See Lewis 1980a.

day application of folk-psychology, in our everyday predicting and explaining of other people. But different in which way?

On the regimented, narrow psychology envisioned a few paragraphs ago, subjects' beliefs can be captured by belief attributions where the content clauses are '*modified Ramsey sentences*' (see Lewis 1972). This can be explained as follows. The Ramsey sentence of a theory is a sentence where the theoretical terms are replaced with existentially quantified variables. So if the theory says, $T(t_1, \dots, t_n)$, the Ramsey sentence of the theory says, $\exists x_1 \dots \exists x_n T(x_1, \dots, x_n)$. The *modified* Ramsey sentence adds a uniqueness condition. It says that there exists a *unique* n-tuple $x_1 \dots x_n T(x_1, \dots, x_n)$. A modified Ramsey sentence is not merely an existential claim, it is the stronger claim to the effect that there exists *exactly one* such-and-such.

On this picture, the regimented, narrow psychology will ultimately produce a belief attribution of the form: S believes that there exists a unique n-tuple $x_1 \dots x_n T(x_1, \dots, x_n)$. Every existential quantifier in this ascription corresponds to an entity or kind that the subject believes to exist. And the belief ascription captures how the subject believes these entities and kinds to be *structurally related* to each other. (The term 'theory,' here, should be taken lightly; it merely indicates that it concerns a collection or totality of beliefs.)

One non-logical term that occurs in such a modified Ramsey sentence is '*instantiates.*' Remember, the quantifiers in such a sentence range over particulars as well as properties or kinds. The Ramsey sentence will say that certain entities that the subject in question believes to exist instantiates certain properties that he believes to

exist. An important question arises: Will this modified Ramsey sentence contain, besides 'instantiates,' other non-logical terms as well? Or will it only contain purely logical vocabulary?

Lewis, in his earlier work, seems to have held the view that intentional content can be represented using only logical vocabulary. He seems to have endorsed a form of functionalism that conceives of mental representation purely in terms of structure.¹⁷ On this view, every single concept is purely defined in terms of its conceptual or functional role. An interpretation is a mapping from the variables in the modified Ramsified belief ascription onto extensions in the world. This mapping is one that maximizes the truth of the ascribed theory. But there are no prior constraints on how the subject's representational structure is to be interpreted.

Besides being an extreme view, the view is also susceptible to Newman's problem (also known as Putnam's model-theoretic argument.)¹⁸ According to this argument, truth and consistency cannot be the only constraints on interpretation. Because if these two were the only constraints, it would be far too easy for the totality of a subject's beliefs to come out true. Error would no longer be possible, because there would always be a mapping from the terms onto extensions in the world that makes the believed theory come out true. In response to this challenge, Lewis changed his view. In Chapter 4, I will return to these issues.

¹⁷ See Chapter 4 for textual evidence that Lewis held this view.

¹⁸ Newman 1928, Putnam 1978.

Putting these questions aside for now, we have arrived at Premise 2 of the argument from Section 1.1. As explained above, a modified Ramsey sentence is an existential statement with a uniqueness condition added. It says that there exists a unique n-tuple such-and-such that is interrelated in so-and-so a fashion. In other words, it can be considered a descriptive statement. What we have here is a version of descriptivism that applies in the realm of thought. On this view, the belief attributions generated by a regimented, narrow psychology contain descriptive sentences that do not contain any names for environmental entities and kinds.

Let me now return to some of the other questions raised earlier. The second question was: Can narrow content always be expressed in the thinker's language? I suggest the more careful answer 'No' instead of 'Yes.' So far, I have argued that the real contents of people's thoughts are captured by the attitude ascriptions of a regimented, narrow psychology. This means that the language of this hypothetical psychologist can capture the content of people's thoughts. But this language need not be the same as the subject's language. There is no guarantee that the subject's language can capture every nuance of the subject's thoughts.

This issue turns out to be relevant to questions of *analyticity* and *definability*. For suppose that the meanings of sentences are the narrow contents of the thoughts they express. Extending this approach to words: suppose the meanings of words are the narrow contents of the concepts they express. If such narrow contents were expressible in the subject's language, one would expect there to be many analytic statements (that are not purely syntactic tautologies). But we all know that it is very hard to come up

with such statements ('meaning postulates,' using Carnap's term). Below, I will suggest that this is because the meanings of terms in a subject's language cannot always be expressed in other terms from that same language. That they can be expressed in the psychologist's language is one thing. But that does not mean that they can be expressed in the thinker's language.

The third question also concerned the relation between thought and language. The question was: Can we know the content of a thinker's thoughts before knowing the meaning of his words? That is, is the evidence that consists of a subject's *non-linguistic* behavior (and the facts about his non-linguistic environment) sufficient for fixing his propositional attitudes? If not, the Gricean assumption that mental content is prior to linguistic meaning might be in danger. We can ask the question using the hypothetical psychologist from earlier. Will he be able to attribute propositional attitudes to people on the basis of *only* their non-linguistic behavior and their non-linguistic environment? If not, the Gricean project is at risk.

It is important to see exactly what this issue amounts to. The question is *not*: Can we explain people's propositional attitudes without making reference to language whatsoever? The answer to this question is clearly 'No,' since people have all kinds of thoughts that involve language. (For instance, I might have a favorite poem.) Rather, the issue is that, since we want to explain psychological intentionality without relying on linguistic intentionality, we can *only* make reference to language *qua* non-intentional phenomenon. For instance, the account of mental content can make reference to language as a bunch of sounds or inscriptions, but not as having intentional properties.

Now, the content of some of our beliefs is closely linked to non-verbal behavior. In order to attribute to an individual the belief that a bear is chasing them, it is not necessary to appeal to the meanings of any words they might utter. Seeing them flee in horror from the bear is enough. But we also attribute beliefs to subjects on the basis of their *verbal* behavior. Similarly, knowing that they hold certain beliefs, we *expect* them to produce certain verbal behavior. The content of such beliefs is closely related to the meaning of what people say. Does this throw a wrench in the current approach? Do these difficulties cause trouble for the Gricean project?

No, for we have to realize that speakers very often use language *deferentially*. That is, speakers use linguistic expressions without knowing their exact meaning. Consider, for instance, Putnam's well-known example of 'elm' and 'beech.'¹⁹ Putnam discusses these terms in connection with the 'division of linguistic labor'—the fact that we rely on experts in delineating the extensions of many of our terms. Putnam's point is that someone can be a competent user of a term, without associating it with information that is detailed enough to fix its exact reference.

To make the example vivid, imagine Putnam going to a nursery to pick up an elm sapling. Arriving there, he says 'I need an elm sapling' and the merchant gives Putnam what he needs. Now, to explain how this task was accomplished, we need to attribute certain beliefs and desires. But at no point do we have to appeal to the *meaning* of 'elm.' Instead, Putnam's successful purchasing behavior can be explained by saying that

¹⁹ Putnam 1975.

he wants a tree that is called 'elm,' that he can get what he wants by saying 'I need an elm sapling,' etc. Putnam writes, "[m]y concept of an elm tree is exactly the same as my concept of a beech tree" (1975: 226). Exactly so.

Not all language use can be deferential, of course. There better be experts on elm trees and beeches who *do* know the meanings of these terms. When it comes to such experts, however, their beliefs about elms about can be accounted for in *non-verbal* behavior. These experts have various behavioral dispositions related to elm trees. They can discriminate them from other species of trees, they can predict where they will grow, etc. A functionalist can maintain that their thoughts are about elms in virtue of being causally connected, in certain ways, to non-verbal behavior and to the non-linguistic environment. To explain their mental state, no appeal to the meaning of their words is needed.

We can even imagine such an expert introducing the word 'elm' into the language. The introducer's mental grasp of the concept *elm* better not be a matter of the meaning of his words—for *there was no such word before he coined the term!* Few philosophers would want to maintain that his introduction of this term magically allows him to entertain thoughts he was not able to think before. To fast-forward a little, in Chapter 3 I will argue that the meaning of terms such as 'elm' is determined by the narrow beliefs of authoritative users of such terms. Introducing a term (a 'baptism') is one type of authoritative name use.

What about beliefs about abstract or logical matters that speakers express by using language non-deferentially? For instance, what about the belief in the law of non-

contradiction? On the current approach, some of these beliefs are not attributed on the basis of either non-verbal or verbal behavior. Instead, *they come for free*. This is a result of rationality being a constraint on interpretation. Simply by being a believer, one counts as a rational subject who *by definition* believes many logical principles. (Perhaps such attributions can be overridden by other considerations. But it seems to me that even when subjects' verbal behavior *appears* to reveal that they disbelieve such principles, we are more likely to think they do not understand what they say.)

Another question briefly worth addressing is whether narrow content deserves to be called 'truth-conditional content.' The answer is: Yes. A narrow belief of a subject represents his environment as being a certain way. If it is that way, the narrow belief is true; if it is not, the belief is false. This deserves to be called 'truth-functional content.' Unfortunately, many philosophers use the terms 'proposition' or 'truth-conditions' in a restricted way that excludes this use. Lewis describes such a philosopher of language as committing a type of "terminological piracy. He transforms one term after another into a mere synonym for 'singular proposition.' He has taken 'content.' He has taken 'proposition.' He is well on the way to taking 'truth-condition'" (1994: 318).

Finally, let me bring up *indexicality*. In the 1960s, John Perry called attention to the puzzling nature of indexical belief.²⁰ Lewis weighed in with an article called 'Attitudes *De Dicto* and *De Se*,' and provided a thought experiment that clearly

²⁰ Inspired by earlier work of Hector-Neri Castañeda and Peter Geach.

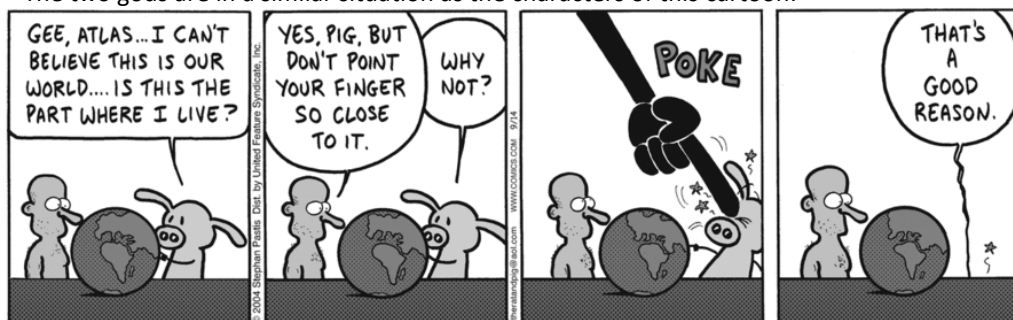
illustrates why indexical belief is special.²¹ In his example, there are two gods that know all qualitative facts about the world they live in. Both know that their world contains two gods: one on top of the tallest mountain and one on top of the coldest mountain. However, they do not know which one of the two gods they are. When they discover which one they are, there is something new that they come to believe. What type of belief do they gain?

Lewis argues that such *de se* beliefs are best represented with *properties*. When one of the gods comes to know he lives on the tallest mountain, this is best understood as that god self-attributing the property of being the individual on top the tallest mountain in a world with such-and-such a nature.²² And the other god self-attributes the property of being the individual on top of the coldest mountain. Since both gods have the properties they self-attribute, their beliefs are true. The content of these *de se* beliefs is evaluated for truth *at* individuals, not at worlds.

Applying this account to an example of Perry's, consider the case of the crazy Heimson who believes himself to be Hume.²³ This nut-job is so deranged that his inner

²¹ Lewis 1979.

²² The two gods are in a similar situation as the characters of this cartoon:



²³ Perry 1977.

life is indistinguishable from Hume's. When Heimson says, 'I wrote the *Treatise*,' his belief has the same content as the one that Hume expresses when he utters that sentence. Their beliefs have the same property as their narrow content. In this case, the single thought that they entertain is false at Heimson, but true at Hume. Heimson does not instantiate the property that provides the content of the thought, whereas Hume does.

In this work, we adopt this Lewisian treatment of indexical belief. In Chapter 3, I will return to the issue of indexicality when considering the meaning of pure indexicals such as 'I,' 'here' and 'now.' An important question is: If we understand meaning in terms of thought, then what exactly is the contribution of pure indexicals, such as 'I,' 'here,' and 'now,' to the meaning of the sentences in which they occur? It will turn out that pure indexicals differ in important ways from bare demonstratives and proper names, and that they need a very different treatment in the theory of meaning.

1.3 Conclusion: Descriptivism

Let me now return to the argument from Section 1.1 and spell out the conclusion in some more detail. According to this conclusion, the correct theory of meaning is a form of descriptivism. In this section, I will provide more support for the central argument and formulate a plausible version of descriptivism. In Chapter 3, this form of descriptivism will be applied to different categories of expressions. The theory will be applied to indexicals, demonstratives and names.

In Section 1.1, I introduced Premise 1 and argued that the meanings of sentences are the thoughts people use them to express. I claimed that this characteristic of meaning is definitive of the concept, and that as such it is not up for discussion. Reflecting on how speakers use language to transmit information to others—information they possess because they are believers—provides support for the view that sees the meanings of sentences as the thoughts that they express.

Thoughts, according to the theory of functionalism discussed in the previous section, are theoretical entities posited to do a certain explanatory job. ‘Belief’ and ‘desire’ are terms from folk-psychology for the internal states of people that give rise to behavior and that are the results of environmental effects. This folk-psychology implicitly defines ‘thought.’ What thoughts really are, on this view, is determined by a properly regimented version of this psychology. This is a narrow psychology; i.e. the *ceteris paribus* generalizations of this psychology will not contain names for environmental things.

Thoughts are best captured with modified Ramsified belief ascriptions of the form: S believes that there exists a unique n -tuple $x_1 \dots x_n$ $T(x_1, \dots, x_n)$. Every quantifier in the content clause corresponds to an entity or kind that the subject believes to exist. If the narrow contents of thoughts can be represented with such descriptive sentences, and if the meanings of sentences are the narrow contents of thoughts, then *meanings can also be represented with such descriptive sentences*. So we arrive at a version of descriptivism. As emphasized earlier, it is not part of the view that these meaning-capturing, descriptive sentences are from the same language as the expressions the

meaning of which they capture.

So far, we have been mostly concerned with the meanings of complete sentences, instead of the meanings of words. Let us briefly address how the current approach treats sub-sentential expressions, for instance names of individuals and properties. Here, too, we can take our inspiration from Lewis. He showed that it can be useful to transform Ramsey sentences in certain ways.²⁴ For instance, we can single out a particular entity that the Ramsified theory is about by saying: the i^{th} member of the sequence that is the unique realizing sequence of $T(x_1, \dots, x_n)$.

Now, suppose a thinking subject employs a name 'n' or something he believes to exist. Then we interpreters can provide the meaning of the name by saying:

The meaning of 'n' is: the i^{th} member of the sequence that is the realizing sequence of $T(x_1, \dots, x_n)$,

where $T(x_1, \dots, x_n)$ is the content clause from the belief attribution that captures the subject's beliefs. And if the subject produces a sentence like 'n is F,' then we interpreters can provide the meaning of the sentence by saying:

The meaning of 'n is F' is: the i^{th} member of the sequence that is the realizing sequence of $T(x_1, \dots, x_n)$ *instantiates* the j^{th} member of the sequence that is the

²⁴ Lewis 1970, 1972.

realizing sequence of $T(x_1, \dots, x_n)$.

Of course, not the whole believed theory is usually part of the meaning of the terms 'n' or 'F,' just a certain relevant part. What is the 'relevant part'? There is no one-size-fits-all answer for that question. This is an issue where a high level of specificity or precision should not be expected. To an extent, meanings are also negotiable and up to speakers themselves. (As seen in 'semantic decisions' about what to call certain instances of kinds. See Chapter 4's discussion of 'jade'.)

The account of meaning just sketched is 95% that of Lewis (except for a few tweaks to be addressed below). Lewis writes, "This is what I have called functional definition. The T-terms have been defined as the occupants of the causal roles specified by the theory T; as *the* entities, whatever those may be, that bear certain causal relations to one another and to the referents of the O-terms" (1972: 255). Names, on the picture that Lewis favors, "can be treated as definite descriptions" (252).

We depart from Lewis in that meanings are expressible in the interpreter's language, but not always the speaker's language. The meaning-providing expression 'the i^{th} member of the sequence that is the realizing sequence of $T(x_1, \dots, x_n)$ ' is an expression in the language of the hypothetical psychologist. As noted earlier, this has certain repercussions when it comes to questions of analyticity and definability. For Lewis, there are many analytic truths of the form: If there exists a unique entity that plays such-and-such a role (e.g. the water-role), then that entity is N (e.g. water). On our view, there is no guarantee that there are many of these statements, for the theory that

defines expressions is *believed* by speakers and may not be expressible in these speakers' language.

Frank Jackson agrees that descriptivism is not committed to the claim that the meanings of terms must be expressible in that same language. When talking about meanings, he writes, "[i]t is not an essential part of the theory that we should have words, or 'other words,' for these properties" (1998: 203-4). And in one of his descriptivist phases, even Kaplan argued for a similar point. He writes that "stored images are simply one more form of description, worth perhaps a thousand words, but thoroughly comparable to words" (1968: 236).

Because of his view on meaning, Lewis holds the strong view that even *color terms* are definable. But what would a definition of such a term look like? According to Lewis, the definition of 'red' differs from that of 'green.' But exactly what is the difference? In his article 'Naming the Colors,' Lewis considers this question. He suggests that it is part of the meaning of 'red' that certain things are red (e.g. fire trucks, tomatoes), and that it is part of the meaning of 'green' that certain *other* things are green (e.g. grass, limes). This seems a counter-intuitive view. Suppose I believe that X's, Y's and Z's are red. I can then say, 'X's and Y's and Z's are not red.' Taken by itself, this statement does not appear to be a contradiction. What is more, I am in full cognitive possession of the meaning of 'red' (that is, I do not use the term deferentially).²⁵

²⁵ Lewis 1997. Lewis also discusses the following problem. A group of language users may be separated into two sub-communities. For instance, one sub-community that lives in a region where Xs, Ys and Zs are red and where it is part of the meaning of 'red' that they are. And another sub-community from a region where instead As, Bs and Cs are red and where it is part of the meaning of 'red' that they are. Now, if a member of the former community uses the term 'red' in a conversation with a member of the second community, will communication not falter? There is no common knowledge of what they mean by the

A more common-sensical approach to color terms is that the meaning of 'red' is partly determined by our beliefs about how it makes things look. That red is a surface property of objects that makes things look a certain way is included in its meaning. This is part of the conceptual role of the color term 'red.' On this account, Jackson's color-blind neuroscientist Mary does not know the full meaning of the term 'red.' The fact that color-*seeing* people cannot express this part of the 'red'-theory in words does not mean it is not part of the theory. It is part of the *believed* theory. In Chapter 4, I return to these issues.

Over the past 50 years, description theories of meaning have been much maligned. Many philosophers believe the view has been shown to be untenable. Descriptivists, on the other hand, consider many of the criticism of the view not as knock-down arguments, but rather as revealing that the meanings of various terms are more complicated than supposed. In the remainder of this section, I will consider some of these criticisms. Many of them have a common structure. First, claims are made about how the reference of certain expressions is constrained (e.g. by causal factors). Second, it is argued that the beliefs we associate with these expressions do not involve these constraints. It is then concluded that the associated beliefs do not determine

word. Lewis attempts to solve the issue in terms of *existential* common knowledge: both speakers know that i) some definition of 'red' is employed by the one group, ii) some definition of 'red' is employed by the other group, and iii) the two definitions of 'red' agree on which things are red. While this may solve the issue of communicative coordination, it seems to me that it runs into trouble when we consider the function of language as representing the nature of the world. Imagine a speaker from the first and a speaker from the second group facing a red wall. When they both say, 'That is red,' they represent the world as being the *same way*. Furthermore, this representation is richer in content than Lewis can allow with his notion of 'existential common knowledge.'

reference.

Many of the criticisms originate from the work of Kripke and Putnam.²⁶ Let me discuss *three* different ones. The first centers around *causal* constraints, the second around *naturalist* constraints, and the third deals with *indexical* constraints. The reply to Kripke and Putnam on behalf of the descriptivist is roughly similar in all three cases. The proper response for the descriptivist is simply to claim that the beliefs that provide the meanings of certain expressions *do* involve these constraints. Kripke and Putnam's arguments have merely shown the beliefs that provide the meanings of certain terms to be more sophisticated than many initially thought.

First, *causality*. A common complaint about descriptivism is that it does not give causality its proper place. To take an example, consider Kripke's example of Leverrier (on which more later). Leverrier was a French astronomer who coined the name 'Neptune' for the planet that caused changes in the orbit of Uranus. The criticism then is: the beliefs that speakers associate with a name like 'Neptune' are usually not couched in causal terms. Yet, the reference of a term like 'Neptune' is largely dependent on causal constraints. Therefore, the beliefs that speakers associate with many terms do not determine their reference.

Descriptivists standardly reply with *causal descriptivism*, which has been called by Brian Loar, "the causal theory made self-conscious" (1980: 86).²⁷ On causal

²⁶ Kripke 1972, Putnam 1975.

²⁷ Kripke discusses such a view as well, see 1972: 88fn. See also Loar 1976 and Kroon 1987.

descriptivism, the meaning-providing beliefs that we associate with many of our terms *do* include causal constraints. In other words, causality plays a role that is *internal* to the theory, instead of external. This view receives further support from the realization that among the many beliefs we have about entities in the environment, some of the most central ones are beliefs about how we are causally related to these entities. There are few things that we are more certain of than that we are causally related in various ways to the things in the world around us.

Second, there are what can be called ‘naturalist’ constraints. Both Kripke and Putnam’s arguments show that whether or not something is part of the extension of ‘water’ or ‘tiger’ is not a matter of its superficial appearance. Rather, whether a sample or instance belongs to these kinds depends on its underlying structure. If what ‘water’ referred to were merely a matter of appearance, then ‘water’ would refer to any transparent liquid (which includes ethanol). It would not refer to muddy water. And if what ‘tiger’ referred to were merely a matter of appearance, then ‘tiger’ would apply to realistic-looking robot-tigers, and not to albino tigers.

The typical criticism of descriptivism, then, is the following. People associate with expressions like ‘water’ or ‘tiger’ characteristics about the superficial appearance of these kinds. But what these terms refer to is not a matter of superficial appearance. Therefore, what people associate with ‘water’ and ‘tiger’ does not determine the reference of these terms. The thought is that reference is fixed by theory-external constraints that involve *naturalness*. Things that share their underlying nature (but not things that present a similar superficial appearance) form a sufficiently natural collection

and, so, are suitable candidates for reference.

The reply on behalf of the descriptivist is similar to the one before. According to ‘naturalist descriptivism,’ the meaning-providing beliefs that we associate with many of our terms *do* involve constraints on whether its extension is a natural collection. In order for something to be the extension of ‘water’ or ‘tiger,’ it has to be a sufficiently natural collection.²⁸ When the descriptivist has made this move, the criticism loses its punch. (Of course, questions arise. “What exactly is naturalness?” “What does the average speaker know about naturalness?” I will return to questions about the analysis of naturalness in Chapter 4.)

At this point, critics of descriptivism may accuse the theory of *over-intellectualization*. Surely, it is implausible—they say—to think that your average language-user has such complex referential intentions. Can speakers not just pick up various words that they encounter and proceed to use them without these highly sophisticated intentions? This charge has merit. But the charge of *over-intellectualization* does not threaten descriptivism, once we distinguish authoritative from deferential language-use. Deferential language use does indeed require very little, *not even* knowledge of the meaning of the expression in question.

I will argue in Chapter 3 that the distinction between authoritative and deferential language use is much more important than often is supposed. Drawing this distinction can help shed light on the various purposes for which we use language. I will

²⁸ McDermott 1988.

argue that deferential language use is very central to linguistic communication. There is even a sense in which deferential language use allows us to entertain thoughts that we would not be able to entertain without language. I will postpone further discussion of this important issue until Chapter 3.

Let me now turn to a third criticism of descriptivism. This criticism is based on Putnam's Twin-Earth thought experiment.²⁹ In this well-known thought experiment, we are asked to consider a planet called 'Twin Earth' that is similar to Earth except that the transparent stuff that fills the lakes and falls out of the sky is XYZ instead of H₂O. Oscar on Earth uses 'water' to refer to H₂O, while his twin on Twin-Earth uses 'water' to refer to XYZ. Suppose further that Oscar and his twin live in the year 1750, when nothing was known about the chemical composition of water.

Oscar and Twin-Oscar are alike in many respects: they exemplify the same functional/causal organization, they have the same qualitative evidence available, etc. In fact, the psychology that we envisioned earlier in this chapter would ascribe to Oscar and his twin the exact same thoughts. Their thoughts—about water, about anything else—have the same narrow content. Putnam concludes that if meaning determines reference, grasping a meaning cannot be a matter of being in a narrow psychological state. He concludes that "meaning just ain't in the head!"

The standard reply on behalf of the descriptivist is to appeal to *indexicality*. As discussed at the end of Section 1.2, narrow contents should be understood as

²⁹ Putnam 1975.

irreducibly *de se*. The psychological contents of thoughts are best regarded as properties, as Lewis argued. These contents are evaluated at individuals instead of at worlds. The narrow content of Oscar and Twin-Oscar's thought is indeed the same. But this content, when evaluated at the two different individuals, Oscar and Twin-Oscar, delivers different referents: H₂O in Oscar's case and XYZ in Twin-Oscar's case. If Oscar and Twin-Oscar both say, 'Water consists in part of hydrogen,' Oscar's utterance is true and Twin-Oscar's is false. Yet, they have the same meaning.

An early descriptivist like Peter Strawson already emphasized the importance of indexicality. He wrote that "the 'identifying description' may... include demonstrative elements, i.e. it need not be framed in purely general terms" (1959: 182fn). He even imagined a scenario resembling Putnam's. A Twin-Earth *avant la lettre*. When singling out particulars by their properties, writes Strawson, "there might be another particular, answering to the same description, in another section of the universe. ... [T]his possibility of massive duplication remains open." But to think that this is a problem, "supposes that... [the particular's] identification must rest ultimately on description in purely general terms. But this supposition is false. ... [I]t may be identified by a description which relates it uniquely to another particular which can be demonstratively identified" (1959: 20-1; italics added).

I bring up Strawson not to suggest that he had anything like Putnam's argument in mind or that he had a detailed view on *de se* belief. Rather, I merely want to indicate that early descriptivists were well aware of the fact that reference should not be understood as a purely qualitative characterization that determines a unique entity in a

world. In addition, the fact that a certain pre-Putnam view of situated reference can be used to deal with Putnam's Twin-Earth argument only makes the case for this view stronger. I will return to these issues in later sections when I discuss referring with pure indexicals such as 'I,' 'here' and 'now.'

CHAPTER 2

REFERENCE

2.1 An Argument Relating *De Re* Belief and Reference

Just like in Chapter 1, I will again present a three-part argument. This time, the argument deals with reference. The first premise relates *de re* belief to reference, the second premise concerns the correct analysis of *de re* belief, and the conclusion spells out what this means for reference. The structure of the argument is similar to that of the previous chapter. In Chapter 2, the nature of belief was used to understand on the notion of meaning. In this section, the nature of *de re* belief is used to shed light on the nature of reference.

Without further ado, here is the argument:

An Argument Relating *De Re* Belief and Reference

Premise 1: Linguistically expressing a belief that is *de re* with respect to an entity is to refer to that entity.

Premise 2: That a belief is *de re* with respect to an environmental entity is a contingent property of that belief. Furthermore, whether a belief counts as *de re* depends upon the attributor's context.

Conclusion: Therefore, that an expression refers to an entity is a contingent property of that expression. Furthermore, whether an expression counts as

referring depends upon the attributor's context.

Just like before, I will spend the current section discussing Premise 1 and the general structure of the argument. The next section, Section 2.2, is devoted to Premise 2 and takes a look at different accounts of *de re* belief. Finally, in Section 2.3, I spell out the conclusion by sketching the implications for the concept of reference.

What exactly do I mean by 'belief' and 'expression' in the premises and conclusion? Belief and expression *types* or *tokens*? The answer is: types. Beliefs are typed psychologically, as discussed in the previous chapter. What about expression types? They are individuated in terms of their meanings. For unambiguous names, this is straightforward. Individuating them by meaning has the same results as individuating them by word shape. But for context-sensitive expressions, this does not work. So in order to individuate a demonstrative like 'that' in the right way, we take into account the word shape but *also* the conversational context. The conclusion of the argument says that the referential properties of an expression *as used with a certain meaning* are contingent. More on this below.

Does the argument from this section require that the reader accepts the argument from the previous section? No, it does not. The view on narrow content from the previous content and the view on non-narrow (i.e. wide) content from this section *do* make a good package deal, but they are not dependent on each other. Someone unwilling to accept the functionalist account of narrow content from the previous section can accept many of the arguments and claims presented in the current one. And

vice-versa, a friend of narrow content could refuse to accept any of the claims from this section.

First, a few words on some of the terms from the argument: '*de re* belief' and 'reference.' The term '*de re* belief' is a technical term. This means that there is a risk of turning the first premise into an empty claim, because it amounts to nothing more than a stipulation of how a technical term will be used. This would not move things forward. There is also the risk of attempting to explain a common-sense notion ('reference') in terms of a technical notion ('*de re* belief'). How useful can that be? Do we not already know what reference is? In order to avoid such problems, we must connect the term '*de re* belief' to some common-sense notions, in order to make sure Premise 1 is substantive and clarifying.

The notion of *de re* belief can be regarded as a way of *typing* belief tokens in terms of what they happen to be about. As a starting point, let us say that a *predication* is the attribution of a property to an individual. Predications, then, are individuated in terms of individual-property pairs. When two subjects perform the same predication, they stand in a relation to the same individual-property pair, or the same singular proposition. Consider now a *de re* belief attribution of the form, 'S believes that a is F'. If this attribution is true, and if 'a is F' expresses a singular proposition, and if 'a' is co-referential with 'b,' then the attribution 'S believes that b is F' will be true as well. Both are claims that a certain relation holds between a subject and a singular proposition.

We have learned a lot about these matters from Quine and Kaplan. They discuss *de re* belief in two seminal articles: Quine's 1956 "Quantifiers and Propositional

Attitudes” and Kaplan’s 1969 “Quantifying In.” The well-known example from Quine’s article is that of Ralph, who has seen Bernard J. Ortcutt behave suspiciously and has him pegged for a spy. Ralph believes *of* Ortcutt that he is a spy. This example is a Frege-case, for Ralph later encounters Ortcutt again, not realizing it is one and the same person. Kaplan, in his discussion of the example, claims that Ralph is in possession of a ‘vivid name’ for Ortcutt.³⁰ A common view is that one can have vivid names of individuals one perceives or is in another way intimately causally connected with.

Because Ralph has a vivid name for Ortcutt, says Kaplan, the sentence ‘Ralph believes that Ortcutt is a spy’ implies the sentence ‘Ortcutt is such that he is believed by Ralph to be a spy.’³¹ The name ‘Ortcutt’ can take wide-scope without changing the truth-value of the sentence. Such belief attributions ascribe *de re* beliefs. On a view inspired by these remarks, having a vivid name of an individual allows one to entertain *de re* beliefs about that individual. When a person entertains such a *de re* belief about an entity, he stands in a certain relation to a singular proposition involving that very entity.

What about the term ‘reference’? While this term is not as technical as the notion of *de re* belief, let me nevertheless say a couple of things before proceeding. Peter Strawson once said in response to Russell that “‘mentioning,’ or ‘referring,’ is not something an expression does; it is something that someone can use an expression to

³⁰ Kaplan 1969.

³¹ Just like vivid designators allow exportation out of propositional attitude contexts, rigid designators allow exportation out of modal contexts. The *de dicto* modal statement ‘Necessarily, Aristotle is human’ implies the *de re* modal statement ‘Aristotle is such that he is necessarily human.’ See Quine 1977.

do" (1950: 326). My usage here will correspond mostly to that of Strawson's. Referring, in the first place, is a *speech act*. It is something that speakers do, with an expression, usually in the presence of an audience, while uttering a sentence in some language.

There is also a sense of 'reference' that applies to *expression types*, not speakers. For instance, the name type 'Aristotle' can be said to refer to Aristotle. And expression types 'I' and 'that' can be said to refer, relative to a context of use, to the speaker or to a contextually salient object. I consider these uses of 'refers' as *derivative*. That is, this derivative sense of 'refers' ultimately is to be explained in terms of the primary sense of 'refers' that applies to speech acts. In what follows, I will occasionally switch from using the one use of 'refers' to the other, and even use both at the same time, but this should not cause confusion.

Let us now return to Premise 1: *Linguistically expressing a belief that is de re with respect to an entity is to refer to that entity*. The best way to support this premise is by considering how both *de re* beliefs and acts of reference are individuated. Both *de re* beliefs and acts of reference are individuated in terms of 'predications,' or singular propositions. This supports the view that linguistically expressing a belief that is *de re* with respect to an entity is to refer to that entity.

Consider Ralph again, who believes *of* Bernard J. Ortcutt that he is a spy. Ralph can express this belief by saying *of* Bernard J. Ortcutt that he is a spy. Both these phenomena (Ralph's believing and Ralph's act of saying) are individuated in the same way. Ralph's belief *of* Bernard J. Ortcutt that he is a spy is the same as his belief *of* a certain pillar of the community that he is a spy. And Ralph's act of saying *of* Bernard J.

Orcutt that he is a spy is the same as his act of saying *of* a certain pillar of the community that he is a spy. In fact, his act of referring is nothing but the linguistic expression of this *de re* belief. It appears, then, that Premise 1 is plausible and well-supported.

2.2 The Second Premise: *De Re* Belief

On the functionalism sketched in the previous chapter, propositional attitudes are narrow—i.e. they supervene on the local, intrinsic state of the subject. A question that arises is: How do these narrow propositional attitudes relate to so-called *de re* beliefs about environmental entities? How do they relate to *wide* propositional attitudes that depend partly on the world surrounding the subject?

In the 1970s, philosophers debated whether or not such *de re* attitudes are reducible to *de dicto* attitudes. Roderick Chisholm, for instance, argued in favor of such a reduction, whereas Tyler Burge did not think it possible.³² Lewis sided with Chisholm, but did not defend a reduction to *de dicto*. He writes that, “the analysis of belief *de re* that I propose... is a reduction of *de re* generally to *de se*. [...] [T]here can be no reduction of *de re* to *de dicto*.” (1979: 157). Lewis does not think that a reduction to *de dicto* can be carried out, because of the irreducible nature of indexicality discussed above.

³² Chisholm 1976, Burge 1977.

How exactly do *de re* beliefs reduce to *de se* beliefs? Lewis says, “Beliefs *de re* are not really beliefs. They are states of affairs that obtain in virtue of the relations of the subject’s beliefs to the *res* in question” (1979: 152). According to the functionalist, *de re* beliefs are not narrow attitudes that are related to environment, action, and other internal states. Whether the person Ralph believes to be a spy is Bernard J. Ortcutt or his twin brother Robert F. Ortcutt makes no difference to Ralph’s actions, inferences, etc. Ralph’s narrow belief can be represented by something like: *the unique person I am acquainted with under such-and-such a mode of presentation is a spy*. He counts as having the *de re* belief that Ortcutt is a spy in virtue of the environmental fact that the person he is so acquainted with is in fact Bernard J. Ortcutt.

On this picture, there are two ways to *type* the contents of beliefs—narrow and wide. Typing a belief in a certain way is to count certain other belief tokens as having the same content as that belief. One option is to type beliefs in terms of their cognitive role in the thinker’s psychology. This means we regard a certain class of belief tokens as having the same content as that belief (even though these tokens may be about different entities). The other option is to consider the believed-about entity as constitutive of the belief content. This means we regard some *other* class of belief tokens as having the same content as that belief (even though these tokens may play different roles in these thinkers’ psychologies). But in both choices, we miss out on something. Either we miss out on the identity of the believed-about entity, or we miss out on the mode of presentation of that entity.

The question arises: Why do we attribute *de re* beliefs to people? A promising

line is the suggestion that when we are interested in people as carriers of information about particular individuals or kinds in the environment, we tend to characterize their beliefs as *de re* about these things. When both Ralph and his neighbor suspect Ortcutt to be a spy, then even though they conceptualize Ortcutt in different ways, there is a sense in which they believe the same thing. The FBI will be interested in talking to Ralph and his neighbor because both are carriers of information about Ortcutt. The FBI is interested in what Ralph and his neighbor believe characterized *widely* (*de re* with respect to environmental entities), not characterized *narrowly* (*egocentric, de se*).

When it comes to beliefs about things we are perceiving, or things we are causally impacted by, or things we have extensive mental ‘dossiers’ about, we tend to type beliefs in a *de re*-manner. Brian Loar sketches such a picture when he claims that some modes of presentation are *identifying*, while others are not. He writes that certain ways of conceiving individuals, “[form] a class of... individuals concepts which are naturally regarded as identifying *simpliciter* because they are identifying relative to those pragmatically important classificatory interests” (1976: 364) Kaplan makes a similar point; he writes that the attribution of *de re* beliefs depends on “conditions relative to the topic, interests, aims, and presuppositions of a particular discourse” (1989: 606).

Consider Sherlock Holmes investigating a murder. At first, Holmes merely believes there to be a unique murderer who did such-and-such. Few would be willing to consider this a *de re* belief about the murderer. As his inquiry progresses, he comes in contact with more and more causal traces left by the murderer. Maybe he sees a

footprint. Maybe he encounters the killer in disguise. When Holmes eventually apprehends the culprit, we are willing to attribute a *de re* belief. Exactly when does this change? There is no principled way to draw a line. Brian Loar writes, “somewhere along the line truth-conditions in this range of cases flip over from being a generalized proposition to being a [singular proposition]. But at what point is that motivated? Of no adjacent pair of these cases is it plausible to suppose that their contents are of such radical different kinds.”³³

If the attribution of *de re* beliefs is governed by such explanatory and classificatory interests, one would expect there to be a grey area. One would expect attributors in different contexts to disagree on whether a given subject counts as having a *de re* belief. This is exactly what we find. Consider, for instance, Kripke’s example of Leverrier who introduced the name ‘Neptune’ for the unique planet responsible for perturbations in the orbit of Uranus (1972: 79). Did Leverrier have *de re* beliefs about Neptune? Some argue that he did, because he stood in indirect causal relations to Neptune. Other philosophers argue that such ‘thin’ causal relations are not sufficient for being in a position to have *de re* beliefs about the entity in question. Who is right?

Something similar applies to Kripke’s puzzle about belief. In this puzzle, the French Pierre hears about a beautiful city in England called ‘Londres.’ He is inclined to say ‘Londres est jolie.’ Later, Pierre moves to London where he rents an apartment in an ugly neighborhood. After learning English, he is inclined to say ‘London is not pretty.’

³³ Loar 1976: 357. See also Lewis 1979: 381.

Pierre never realizes that the name 'Londres' picks out the same thing as 'London'. Now, does Pierre, or does he not, believe of London that it is pretty? In presenting his puzzle, Kripke insists that he is not concerned with *de re* belief (1979: 424). Others disagree and claim that the puzzle does involve *de re* belief (Lewis 1981). Who is right?

In both these cases, a promising strategy seems to endorse a thesis of context-sensitivity. Of course, one should not simply invoke context-sensitivity whenever there is disagreement among philosophers. But the disagreement between the philosopher who claims that Leverrier has *de re* thoughts about Neptune and the philosopher who insists that he does not feels insubstantive. It seems plausible to approach this disagreement by saying that it is verbal only. The disagreement between the two philosophers appears to be a dispute about the application of a term that is sensitive to context.

Quine also regards such *de re* belief attributions as context-sensitive. He writes that whether one counts as having a *de re* belief about Ralph depends on the relevant interests. "The notion of... believing who or what someone or something is, is utterly dependent on context. Sometimes, when we ask who someone is, we see the face and want the name; sometimes the reverse. Sometimes we want to know his role in the community. Of itself the notion is empty" (1977: 121).³⁴

(Note that the suggestion is not merely that the *believer's* context determines

³⁴ See also Hawthorne and Manley who write, "[W]e are completely puzzled by the kind of policing or gate-keeping that goes on when philosophers make speeches of the form, 'In order to have a genuine singular thought/have a genuine *de re* thought one must satisfy constraints X, Y, and Z.'" (2012: 26).

whether he has a *de re* belief about a certain individual. This is a correct, but uncontroversial, claim. The view under consideration is the stronger one according to which the *attributor's* context determines whether the believer has a *de re* belief at all. This stronger view holds that the explanatory and pragmatic concerns of the attributor's context determine whether it is appropriate to ascribe *de re* beliefs to subjects.)

The previous remarks about *de re* belief can be connected with a view called '*liberalism*.' According to liberalism, thinkers can have *de re* thoughts about anything that they can qualitatively single out in thought. There are no restrictions in terms of causal, historical, or acquaintance-relations. Liberalism can be clarified using Kaplan's 'dthat' operator.³⁵ This operator, when applied to a definite description, turns the description into a directly referential expression that contributes the singled-out object to the proposition. The sentence 'Dthat[the shortest spy] is a spy' expresses the singular proposition that Ortcutt is a spy, instead of the general proposition that there exists an individual that co-instantiates *being the shortest spy* and *being a spy*. 'Dthat', applied in the realm of belief, would turn the *de dicto* belief that the shortest spy is a spy into the *de re* belief that Ortcutt is a spy.

Usually, 'the shortest spy' does not count as a vivid name that allows its user to entertain a *de re* belief about Ortcutt. But the name 'Ortcutt,' or the demonstrative 'that' accompanied by a pointing gesture to Ortcutt, *does* count as a vivid name. In the philosophy seminar room, we are free to set aside such contextual concerns and count

³⁵ Kaplan 1977.

any qualitative means of singling out an individual as a ‘vivid name.’ Even speakers who use definite descriptions attributively are able to entertain *de re* thoughts about the individuals singled out by these descriptions. Kaplan writes, “If pointing can be taken as a form of describing, why not take describing as a kind of pointing?” (1978: 392).³⁶ On liberalism, we can use ‘dthat’ to “apprehend singular propositions concerning remote individuals (those formerly only known by description)” (Kaplan 1977: 560).

Functionalism about mental content is independent of contextualism about *de re* belief, but the two views fit well together. On the type of functionalism considered here, beliefs are by definition narrow. There is *no difference in kind* between Sherlock Holmes’ belief at the beginning of the investigation and his belief when he has arrested the suspect. In many contexts, we are unwilling to attribute to Sherlock Holmes a *de re* belief about the culprit in the former situation, and we are willing to attribute to Sherlock Holmes such a *de re* belief later on. This reveals nothing important about Sherlock’s psychological state, however. It merely reflects the interests and concerns that are operative in the context.

The fact that some *de re* belief attributions are sensitive to context does of course not show that they all are. There may be beliefs that count as *de re* in all contexts, and there may be beliefs that do not count as *de re* in any context. For instance, one could argue that it is impossible to have *de re* beliefs about future

³⁶ The term ‘liberalism’ most likely originates from Kaplan. In “Demonstratives,” he defends a view that allows a lot of ‘liberality’ when it comes to the introduction of names (1977: 560). For a recent defense of liberalism, see Hawthorne and Manley 2012.

individuals. But even here, it seems to me, there is room to argue. Imagine a community of clairvoyants who can reliably 'see' into the future and predict that certain events will happen. In the right kind of context, I see no reason why we would not ascribe them *de re* beliefs about future individuals (for instance, a *de re* belief about the first child born in the 22nd century, Newman I.³⁷)

Now, why would one accept the foregoing view on narrow content and *de re* belief? One source of support is that alternative views have not developed into promising research programs. In the 1980s, for instance, Fred Dretske and Jerry Fodor attempted to construct causal/informational theories of mental content.³⁸ Some of these approaches propose explanations of mental content in terms of reliable indication, asymmetric dependence, information-theoretic notions, and other concepts. These forms of 'psychosemantics,' however, turned out to face serious challenges and have not flourished. Of course, this is not to say a causal theory of *de re* belief cannot be constructed. But perhaps the lack of success provides a reason to look at competitors, such as the current approach that takes narrow content as primary.

It is now time to return to indexicality and *de se* belief. How do the *de se* beliefs that speakers express using pure indexicals like 'I' relate to *de re* belief? As mentioned at the end of the previous section, Lewis showed that the psychological content of such beliefs can be represented with properties. This psychological content is independent of

³⁷ Kaplan 1977: 560; Kaplan 1978: 397.

³⁸ Dretske 1981, Fodor 1987.

the subject's environment (i.e. it is narrow) and captures the belief's cognitive significance. But when a speaker uses 'I,' he also expresses a *de re* belief where the *res* is the believer himself. Such a *de re* belief has as its content a singular proposition that involves the believer. This singular proposition is independent of the subject's environment.

Will a regimented psychology that captures people's real propositional attitudes ascribe beliefs that have as contents singular propositions containing these very people? No. Even though people's beliefs that are *de re* with respect to themselves are *narrow*, they nevertheless fail to capture cognitive significance. If the regimented psychology would ascribe beliefs in such singular propositions, it would be individuating beliefs in *too fine-grained* a manner. (Usually the complaint about singular propositions is that they are too course-grained—but here we have the opposite problem. They are too fine-grained.) Heimson and Hume entertain the same thought—not different thoughts.

Heimson stands in the relation of identity to himself, so when Heimson uses a sentence containing 'I,' we have no qualms about attributing to him a *de re* belief about himself. Identity is a relation that is intimate enough to meet the contextual requirements for *de re* belief. So when it comes to the attribution of *de re* beliefs about oneself, contextual pressures do not play the important role that they play in attributions of *de re* beliefs not about oneself. Nevertheless, we maintain that narrow contents are never correctly represented by singular propositions containing the thinking subjects themselves.

This is one place where the view favored here differs from Russell's. Even though

I use the label 'Russellian descriptivism,' we do not share with Russell the view that, when it comes to mental content, persons can be constituents of propositions. In a famous passage, Russell imagines Bismarck having an indexical belief. Russell writes that there is "a judgment which Bismarck alone can make, namely, the judgment of which he himself is a constituent" (1910: 218). Here, we depart from Russell. Bismarck is not a constituent of the narrow content of any of his thoughts. (But can *any* entities be constituents of such contents? Again, see Chapter 4 for more discussion.)

2.3 Conclusion: A View On Reference

It is now time to move on to reference and draw conclusions from the two premises. In Section 3.1, I argued that linguistically expressing a *de re* belief is to refer. Section 3.2 dealt with *de re* belief. I argued that the notion of *de re* belief can be seen as a non-psychological way of typing belief tokens in terms of what they are about. Given the psychological individuation of beliefs explained in the previous chapter, the *de re* characterization of beliefs captures only contingent properties. That a belief is *de re* with respect to some entity is not essential to that belief. Furthermore, whether beliefs count as *de re* at all depends upon the attributor's context. In this section, I will spell out what this means for reference. The resulting view will also be supported with some examples.

The arguments from this section also reveal, in my opinion, that reference is less central to the theory of meaning than often is supposed. If referential properties are both non-essential and furthermore context-sensitive, then they seem for that reason

not as fundamental. What is central to the theory of meaning is the stable nature of linguistic representational acts across contexts. Less important is the question whether these linguistic representational acts amount to reference. In the philosophy seminar room, we can even be liberalists and adopt a Fregean, permissive standpoint where very many speech acts and expressions count as referential.

(There is another issue, however, that is related to reference and that is central to the theory of meaning. This issue can be called—somewhat misleadingly—the question of ‘mental reference.’ In the discussion of Russell, we saw that Russell employs very strict acquaintance requirements for reference. For him, very few expressions are logically proper names. In fact, when Russell speaks of ‘logically proper names,’ he is not concerned with language only, but also with *thought*. Russell’s logically proper names are those elements of thought that represent directly. They do not single out entities by their qualities, but represent them without mediation.

A version of this thesis returned in our discussion of functionalism in Section 1.2. There, the question arose: Will the attitude attributions of a regimented psychology contain names? That is, will the Ramsified belief attributions impute purely ‘structural’ attitudes or will they also contain non-logical expressions? Lewis’s view at some point appears to have been that as far as behavioral and environmental evidence is concerned, only belief attributions with structural content clauses are warranted. I will return to this issue in Chapter 4, where I will provide textual evidence that Lewis held this view, criticize it, and sketch an alternative.)

Now, let us return to the central argument. I will first address the contingency

claim, then the context-sensitivity claim. When a speaker refers, he uses an expression with a certain meaning. An expression's meaning is whatever it contributes to the thought expressed by the sentence in question. That the speaker's thought is *de re* with respect to a certain individual in a certain context is contingent (i.e. it can be *de re* with respect to another individual in another context). That an expression's meaning is '*de re*' with respect to an individual is then contingent as well. If reference is the expression of a *de re* belief, we conclude that the referential property of the expression is contingent also. The same expression be used with the same meaning to refer to another individual in another context.

In the next chapter, this view will be spelled out in more detail. There, I will claim argue that an expression's meaning is essential to it (*qua* expression of the language to which it belongs), but that an expression's reference is contingent. The meanings of expressions, on this approach, are semantic; the referential properties of expressions are not. They are extra-semantic, they do not contribute to the character of the language in question. On this picture, linguistic competence consists of knowledge of sentence-thought relations, i.e. knowledge of meaning. Knowledge of reference is not merely a matter of linguistic competence; it also requires knowledge of the non-linguistic world.

Now for the context-sensitivity claim. In the previous section, I argued that whether a belief counts as *de re* depends upon considerations operative in the attributor's context. If this is so, then what counts as linguistically *expressing* a *de re* belief is a matter of context as well. This means that whether or not a linguistic

representational act counts as reference or not depends upon the attributor's context. In order to make this conclusion more plausible, let me consider some example utterances. I will look at speech acts involving demonstratives as well as ones involving proper names.

First off, consider 'this' and 'that.' These bare demonstratives have non-referential, 'attributive,' uses. Brian Loar considers a speaker who sees a large empty shoe and says about its owner, whoever it may be, "He is rather big" (1976: 357). According to Loar, this is an attributive use of 'he.' No singular proposition is expressed, no reference has taken place. Now, this intuition is actually debatable. Does Loar refer to the owner of the shoe? Does Loar say, *of* the wearer of the shoe, that he was rather big? Loar suggests that he does not.

The correct approach, in my opinion, is to say that whether or not reference took place is a context-sensitive affair. In some contexts, Loar counts as referring; in some, he does not. Does Loar express the belief that a particular person is big, or merely the belief that somebody or other co-instantiates being male and big? It depends. In some contexts, seeing a person's shoe is not sufficient for knowing *which* person is in question. In other contexts, seeing a person's shoe *is* sufficient for knowing which person is in question.

We can imagine contexts where the connection between a shoe and its wearer is not very robust: for instance, a situation where people discard shoes constantly and leave them lying around. Perhaps in such a context, Loar would not count as referring because the information available to him would not allow one to track down the

individual. In a situation where this connection is more robust, we would be willing to describe Loar as referring to a specific individual.³⁹

Next up, proper names. Already in *Naming and Necessity*, Kripke calls attention to so-called ‘descriptive names’—names that are introduced by description. A well-known example is ‘Neptune.’ Leverrier introduced the name ‘Neptune’ for the planet that caused perturbations in the orbit of Uranus.⁴⁰ The French astronomer created a term for an entity that he believed to exist. We can even imagine how such a process might take place: someone can single out an object qualitatively, perhaps by using a definite description, and perform a baptism whereby the object is given a name. (See Chapter 3 for more on baptisms.)

Gareth Evans also discusses names introduced by description. His example is ‘Julius,’ which is introduced as follows: “Let us use ‘Julius’ to refer to whoever invented the zip” (1979: 163). In the literature, people debate back and forth on whether

³⁹ Another example of a non-referential use of a demonstrative is from Geoffrey Nunberg. He considers an example of a speaker who says, while pointing at the pope, ‘He is usually Italian’ (2004). In this sentence, ‘usually’ is an adverb of modification that applies to the adjective ‘Italian.’ Clearly, the speaker does not intend to say that one and the same person is oftentimes Italian but sometimes not. Rather, he means that a certain ‘office’ is oftentimes occupied by an Italian person but sometimes not. So if we allow reference to offices or roles (i.e. if they can be constituents of singular propositions), then we can say that in Nunberg’s example reference did occur. This example is different in kind from Loar’s.

Hawthorne and Manley also consider a non-referential use of a bare demonstrative. Consider a student who just did poorly on a calculus exam. He is bad at math, whereas history is his strong point. The student might utter, “If that had been the history exam, I’d have done well” (2012: 207). They maintain that no reference has taken place. Their reasoning is presumably the following: If the student had done a history exam, he would not have done *that* exam (events such as exams have rich essences). Nevertheless, the counterfactual has to come out true—so we better suppose that not a singular but a general proposition was expressed. Since I have been trying to avoid discussing issues related to modality and essences, I will stop here, but Hawthorne and Manley’s diagnosis of this example is not compulsory.

⁴⁰ Kripke 1972: 79.

speakers *refer* to the inventor of the zip when they use the name 'Julius.' What is the right answer? Do such speakers refer, or do they not? In my opinion, the more important question is: Is this not a futile debate, where one runs the risk of drawing an arbitrary distinction among cases where there is no real difference?

What if, for instance, we introduce 'Julius' as: the creator of *this particular* zip? Or as any of the following: the owner of this zip, the last person to touch this zip, the person whose hand is now holding this zip...? It appears that it is fairly easy to think of a spectrum of cases where names are introduced 'descriptively'—yet where we are unwilling to draw a strict distinction between referential and non-referential names. The best way to look at these things, in my opinion, is to regard referring as a contextual affair.

Some philosophers of language insist that 'descriptive names' do not allow its users to refer, or to entertain *de re* thoughts. Kent Bach, for instance, says about the matter, "Using such sentences [does] not enable one to have singular thoughts about [the individual] and thereby be in a position to refer to him" (2004: 212). This strikes me as problematic. With respect to most of the names we use, we do not know *how* they were introduced. Perhaps they were introduced by ostension, perhaps they were introduced by description. The nature of a speech act should not depend on how exactly the expressions that are used once entered the language. The nature of a speech act should depend on mostly *local* matters—for instance, on the psychological state of the

speaker and audience, and the relevant nature of the informational transaction.⁴¹

At this point, the reader may wonder, “If both demonstratives and names have descriptive uses, what about pure indexicals like ‘I,’ ‘here’ and ‘now’? Can speakers utter these expressions and produce descriptive utterances?” In the next chapter, I will argue that there are no descriptive uses of indexicals. In the literature, some people have provided examples of purported descriptive uses of indexicals. I will consider these examples and argue that indexicals differ in important ways from demonstratives and names, due to the special nature of indexical belief. As a result of this, they have no descriptive uses.

⁴¹ There are even philosophers who have argued that *ostensively*-introduced names of people have uses where speakers do not refer or express *de re* beliefs. Kent Bach discusses a case where two men stumble upon a briefcase that contains a lot of money. They proceed to steal the money from the briefcase, which is labeled ‘Cassius King’. They then say, “Cassius King won’t be happy, but at least he’ll have his briefcase.” According to Bach, no singular proposition is expressed; no reference has taken place (2002: 87). A.P. Martinich considers a similar case. Suppose a drawing has just taken place for the grand prize in a raffle. The chairman of the raffle committee announces the winner by saying, “Jane Smith has won the grand prize.” According to Martinich, no reference has taken place in such a speech act (Martinich 1977: 161). As suggested in the main text, it is better to regard this as a context-sensitive affair.

CHAPTER 3

INDEXICALS, DEMONSTRATIVES, AND NAMES

3.1 Taxonomical Issues

The purpose of this chapter is to apply the descriptivism formulated in Chapter 1 to three types of linguistic expressions: indexicals, demonstratives and names. Some questions that I address in this chapter are: What are the meanings of indexicals, demonstratives and names? And to what extent are the processes that determine the meanings of indexicals, demonstratives and names semantic and to what extent are they pragmatic? In order to answer this question, we will have to consider certain taxonomical questions about where semantics begins and ends.

After discussing indexicals and demonstratives, I will spend some time on names. When it comes to names, it will turn out that understanding deferential name use is important in understanding language use in general. There is even a sense, I claim below, in which deferential name use allows people to entertain thoughts they would not be able to entertain without a language. Deferential name use expands the range of thoughts that people are able to entertain. This is one main difference between names, on the one hand, and demonstratives and indexicals on the other. Demonstratives and indexicals do not so expand one's range of thoughts.

Before I begin, a few remarks on terminology. I will use 'indexical' to mean: pure indexical. Any expression that is not 'I,' 'here' or 'now' will not count as an indexical.

(The expressions ‘you,’ ‘there’ and ‘then’ do not count as indexicals.) When I use ‘demonstrative,’ I will in the first instance be concerned with *bare* demonstratives, expressions like ‘this’ and ‘that.’ However, many of the things I say about these demonstratives also apply to complex demonstratives, personal pronouns, and related expressions.

Let me now address the semantics/pragmatics distinction and explain how I will apply this distinction in the sections up ahead. I propose to count as *semantic* those features of linguistic expressions that are *language-individuating*, and as non-semantic those features that are not. For instance, that the word ‘snow’ has the meaning that it has is something that makes English the language that it is. The meaning of an expression is, obviously, a semantic feature of that expression. But the word ‘snow’ has also certain non-semantic properties that do not depend in any way on English being the language that it is.

In addition to individuating languages, the semantic characteristics of linguistic expressions are related to *linguistic competency*. Speakers of a language possess certain knowledge in virtue of which they count as competent users of that language. In theorizing about meaning and semantics, it is important to keep this in mind. Whatever a theory ends up saying about ‘meaning’, it should be closely related to speakers’ knowledge that makes them competent users of the language and in virtue of which they understand the utterances of others.

Below, I will argue that among the non-semantic properties of linguistic expressions, there are properties that are nevertheless relevant to communication.

Some of these are *pragmatic* properties of expression tokens. When people use English to communicate something to an audience, then not everything that contributes to what is communicated depends on English being the language that it is. Not everything that contributes to what is communicated is semantic; this results from our definition of 'semantic' as language-individuating.

At this point, the reader may think, "Sure, I've heard about Gricean⁴² particularized (and generalized) pragmatic implicatures. Such implicatures make a difference to what is communicated, yet they are not semantic." Grice's classic example involves a professor who, in a letter of recommendation, writes 'The student is very punctual and has excellent handwriting.' The professor says and means one thing (viz. that the student is punctual), and in doing so, means *another* thing (viz. that the student's abilities are lacking). Gricean conversational implicatures must be derivable from apparent violations of conversational maxims, and furthermore are cancelable.

As will become apparent below, the view favored here is stronger than that. Not only are pragmatic processes relevant to what speakers conversationally implicate in addition to what their sentences mean, these processes can make truth-functional contributions to *what their sentences mean*. Below, for instance, I will sketch a descriptive view of demonstratives. Tokens of demonstratives, on this view, have descriptive meanings. However, this is not a *semantic* association, but a pragmatic one. When audiences interpret demonstrative utterances, they rely on knowledge of how

⁴² Grice 1967.

speakers reveal their referential intentions. This is knowledge of communication, not knowledge of English.

Names differ in two ways from demonstratives, on the account favored here. First off, names have context-insensitive meanings whereas demonstratives have context-sensitive meanings. Secondly, names have semantically determined meanings, whereas demonstratives have pragmatically determined meanings. To return to our example, that 'snow' means what it means (viz., frozen precipitation that falls from clouds) is part of what makes English the language that it is. But that Tom's token of 'that' means what it means (for instance, the expensive car that he and his passenger just drove past) is not part of what makes English the language that it is.

Speakers know the meanings of names merely in virtue of speaking the language. (This claim will be qualified somewhat in Section 3.4). Demonstratives are different. Speakers do not know the meanings of demonstrative tokens merely in virtue of speaking the language. Instead, they know their meanings partly because they speak the language and partly because they are well-versed in *communication more generally*. To take an example, consider the act of pointing. The fact that speakers reveal their referential intentions by pointing is not a fact about English. Instead, it is a fact about human communication in general.

What about pure indexicals? The general approach of this dissertation—to understand meaning in terms of thought—requires a somewhat non-standard approach to indexicals like 'I,' 'here,' and 'now.' The view defended here is that they have context-insensitive meanings. I will argue that pure indexicals contribute propositional

constituents that can be understood in a way similar to *variables*. The meanings of sentences that speakers express using indexicals are best represented by open sentences that are true *at* individuals. This approach to indexical meaning flows out of the view of *de se* belief that was discussed in Section 1.2.

What about *reference*? To what extent are referential properties of expressions semantically determined? On my view, *not at all*. On the view favored here, semantics fixes the meaning of certain expressions (either completely, as in the case of names; or partially, as in the case of demonstratives), but it never wholly determines the reference. As argued earlier, referential properties of expressions are accidental and non-semantic characteristics of those expressions. This is a consequence of understanding meaning in terms of thought and reference in terms of *de re* thought.

That a context-insensitive sentence in a language expresses a thought (i.e. has a certain meaning) is an essential feature of that sentence (qua sentence of that language, not qua inscription or sound). As I argued earlier, that a thought is *de re* with respect to some entity is a contingent, non-psychological feature of that thought. So that a sentence is '*de re*' with respect to some entity (i.e. contains an expression that refers to it) is a contingent, non-semantic feature of that sentence. On our picture the language-thought relation belongs to the semantic realm, while the language-world relation belongs to the extra-semantic realm.

(Context-sensitivity complicates this picture a bit. For context-*insensitive* expressions, the meaning of a sentence is the thought it expresses, and the meaning of a word is the concept it expresses. 'Water' means the same on Earth and on Twin-Earth.

That 'water' has the meaning that it has is essential to it, but that it has the referent that it has is accidental. In order to apply this view to context-sensitive expressions, we have to relativize to conversational contexts. The meaning of the demonstrative 'that,' in a certain conversational context, is the concept it expresses. That 'that'-in-a-certain-conversational-context has the meaning that it has is essential to the communicative act, but that 'that'-in-a-certain-conversational-context has the referent that it has is accidental.)

To make these claims more concrete, let me quote Kaplan from his article "Afterthoughts." He writes that it is a semantic fact about Pig-Latin that 'Ohsnay' means *snow* (1989: 573). This statement can be interpreted in a number of ways. First off, the statement could mean that it is a semantic fact that 'Ohsnay' *refers* to snow. On this reading, we disagree with Kaplan. That 'Ohsnay' refers to snow is a referential fact that depends partly on the contingent, non-linguistic world. It is not merely a semantic fact that follows from Pig-Latin being the language that it is.

Secondly, it could mean that it is a semantic fact that 'Ohsnay' means the same as 'snow'. On that reading, we agree with Kaplan. Or rather, that 'Ohsnay' means the same as 'snow' depends on two semantic facts: one about Pig-Latin and one about English. The statement that 'Ohsnay' means the same as 'snow' says that a certain expression in Pig-Latin has the same meaning as a certain expression in English. This meaning is provided by the beliefs we have about snow (*viz.*, that it is the precipitation that falls from clouds as ice particles).

At this point, the reader may object. Is it not patently implausible to hold that a

statement like ‘Snow’ refers to snow’ is non-semantic or contingent? Does it not look *a priori*? The answer to that question depends upon how *substantive* knowledge that ‘R’ refers to R really is. It is part of the meaning of ‘reference’ that it allows disquotational statements of the type: ‘R’ refers to R. If knowledge that ‘R’ refers to R is simply knowledge that it is an instance of this schema, then it is semantic knowledge.⁴³ But there is also a more substantive understanding. For instance, speakers of English in 1750 did not know that ‘water’ refers to H₂O. It was a substantial discovery to find out that the transparent liquid that fills the lakes and falls out of the sky as rain consists of H₂O.⁴⁴ On this understanding, the statement that ‘R’ refers to R partly depends on non-semantic, contingent facts.

3.2 Indexicals

Let me start with perhaps the trickiest category of expression, pure indexicals. Kaplan writes about pure indexicals, “The linguistic rules which govern *their* use fully determine the referent for each context. No supplementary actions or intentions are needed. The speaker refers to himself when he uses ‘I’, and no pointing to another or believing that he is another or intending to refer to another can defeat this reference” (1977: 491; *italics original*). It is a fact about English that the expressions ‘I’, ‘here’ and ‘now,’ are

⁴³ Some authors argue that our understanding the concept of reference consists in nothing but our inclination to accept such instances. See Horwich 1998.

⁴⁴ See below for more discussion of Putnam’s Twin-Earth argument.

used to refer to the utterer, the location, or the time of the utterance. (Remember that, for our purposes, no other expression counts as indexical besides those three.)

Perry 1997 calls expressions of this type ‘automatic indexicals.’ There are quick and dirty rules that allow the audience to figure out what the utterers of indexicals express. The audience does not need to rely on non-linguistic background knowledge, or entertain competing hypotheses about the speakers’ communicative intentions. Instead, all they have to do is apply a simple rule that every competent speaker of the language knows.

Kaplan identifies the meaning of a *token* of ‘I’ with its referent (i.e. the speaker). On this approach, indexicals are so-called ‘directly referential’ expressions. Indexicals are associated with a *character* (a function) that takes a certain ‘contextual coordinate’—the producer, the location, or the time of the utterance—and contributes this to the singular propositions that is meant. On the approach favored in this dissertation, the meanings of expressions depend on the thoughts that speakers express with these expressions. The question now arises: Can Kaplan’s view of indexicals be reconciled with a descriptivist approach that tries to understand meaning in terms of thought?

The answer is ‘No.’ At the end of Section 1.2, I considered the functionalist approach to indexical belief. On the Lewisian view, the contents of indexical beliefs are represented by properties. We looked at Perry’s example of Heimson, a crazy man who believes himself to be Hume. When Hume and Heimson entertain a thought they would express with ‘I wrote the *Treatise*,’ their thoughts share their psychological content. On

the Lewisian approach, they self-attribute the same property (viz. the property of having written the *Treatise*). Evaluated at Hume, this content is true; evaluated at Heimson, this content is false.

We can now see that if we want meaning to be closely related to the thoughts speakers express, we cannot go along with Kaplan. The issue with Kaplan's view is *not* that certain meanings are individuated too coarsely. Instead, these meanings are individuated too *finely*. On Kaplan's approach, what Hume means when he says 'I wrote the *Treatise*' is different from what Heimson means when he produces the exact same utterance. But we want the two meanings to be the same. They both say of themselves that they wrote the *Treatise*. Does this mean we should adopt Lewis's approach?

Yes. However, there is one small wrinkle. We do not want tokens of 'I' to contribute the utterer to the proposition expressed, but we do want tokens of 'I' to have *some* meaning or other. Suppose we straight-forwardly applied the Lewisian account of indexical belief to the view of meaning favored here. Then an utterance like, 'I am hungry,' would have as its meaning: the property of being hungry. But then the sentence 'I am hungry' would have the same meaning as the predicate 'hungry.' This is a minor technical wrinkle that can be resolved fairly easily.

We can simply stipulate that all tokens of 'I' express the propositional constituent '[SELF]'. This constituent '[SELF]' can combine with properties to form propositions that are true just in case the utterer of the sentence in question exemplifies the property. It performs the functions we want the meaning of 'I' to perform: A sentence of the form 'I am F' is true just in case the speaker of the sentence

is F, and if two different people utter 'I am F,' their sentences have the same meaning. On this approach, there is a sense in which 'I' is a context-*insensitive* expression.

To see this view in action, let us consider the position of the audience. How do audiences decipher utterances that involve the indexical 'I'? Perry considers a case where the audience does not know who said 'I'. He imagines receiving from Ellsworth a postcard from Hawaii that says, 'I am having a good time now'.⁴⁵ The postcard got wet and the name of the sender is illegible. Perry writes, "I do not know which proposition it expresses" (1988: 197). I want to take issue with this conclusion of Perry. Why does the audience not know which proposition it expresses?

Perry says, "Ellsworth did not plan for me simply to be able to think something like "How nice, the writer of this postcard was having a good time when he wrote it," but something like, "That old so-and-so Ellsworth is having a good time in Hawaii"" (1990: 198). But *what is it* to think that old Ellsworth is having a good time? Is it to update one's internal 'mental file' about the man named 'Ellsworth'? But surely, the meaning of Ellsworth's token of 'I' has nothing to do with the *name* 'Ellsworth.' Audiences can understand utterances involving 'I' perfectly fine even in cases where they do not know the name of the speaker.

Of course, utterers of 'I' usually *expect* their audience to form certain thoughts. Ellsworth may expect his audience to entertain the thought that the bearer of the name 'Ellsworth' is having a good time, and the thought that the utterer of the sentence is

⁴⁵ Perry 1988:197.

having a good time. But the thoughts that Ellsworth expects his audience to form are not the same as the thoughts that he communicates. On the view proposed here, the meaning of 'I' is the propositional constituent [SELF] that works as laid out two paragraphs ago.

Another way to see why the utterance 'I' does not need a meaning that is richer than [SELF] is by noticing that one cannot construct Frege cases with 'I.' This points to the fundamental difference between indexicals, on the one hand, and demonstratives and names, on the other hand. A subject who attributes to himself a property by using an indexical (for instance, by saying 'I am F') cannot rationally attribute to himself the explicit negation of that property by using an indexical ('I am not F). With demonstratives, this can happen. A person can say 'This is F' and 'This is not F,' and be making contradictory predications. With names, it can happen as well.

Audiences who hear a speaker use 'I' know—merely in virtue of speaking the language in question—the cognitive contribution of that term to the thought expressed in the utterance. But audiences who hear a speaker use a demonstrative do not know the thought expressed merely in virtue of being competent speakers. For instance, an audience who hears someone say 'That is F,' but who is not hooked into the conversational context in the right way, will usually not know the exact thought expressed. Yet, the audience does not lack semantic knowledge. This shows that the meanings of demonstrative expressions are not purely semantically determined.

Consider, for instance, the following variation on Perry's case. Suppose Ellsworth's postcard did not get wet. The name of the sender is perfectly legible.

Furthermore, suppose Ellsworth wrote on his card, 'That is beautiful.' The first word on the card is a discourse-initial occurrence of 'that.' In a case like this, the recipient will have no clue what is meant or referred to. Not only did the communication not go as planned, it utterly failed. There is no way for the audience to recover what Ellsworth meant, what he intended to communicate or refer to.

It is a non-starter to suggest that Ellsworth communicated to the audience something along the lines of: *the object of my attention* is beautiful. This will not work. We should keep in mind that one of the goals of communication is to get our thoughts across to audiences. These thoughts cannot have as their content: the object of this thought is such-and-such. This would result in an interpretative circle. The audience would not be able to figure out what information the speaker is trying to communicate. As we will see, this points to a fundamental difference between indexicals and demonstratives.⁴⁶

Note that I am not suggesting that all uses of demonstratives (or names, for that matter) have non-indexical meanings, that their meanings can be captured with non-indexical descriptive sentences. Rather, saying that these expressions have descriptive meanings is to claim that their meanings are best captured by descriptive sentences—

⁴⁶ Someone could suggest that what is meant by utterances involving pure indexicals *also* depends on speaker intentions. For instance, do speaker intentions not determine the duration of the timeframe that counts as *current*? And do they not determine the size of the location that counts as *here*? It strikes me that this is a fundamentally different type of context-sensitivity. One can teach someone the meaning of 'here' and 'now' by saying in a meta-language, "The term 'here' is used to speak of the location of the utterance." *Insofar as the meta-language is determinate*, this semantic rule results in determinate truth-conditions. Speaker intentions play no role. Things are different with demonstratives. One can teach someone the meaning of 'that' by saying in the meta-language, "The term 'that' is used to speak of the most contextually salient object." Even with a determinate meta-language, speaker intentions must do a lot of work in creating determinate truth-conditions.

where these sentences usually are indexical as well (for instance, by containing propositional constituents [SELF]). For example, when Ralph says, 'He is a spy,' he may be meaning something along the lines of, 'The man / saw at the beach behaving suspiciously is a spy.' More on demonstratives in the next section.

Let me now address the question whether there are descriptive uses of indexicals. If there are, our account that analyzes the meaning of 'I' as [SELF] might be in trouble. Geoffrey Nunberg has provided examples of utterances that supposedly involve descriptive uses of pure indexicals (Nunberg 2004). His examples include the following: 1) Today is always the biggest party of the year (uttered on the fourth of July, for instance), 2) I am traditionally allowed to order what I want for my last meal (uttered by a prisoner). While Nunberg does not provide examples involving 'here', it is not too hard to think of an example, for instance: 3) Here is usually where the bathroom is (uttered by a real estate agent).

The utterers of sentences 1) – 3) do not use indexicals to merely refer to the day of the utterance, the producer of the utterance, or the location of the utterance. Now, the question that arises is this: Do they not refer at all, or do they perhaps refer to other entities that are different from, yet *related to* the day, the producer, or the location of the utterance? Nunberg claims that these speakers still refer, but to *roles* instead of to what pure indexicals normally are used to talk about.

However this issue is to be resolved, I want to suggest that sentences 1) – 3) contain *non-literal* uses of pure indexicals. In a sense, speakers of 1) – 3) are using indexicals not in line with the linguistic conventions that govern them. So a semantic

theory that does not accommodate utterances of 1) – 3) is not thereby threatened as a result. These special uses of indexicals are fairly non-standard. (More non-standard, it seems to me, than the descriptive uses of demonstratives and names that we mentioned in Section 2.3 and will again discuss below.) So our response to these supposedly descriptive uses of indexicals cannot be accused of being *ad hoc*.

3.3 Demonstratives

The strategy so far has been to understand meaning in terms of thought. How does this strategy apply to demonstratives? In this section, I will argue for the following theses. First, the meanings of demonstratives are given by certain referential intentions of speakers. These meanings determine, relative to contexts, what the demonstratives in question refer to. Secondly, that demonstratives have these meanings is *not* merely a matter of their semantics. Instead, it is partly semantic and partly pragmatic. It is partly determined by semantic factors that pertain to the language in question, and partly by pragmatic factors that concern communication more generally.

For example, consider a conversational situation where a speaker utters a sentence containing a demonstrative while simultaneously pointing. Exactly what knowledge does the audience rely on in grasping the speaker's meaning? Two sorts of knowledge are involved: first, semantic knowledge about the expression type in question, and second, pragmatic knowledge concerning how people reveal their referential intentions to their communicative partners. The latter is not knowledge of

English. As suggested above, knowing that pointing is a way for speakers to direct their audience's attention is non-semantic, even though it concerns communication.

Such an approach to demonstratives is not new. Many authors have claimed that speaker intentions determine the referents of demonstratives. Some of these authors also hold that this is not a completely semantic process. Kent Bach, for instance, is an author who has long argued for this position (Bach 1987: 178-9). However, my account differs in that I maintain that demonstrative utterances express descriptive propositions that capture the cognitive significance of the thought expressed. Furthermore, I present a number of new arguments for the view that demonstratives have descriptive, non-semantic meanings.

Let us start with Kaplan's "Demonstratives." Here, Kaplan considers (but does not endorse) a directly referential account of demonstratives. He writes, "just as we can speak of agent, time, place and possible world history as features of a context, we may also speak of... demonstratum[s]... as features of a context" (1977: 528). On the basis of these remarks, we can imagine a view where demonstratives are semantically associated with a function that takes a contextual coordinate (viz. a demonstrated object) and contributes this entity entities to the singular proposition that the utterance expresses.⁴⁷

As already indicated above, the account favored here diverges from this directly

⁴⁷ This is actually not Kaplan's view in "Demonstratives." He maintains what he calls a 'neo-Fregean' view where complete tokens of 'this' and 'that' consist of these words *plus* their associated demonstrations (1977: 527). Such expression-plus-demonstration-combinations have characters that semantically determine referents relative to conversational contexts.

referential account of demonstratives in *two* ways. First, on the account favored here demonstrative utterances do not express singular propositions, but instead descriptive propositions. Second, that demonstrative utterances have the meanings that they have is not simply a semantic affair. It is partly governed by pragmatic, non-semantic processes. These processes concern communication, but they are not semantic in that they hold in virtue of English being the language that it is.

As a source of support, let me first introduce a case from an article by Brian Loar. “Suppose that Smith and Jones are unaware that the man being interviewed on television is someone they see on the train every morning and about whom, in that latter role, they have just been talking. Smith says, ‘He is a stockbroker,’ intending to refer to the man on television; Jones takes Smith to be referring to the man on the train. Now Jones, as it happens, has correctly identified Smith’s referent, since the man on television is the man on the train; but he has failed to understand Smith’s utterance” (1976: 357). Our account can explain the failure of communication. The descriptive proposition that Jones understood was not the one Smith meant.

What determines the reference of demonstratives? In the previous section, we saw that indexicals refer without any mediation of speaker intentions. A person who utters ‘I’ refers to himself, regardless of what he thinks or wants or intends. Do demonstratives work similarly? Are they, in Perry’s term, ‘automatic indexicals’ in that they are associated with semantic mechanisms that determine a referent without the mediation of speakers’ intentional states? In the literature on demonstratives, there are two opposing views on this issue. They are called ‘contextualism’ and ‘intentionism’. We

join the side of the intentionists.

According to the contextualist, demonstratives are associated with semantic functions from non-intentional features of the context to their referents. Early David Kaplan and Marga Reimer are contextualists.⁴⁸ According to the intentionist, on the other hand, speaker intentions play a large part in determining what speakers using bare demonstratives refer to. Later Kaplan (from 1989 on), Kent Bach, and Susanna Siegel are intentionists.⁴⁹ The later Kaplan writes, “I am now inclined to regard the directing intention, at least in the case of perceptual demonstratives, as criterial, and to regard the demonstration as a mere *externalization* of this inner intention” (1989: 582).⁵⁰

Early Kaplan provides an argument that allegedly shows that speaker intentions do not determine the referents of demonstratives. In the example, Kaplan points (without looking) to the wall behind him and says, ‘That is a picture of one of the greatest philosophers of the twentieth century.’ Unbeknownst to Kaplan, his picture of Carnap is replaced with a picture of Spiro Agnew. Kaplan intended to demonstrate Carnap’s picture, but in fact pointed at Agnew’s picture. This means that Kaplan’s intention does not determine the referent of the demonstrative, because if it did, the

⁴⁸ Kaplan 1977, Reimer 1991.

⁴⁹ Kaplan 1989, Bach 1987, Siegel 2002.

⁵⁰ Kent Bach makes a similar point. He writes about a Kaplan-inspired view that treats demonstratives on the model of indexicals, “Since the meaning of an indexical is supposed to specify the contextual parameters that constrain the referential intentions with which the indexical can be used (literally), the theory cannot allow the speaker’s intention to count as a separate parameter of context. If it did allow this, it could not explain how that intention is constrained by the meaning” (1987: 178).

referent would be Carnap's picture. Or so the argument goes.⁵¹

But as many intentionists have pointed out, Kaplan overlooks an important intention, viz. the intention to demonstrate the picture *behind him*. The two intentions conflict: the intention to demonstrate Carnap's picture and the intention to demonstrate the picture behind him. Which intention is more important? Kaplan's intention to talk about the object behind him *outweighs* the intention to talk about Carnap's picture. Among the plethora of beliefs about our environment, few are more central than those about how we are causally and perceptually related to things in our environment.

To see that the intention to refer to the picture behind him outweighs the intention to refer to Carnap's picture, we can imagine Kaplan discovering that Carnap's picture is not there. He will judge his previous utterance to be *false*, because the thing he demonstrated (the picture behind him) does not instantiate the property he attributed to it (being of the greatest philosopher of the twentieth century). (The reason it is false is *not* that there is no object that satisfies both 'being a picture behind him' and 'being Carnap's picture. If Kaplan had said—with the very same intentions—'That is a picture of a man,' he would have said something true.)

It appears, then, that communication with demonstratives works as follows. A speaker uses a demonstrative. Both speaker and audience know that demonstratives are used to speak of contextually salient entities. This is semantic knowledge that

⁵¹ Kaplan 1978.

pertains to an expression *type*. The audience will then search for the relevant contextually salient entity. In doing so, the audience will attribute referential intentions to the speaker. This may be based on gestures, but it also may be based on what reasonably can be supposed to be *relevant*. And in the search of what is relevant, the audience uses all kinds of knowledge that is not about language at all.

This supports the view that demonstratives have *pragmatic* meanings. That is, the speaker intentions that provide the meaning and determine the reference of demonstratives are not purely semantically, but also partially pragmatically determined. Speakers—in getting their audience to figure out the meaning of a demonstrative expression—do not purely rely on the audience’s linguistic knowledge. Audiences use practical knowledge about how speakers reveal their referential intentions—but this is not knowledge that pertains to a specific language. It is knowledge of communication, but not of semantics.

Some readers may object to this use of ‘semantic’. Many philosophers of language are inclined to classify every factor that affects the truth-conditions of (or proposition expressed by) an utterance as ‘semantic.’ For instance, Stanley and Szabo write, “As we use the term ‘semantic’, it includes any contextual contribution to the proposition expressed by an occurrence of a grammatical sentence, whether traceable to a constituent in that is uttered or not” (2000: 230). On this view, lots of factors that determine what speakers mean when they use language count as semantic.

There is also a competing view, sometimes called ‘minimalism,’ according to

which semantic context-dependence is much more limited.⁵² It is restricted to such things as resolving lexical ambiguity, assigning referents to pure indexicals, and recovering sentence ellipsis.⁵³ On this approach, effects on the truth-conditions of a sentence that do not belong to these categories are pragmatic. As should be apparent, I side with the minimalist approach to semantics. This is not the place to go into a lot of detail about minimalism vs. the competition, but let me briefly mention a few sources of support.

Firstly, semantic knowledge is just that, semantic knowledge. During communicative exchanges, audiences very often rely on knowledge that does not directly involve language. In order to maximize the effort-to-result ration, communicators often take shortcuts by relying on things that are mutually known between speaker and hearer. Practically all of our knowledge of the *Lebenswelt* can be utilized in communicative exchanges. But, clearly, semantic knowledge is not knowledge of everything. Since we have to draw the line somewhere, it is best to make a fairly strict distinction between the pragmatic and the semantic. This is one of the attractions of minimalism.

A related reason to favor the minimalist approach has to do with the amount of *systematicity* that can be expected in theorizing about these matters. About the speaker

⁵² Minimalist approaches are defended in Emma Borg 2004 and Cappelen & Lepore 2005.

⁵³ One does need pragmatic knowledge to figure out the correct interpretation of ambiguous expressions like 'bat.' In cases like these, once the audience knows which of the two conventions governing 'bat' the speaker relies upon, semantics does the rest. This is very different from the way pragmatics is involved in interpreting demonstratives.

intentions that determine demonstrative reference, Susanna Siegel writes, “the challenge is to give a systematic theory that identifies that intention as the one that fixes reference” (2002: 7). But how much systematicity can be found here? The process that hearers engage in when they infer speaker intentions is a very delicate process sensitive to a large body of information. A slight change in context may shift the reasonableness from one hypothesis about speaker intentions to another; there are no quick and dirty rules.

Now, some readers may object to the foregoing and insist that any truth-conditional contribution to the meaning of a sentence *must* be semantic. Kaplan, for instance, writes, “[*T*]ruth is a property of contents, and one wouldn’t want to be caught advocating a pragmatic theory of truth” (1989: 575). Kaplan’s worry is not founded, though. Many philosophers of language acknowledge non-semantic context-sensitivity that is relevant to what a *sentence* means (and so is different from what speakers merely conversationally implicate). The truth of sentences with this type of context-sensitivity depends in part on non-semantic factors.

Consider these examples. A speaker who utters ‘John is ready’ is usually understood as saying John is ready for *something*. A speaker who utters ‘John’s car is blue’ is usually understood as saying that the car that is *related* to John in a certain way is blue. A speaker who utters ‘Everything is for sale’ is usually understood as saying that there is a *certain domain* such that everything in it is for sale. A speaker who utters ‘I have not eaten’ is usually understood as saying that there is a *period of time* such that

they have not eaten during that period of time.⁵⁴ A minimalist will argue that these are examples of non-semantic truth-functional effects.

Let me mention an approach to demonstratives that is compatible with the view outlined here. Hawthorne and Manley 2012 formulate a *presuppositional* account of demonstratives. They write, “demonstratives are... existentially quantified expressions presupposed to be restricted in such a way that they have exactly one object/plurality in their extension” (2012: 218). On their account, demonstratives semantically trigger presuppositions. This can be regarded as an updated version of Kaplan’s notion of *character*, since these presuppositions attach to expressions types, not tokens. Hawthorne and Manley’s proposal is compatible with ours, since the *content* of these presuppositions in particular conversational contexts is pragmatically determined. On our view, these presuppositions would have descriptive contents.

Finally, I would like to briefly return to a worry already mentioned in Chapter 1. This worry is the following: If we spell out the meaning of demonstratives in terms of the beliefs of the speaker, will we not end up with a notion of meaning that is way too *egocentric* and *indeterminate*? How a speaker conceives of entities in the environment is a private matter, whereas meaning must be public. There is also the risk of ending up with meanings that are too indeterminate. Speakers entertain a plethora of thoughts about how they are related to entities in their environment. If it is not clear which of these beliefs is expressed, would that not make meanings indeterminate?

⁵⁴ These are common examples from the literature, see e.g. Borg 2004.

Again, I think these worries are not well founded. First off, demonstratives have pragmatic meanings. These meanings are public in the sense that competent speakers that are well-versed in communication can determine these meanings in different conversational contexts. But they are public in a different way than the way in which the meanings of names are public. The meanings of names (that one must know to count as a competent speaker of the language) are completely determined by the semantics of the public language in question. That does not apply to demonstratives.

Given our functionalist view from Chapter 1, in order for an audience to assign a meaning to a demonstrative, he needs to attribute a propositional attitude to the speaker. Audiences attribute such attitudes on the basis of these speakers' behavior and on the basis of how these speakers are embedded in the environment. As we have argued, in attributing such attitudes, speakers make use of a theory of communication, part of which consists of a theory of mind. On this picture, the meanings of demonstratives are exactly as public as these propositional attitudes. There is no hidden 'belief box,' with X beliefs in it. Instead, propositional attitudes (and, so, meanings) exist to the extent that they would be attributed by reasonable, competent interpreters.

3.4 Names

Descriptivism is best known as a theory of the meaning of names. On the one hand, names are easier to understand than indexicals and demonstratives because they are context-*insensitive*. On the other hand, they give rise to several new puzzles: How are

new names introduced into a language? How should we understand deferential name use? It will turn out that considering these questions can help us understand important issues in language use and cognition.

What exactly is covered by the category of ‘names’? Proper names of people and places, but also natural kind terms, common predicates, theoretical terms, and more. Predicates are from a different grammatical theory than names, but we can assimilate the two. Predicates can be turned into names by rewriting a sentence like ‘The tomato is red’ as: the tomato instantiates *redness*. The term ‘redness,’ here, can be regarded as a referring expression. (One can also construct a set-theoretic equivalent replacing ‘instantiates’ with ‘is a member of.’)

Briefly worth mentioning are names with multiple bearers. The name ‘Aristotle’ can be used to talk about the philosopher, but also about the shipping magnate who married Jackie-O. The term ‘bat’ can refer to a type of club used to play baseball, or to the nocturnal, rabies-carrying creature. This context-dependence is fundamentally different than the one covered in the previous section. Audiences use pragmatic knowledge when they encounter such terms. But this is a matter of disambiguation, of figuring out which of the two linguistic conventions that governs ‘Aristotle’ or ‘bat’ is invoked. Once they know which convention is in play, language does the rest.⁵⁵

In Section 1.3, I formulated a version of descriptivism on which meanings consist

⁵⁵ There are also *predicative* uses of proper names, as in ‘John is a Smith.’ See Burge 1973 for discussion. The sentence ‘John is a Smith’ appears to make the statement that John’s last name is ‘Smith’ —in other words, a meta-linguistic claim. These issues are relevant to the linguistic analysis of proper names; I will ignore them in what follows.

of thoughts. Thoughts, I argued, are best represented with modified Ramsey sentences. These are existentially quantified sentences that contain no names for environmental entities or kinds. The argument from that chapter concluded that meanings, as well, can be captured with such descriptive sentences. The meanings of sentences containing names, then, can be represented with those descriptive sentences that capture the thought that people use these names to express. As I noted, there is no guarantee that these meaning-providing sentences are from the same language as the name-containing sentences.

I emphasized that the meaning-providing beliefs about the referents of names may include beliefs about a number of different characteristics. Such beliefs may be *indexical*, but also include beliefs about the referent's *causal* features, and the extent to which the referent is a *natural* entity or kind. (The issue of naturalness will be further discussed in Chapter 4). The descriptivist maintains that when meanings are properly understood along these lines, Kripke and Putnam's supposed counterexamples are harmless.

This section is mostly concerned with Kripke's *semantic argument*. According to this argument, speakers can use a name to refer without associating it with uniquely identifying information. Descriptivism, it appears, requires speakers to be in cognitive possession of such uniquely identifying information. So, the argument concludes, descriptivism is false. Central to answering this argument is distinguishing *authoritative* from *deferential* name use. Ultimately, understanding the difference between authoritative and deferential name use will clarify the nature of names and be useful in

understanding language use and cognition.

An argument I will not discuss here is Kripke's *modal argument*. This well-known argument is designed to show that sentences containing proper names do not have the same *truth-conditions* as sentences where these names are replaced by descriptions that provide their meaning. Kripke's example involves the sentences "Aristotle was fond of dogs" and "The last great philosopher was fond of dogs." He writes, "My view is that 'Aristotle' in (1) is rigid, but 'the last great philosopher of antiquity' in (2) is not. [I]t is a doctrine about the truth conditions, with respect to counterfactual situations, of [...] *all* sentences, including *simple* sentences" (1972: 12). There are possible situations where Aristotle was fond of dogs, but where he was not a philosopher or even called 'Aristotle'. Here, the first sentence is true; the latter is not.⁵⁶

My reason for not discussing Kripke's modal argument and the notion of rigidity is that it requires careful study of *de re* modality. As suggested in the Introduction, if one characterizes or analyzes meaning in terms of modal notions, then one cannot in turn use meaning to understand modal notions—for that would be circular. Meaning and modality, in many people's eyes, are intimately related. As a theorist, one has to choose which of these two families of notions is the more fundamental one, and which one is the more derivative one.

There is a still-respectable tradition of understanding modality (both *de re* and

⁵⁶ Descriptivists have responded to this argument in a number of ways. Some invoke *wide scope* (Loar 1976), some invoke *rigidification* (Lewis 1997). Others have made comparisons with *referentially used* definite descriptions (Bach 2002).

de dicto) in terms of meaning. Earlier, I quoted Kaplan who spoke of the 'Golden Age of Pure Semantics.' This age, he says, culminated in Carnap's aptly-titled book *Meaning and Necessity*. Contemporary philosophers who are sympathetic to this general outlook might, for instance, use some form of counterpart theory to reduce *de re* modality to *de dicto* modality. And they might try to understand *de dicto* modality in terms of, for instance, analyticity and related notions. On such an approach, meaning would be more primitive than modality.

Such an approach to modality has several advantages. For one, it is not committed to any primitive modality. Also, it is in a good position to explain the context-sensitivity of certain types of modal statements. Kripke is well-known for being unsympathetic to counterpart theory. He suggests that we take our everyday modal judgments at face value. About our ordinary modal intuitions, he writes, "I really don't know, in a way, what more conclusive evidence one can have about anything, ultimately speaking" (1972: 42). But, of course, we can be very confident of an ordinary judgment that nevertheless employs derivative concepts. These two things are not mutually exclusive.

(What about the notion of *belief*? Am I not guilty of trying to understand meaning in terms of belief, and am I not thereby excluding approaches that explain belief in terms of meaning? Well, yes. When it comes to this debate, I clearly choose a side. The approach favored here conflicts with the views of Michael Dummett and Wilfrid Sellars who see the intentionality of language as prior to that of thought. But, I maintain, the explanation of meaning in terms of thought has the backing of common-

sense. In addition, there are powerful and attractive theories that employ this explanatory strategy.)

Now, on to names and language use. Uses of names can be separated into deferential uses and authoritative uses. Deferential uses of names are uses that Kripke describes as instances 'reference borrowing' (1972: 91-5). A speaker who uses a name deferentially piggy-backs on an existing reference relation. Kripke claims that certain meta-linguistic intentions are essential to the process of reference borrowing. He writes that such a borrower "must... intend when he learns it to use it with the same reference as the man from whom he heard it" (1972: 96).

By using names deferentially, we *expand* the range of things we are able to think and say. Schiffer, for instance, writes that "the thoughts expressed by sentences containing names like 'Thales' and 'Feynman'... are ones some of us would not have if we did not have the names. This fact is nicely accounted for on the hypothesis that my only knowledge of Feynman is under a description which mentions his name" (Schiffer 1978: 199). Kaplan, whose view on names differs from Schiffer's, makes a similar point when he writes: "On my view, our connection with a linguistic community in which names and other meaning-bearing elements are passed down to us enables us to entertain thoughts *through the language* that would not otherwise be accessible to us" (1989: 603; italics original).⁵⁷

⁵⁷ Notice the difference with indexicals and demonstratives. They seem to be expressions that do not expand the range of thoughts we are able to have. In fact, one could even argue that the thoughts speakers communicate using only indexicals and demonstratives are thoughts they could have had without any language whatsoever.

What does it take to be able to use a name deferentially? Very little. Using a name deferentially seems to require merely two things: knowledge that there is a convention of calling an entity or property by a certain name, and an intention to use that name in accordance with that convention. Scott Soames makes a similar point when he discusses the “information grasp of which explains speakers’ ability to understand [a name], and to be able to use it competently.” He writes that in this sense of ‘meaning,’ “different names have nearly the same meaning, no matter what they refer to” (2002: 56). Maybe this is also what John Stuart Mill means when he says that proper names are “non-connotative,” that they are mere tags.⁵⁸

Since deferential use of proper names requires so little cognitive investment, speakers constantly learn new ones (and forget old ones). Paul Ziff even claims that no proper name is such that familiarity with it is required in order to count as a speaker of the language (1960: 86-7). When applied to proper names of people and places, this is very plausible. I may not be familiar with the terms ‘Aristotle’ or ‘Paris,’ yet count as a speaker of English. However, below I argue that there are *other* context-insensitive referring expressions where one *does* have to know the meaning in order to count as a competent speaker of the language.

(It is odd that theories of proper names have been so influential in philosophical theorizing about meaning and reference over the last 50 years. If Ziff’s point is correct, and such names are not really part of the language after all, it would seem a bad idea to

⁵⁸ Mill 1949: 20.

take proper names of people as the paradigm case when analyzing the meaning and reference of linguistic expressions. It would seem methodologically suspect to let one's conclusions about proper names dictate what one thinks about meaning and reference in general.)

Now, not all uses of a name can be deferential. Searle writes, "reference to an individual may be parasitic on someone else's but this parasitism cannot be carried on indefinitely if there is to be any reference at all" (1969: 170). Kripke makes a similar point when he discusses his non-circularity condition (1972: 68-70). According to this condition, a full account of reference cannot claim that all relations of reference depend upon prior relations of reference. He writes, "[f]or any theory, the account must not be circular. [It cannot] involve the notion of reference in a way that it is ultimately impossible to eliminate" (1972: 68, 97). Devitt and Sterelny write, "this process [of reference dependence] cannot, however, go on forever: there must be some terms whose referential properties are not dependent upon others. Description theories... pass the referential buck. But the buck must stop somewhere" (1999: 60).⁵⁹

So in addition to deferential uses, there are *authoritative* uses of names. When speakers use a name authoritatively, they do not merely rely on existing referential conventions involving that name. Instead, their use partly constitutes the referential convention in question. To make things more precise, let me distinguish three uses of

⁵⁹ Some of these critical remarks can also be understood as criticisms of a version of descriptivism called 'global descriptivism.' See the next chapter for further discussion.

names: purely deferential, partly deferential/partly authoritative, and purely authoritative.

What is a purely authoritative use of a name? A *baptism*. Consider a person who points to a ship and says, 'This will from now on be called 'Titanic,'" or 'The perpetrator of these murders will from now on be called 'Jack the Ripper.'" The baptizer introduces a name for something he believes to exist. The meaning of the name is provided by these beliefs and the referent of the name is the entity or property in the world that these beliefs are true of. In Kripke's terminology, *fixing the referent* of a name is giving the meaning of that name.

Now, when a baptizer introduces a name for something he believes to exist, these beliefs may be perceptual, but they do not have to be. The person who named the Titanic probably saw the thing he named. The person who named Jack the Ripper probably did not see the person he named. Earlier, I mentioned Kripke's example of Leverrier who introduced the name 'Neptune' for the planet that caused perturbations in the orbit of Uranus.⁶⁰ Leverrier created a term for a certain entity that he thought was real and that he believed to have certain characteristics and established a convention.⁶¹ This is another example of introduction by description.

Let us return to name *use*, as opposed to name-introduction. Once a name is introduced, purely authoritative uses of that name are no longer possible. Uses of that

⁶⁰ Kripke 1972: 79.

⁶¹ Does it make sense to speak of a convention when only one person is involved? I do not think much rests whether this deserves to be called a convention or not. There might be some interesting connections with Wittgenstein's private language argument here.

name will be either purely deferential, or partly deferential/partly authoritative. Why are purely authoritative uses of that name no longer possible? Because once a convention has been established to call something by a certain name, subsequent users of that name do not count as using that name if they do not attempt to adhere to this convention. That is, all later users of a name will try to follow the convention (they may fail to do so, but they still will *try* to do so).

This means that many uses of names are *mixed*—partly deferential and partly authoritative. Speakers who use names in a mixed fashion do two things: 1) they intend to adhere to an existing convention whereby a certain entity or kind is called a certain name, and 2) their use of the name in question partly constitutive of that convention. A good example to discuss here is Evans’s example of ‘Madagaskar’.⁶² Initially, this term referred to a part of the African mainland (near the current Somalia). Later users turned it into a name for an island off the East-African coast.

How did this change come about? To ask this question is, in a sense, to ask how *use determines meaning*. Consider the name ‘Madagaskar’ when it was introduced with its initial meaning, before it turned into a name of the island. In order for a change to have occurred, there must have been uses that were not purely deferential. There must have been mixed uses (partly deferential, partly authoritative) that made a difference to the convention in question. Still, how can this occur?

In the case of ‘Madagaskar,’ speakers whose use of the name was partly

⁶² Evans 1973. Some Googling actually reveals that the history of the name ‘Madagaskar’ is not as simple as Evans suggests it to be, but let us ignore historical accuracy.

deferential/partly authoritative must have used the name *non-literally*. That is, they must have started to use the name not in line with the convention governing the expression at that time. One can see how this works. Someone can intend to use a name in line with the convention that governs it (and consider their use constitutive of that convention), yet be *mistaken* about whether they are conforming to the convention. Because speakers used 'Madagaskar' in this way, the meaning of the name changed.

Kripke gives an example of two people who see Smith in the distance and mistake him for Jones. One of them says, 'Jones is raking the leaves'.⁶³ Kripke considers this an example of 'speaker's reference.' On our account, this is a *non-literal* use of the name 'Jones.' The speaker is not conforming to the convention that governs this name. But suppose the conversational participants continue to use the name 'Jones' for Smith for some period of time. At some point, they will have established a convention to refer to Smith using the name 'Jones'. In effect, they will have given Smith a new, second name.⁶⁴

Now, proper names of people and places are such that many of our uses of them are deferential. But when it comes to other terms—e.g. color terms such as 'red' or natural kind terms such as 'water'—our uses are less deferential. That is, most speakers are in cognitive possession of the beliefs that give the meaning of the name, and that

⁶³ Kripke 1972: 25, 1977: 263.

⁶⁴ What about Kripke's example of 'Gödel'? (1972:83) Suppose that Gödel stole the incompleteness proof and published it. Gödel's neighbor did the actual proof. Kripke argues that speakers who believe that 'Gödel' names the prover of incompleteness still refer to Gödel when they use that name, and not to his neighbor. Our account can easily account for this: speakers' intentions to conform to naming practices outweigh their intentions to refer to persons they believe to have certain characteristics.

determine—depending on the contingent state of the non-linguistic environment—the referent of the term. Below, I go into further detail on how this process works.

But before moving onto authoritative name use, I want to consider these questions: Why are uses of proper names of people and places usually purely deferential? Why are uses of kind or property expressions, color terms such as ‘red,’ or natural kind terms such as ‘water,’ usually not purely deferential? In my opinion, this has partly to do with the nature of their referents. Proper names of people and places are names of *particulars*. Property names and natural kind terms are names of *universals* or *kinds*. Due to the different natures of the two, our practice of naming them differs.

Particulars change: they lose and gain both intrinsic and extrinsic properties over time. This means that one cannot track individuals over time if one singles them out qualitatively. Since we are interested in following individuals over time, it is useful to introduce labels for these individuals that are used deferentially by others speakers. During the baptism, a speaker ‘locks in’ the reference, and subsequent users rely on that process when they use the name.

Properties and natural kinds, on the other hand, *do not change*; they do not lose and gain intrinsic properties over time. (Perhaps they do not even lose and gain extrinsic properties over time.⁶⁵) We are also interesting in tracking properties, but there is no

⁶⁵ If say ‘perhaps,’ because whether this is true depends on what one means by ‘extrinsic property.’ We have to exclude ‘Cambridge’-properties, such as *being Tom’s favorite property*. We also have to ignore extrinsic properties of property *instances*, such as *being located in Paris*. But with the right restrictions put in place, it seems to me that one can maintain that properties do not gain or lose extrinsic properties over time. See Chapter 4 for more discussion.

need to use names for properties purely deferentially. We can simply refer to them in virtue of having beliefs about their intrinsic and extrinsic characteristics.

Of course, there are many property names that most speakers use merely deferentially, such as specialized lingo and technical jargon from various domains of discourse. But the reason that speakers use such terms deferentially is that it would require a significant *cognitive investment* to use them authoritatively. The reason is not that authoritative uses of such names would run the risk of losing track of the properties they refer to—for these properties have stable and unchanging characteristics, unlike individuals.

The current topic of deferential language can be connected to the issue of cognitive significance. The overall strategy of this dissertation has been to understand meaning in terms of thought. Thoughts, in turn, are understood in terms of the role they play in perception and action. So the meaning of linguistic expressions tells us something about their cognitive significance. But a speaker who utters a sentence deferentially entertains a different thought, of course, than a speaker who utters the exact same sentence authoritatively. In fact, there is a sense in which deferential language use allows speakers to carry information about the world that they would not be able to do otherwise. The quotes from Kaplan and Schiffer earlier from this section point this out.

At the same time, I have argued that there is a certain set of words the meaning of which people *have* to know in order to count as competent speakers. This set of words does not contain proper names of people and places, nor does it contain

specialized lingo and technical jargon. For this core-set of expressions, though, knowledge of meaning equates to knowledge of cognitive significance. With respect to this set of expressions, we can even agree with Frege's insight that the meanings of expressions allow us to transmit mankind's "common store of thoughts" (1948: 212).

CHAPTER 4

THE MODEL-THEORETIC ARGUMENT

4.1 The Argument

This chapter is devoted to Putnam's model-theoretic argument. This argument presents the most formidable obstacle to the view of meaning outlined so far. I touch upon the following questions: What exactly is Putnam's model-theoretic argument? Exactly which theories are susceptible to the argument? What type of externalist constraints upon reference have been invoked to solve the model-theoretic argument? Does Putnam's 'just more theory'-response to these answers work? Is it feasible to invoke internalist constraints to answer the model-theoretic challenge?

Here is a common reading of Putnam's model-theoretic argument (an early version of which was given by Newman).⁶⁶ Consider a theory T that is consistent and empirically confirmed. A corollary of Gödel's completeness theorem tells us that if T is consistent, it has a *model*. A model is an ordered pair $\langle W, V \rangle$, consisting of a domain and an interpretation function. The domain W is a set of elements that can be considered the domain of discourse of T , and the interpretation function V is a function that maps the terms of T onto sets of elements from W in such a way that (all, or many of) the sentences in T come out *true*.

⁶⁶ Putnam 1978, 1983. Newman 1928. There is some controversy as to the correct interpretation of Putnam's model-theoretic argument. My interpretation resembles that of Anderson 1993, Bays 2008, Lewis 1983a, 1984, and Merrill 1980.

A consistent theory T, then, can be considered true-in-the-model $\langle W, V \rangle$. Putting aside issues of size, the *real world* can be turned into a model of theory T as well. The domain W can be replaced with a set of elements from the real world and the interpretation function V can be replaced with a mapping onto elements from the real world. Consequently, T is true-in-the-real-world. In other words, it is true *simpliciter*. If we identify *reference* with this mapping from terms onto elements in the real world, Putnam's argument shows that every consistent theory T comes out *true*. This is a *reductio ad absurdum*, for it should be possible for such a theory to be false.

A different way to understand the model-theoretic argument is by considering the notion of *functional definition*. The Ramsey-Carnap-Lewis view of theoretical terms provides the default account.⁶⁷ Consider a theory T that contains terms t_1 through t_n . It also may contain already interpreted terms (O-terms). This theory $T(t_1, \dots, t_n)$ can be transformed into its *Ramsey sentence*: $\exists x_1 \dots \exists x_n T(x_1, \dots, x_n)$. This Ramsey sentence, in turn, can be transformed into an open sentence by removing the quantifiers: $T(x_1, \dots, x_n)$. Functional definition, then, is the view that the referent of t_1 is the first member of the unique n-tuple that *realizes*, or comes close enough to realizing, this open sentence. The meaning of t_1 is how the referent of t_1 depends on the state of the world.

Putnam's model-theoretic argument poses a problem for the view that all terms are so functionally defined. On this view—called 'global descriptivism'—there are no already-interpreted O-terms. Lewis writes, "global descriptivism... leads straight to

⁶⁷ Lewis 1970.

Putnam's incredible thesis. For *any* world (almost), whatever it is like, can satisfy *any* theory (almost), whatever it says. ... It is (almost) certain that the world will afford the makings of an interpretation that will make the theory come out true" (1984: 60). He continues, "Global descriptivism... may be part of the truth about reference, but it cannot be the whole story. There must be some additional constraint on reference; some constraint that might, if we are unlucky in our theorizing, eliminate *all* the allegedly intended interpretations that make the theory come true" (61).

Before I discuss proposed solutions, it is important to address the following question: Exactly for what view does Putnam's argument pose a problem? Have philosophers even seriously considered global descriptivism as a theory of the meaning of terms in a public language? In fact, does global descriptivism not commit the error of attempting to analyze linguistic meaning in terms of word-world relations? Is this not a type of error that ignores the Gricean insight that linguistic intentionality should be explained in terms of mental intentionality?

Lewis agrees. He writes, "It would be better, I think, to start with the attitudes and go on to language", but a proper restatement of Putnam's argument, "would *relocate*, rather than avoid, the problem" (1984: 57-8; italics added). Elsewhere, he writes, "If the problem of intentionality is *rightly posed* there will *still* be a threat of radical indeterminacy" (1983a: 49; italics added). So what would the model-theoretic argument as applied to mental content—a 'rightly posed', 'relocated' version—look like?

Some philosophers have defended a language of thought-hypothesis.⁶⁸

According to this hypothesis, human thinking takes place in a symbolic language that is realized in the brain. Thoughts, on this view, are composed of ‘sub-thought’ parts (i.e. concepts) in much the same way as sentences are composed of sub-sentential parts (i.e. words). In fact, Putnam himself considers an application of his argument to a language of thought. He writes, “‘Mentalese’ is thought to be a medium whereby the brain constructs an *internal representation* of the external world... If thinking is ultimately done in ‘mentalese’, then [my argument shows that] *no concept we have will have a determinate extension.*” (1983a: 17; italics original).

Does that mean the model-theoretic argument only poses a problem if we accept a language of thought? No, the argument can be viewed as being *both* about language and mind without having to adopt ‘Mentalese’—here is how.⁶⁹ On the functionalist approach sketched in Section 1.2, ‘belief’ and ‘desire’ are theoretical terms from folk-psychology. This theory consists of generalizations that capture regularities about how the environment affects subjects, how subjects’ internal states affect behavior, and how these states affect each other. A properly regimented, more rigorous version of this folk-theory will ascribe certain beliefs to subjects. Such ascriptions capture what subjects really believe and desire.

This regimented psychological theory will generate belief ascriptions on the basis

⁶⁸ Fodor 1975.

⁶⁹ My claims about the proper target of the model-theoretic argument is inspired by remarks from Michael McDermott. See his 1988. But what follows should not be considered his view.

of behavioral and environmental evidence. As suggested earlier, this psychology will ascribe to a subject a belief in a modified Ramsey sentence. It will have the form: S believes that $\exists!x_1 \dots \exists!x_n T(x_1, \dots, x_n)$. (The exclamation marks add a uniqueness condition.) Putnam's model-theoretic argument can be understood as saying that the world cannot fail to provide a model that makes the Ramsey sentence from this content clause come out true. And this is the *reductio* that has to be avoided. Ultimately, then, Putnam's model-theoretic argument does not pose a problem for theories of reference, but rather a challenge for theories of *intentionality*.⁷⁰

4.2 Structural Naturalness and Future Theorizing

⁷⁰ There are some interpretative difficulties in understanding Lewis's view on intentionality. On my reading of Lewis, interpretation is governed by *four* different principles. (i) First, there are principles relating environmental impact to subject's internal states. People should be interpreted as generally believing true things. In 1974, Lewis calls these 'principles of charity' (117); in 1983a he calls them 'principles of fit' (50); and in 1986, he considers them part of 'instrumental rationality' (39). (ii) Secondly, there are principles relating subjects' internal states to behavior. People should be interpreted as generally acting rationally. In 1974, Lewis calls these 'principles of rationalization' (117). In 1983/86, he considers these also a matter of 'fit' and 'instrumental rationality.' (iii) Interpretation is also governed by principles of consistency on how subjects' internal states interrelate. People generally do not endorse contradictions. Again, this is usually spelled out in terms of instrumental rationality. (iv) The fourth, and most controversial, principle of interpretation is concerned with objective naturalness. In 1983's "New Work...", he *also* calls such principles 'principles of charity' (52). (Notice the change in his use of that label.) In 1984's "Putnam's Paradox," he speaks of eligibility constraints concerning naturalness, and in 1986's *The Plurality of Worlds*, he calls them "principles of rationality more generally" (39).

Some commentators (Schwarz forthcoming) suggest that it is not clear whether Lewis held on to this part of his 1983-86 view of mental content. Perhaps Lewis dropped the naturalness constraint. Schwarz points out that in Lewis's articles on content from the 90s (1994's "Reduction of Mind" and 1997's "Naming the Colors"), he does not discuss naturalness whatsoever. This is accurate, however Lewis's 1997 introduction to his *Papers in Epistemology and Metaphysics* summarizes his earlier views without disavowing them. And in a 2000/2001 paper (posthumously published in 2004), Lewis used the term 'reference magnet.' So it seems to me that Lewis stuck with the view that incorporates naturalness as a constraint on interpretation.

Earlier, I considered the beliefs that provide the meaning of names. These beliefs concern various features of the believed-in entities or kinds, among them the degree to which they are *natural* entities or kinds. I now want to look closer at this naturalness constraint. How exactly does this contribute to meaning and fix reference? I will then relate this issue to the fact that our favored form of descriptivism is a future-oriented one. In Section 1.2, I argued that an interpretative assignment of content to internal states takes into account *future* subjects. These two issues turn out to be importantly related.

In order to get the discussion started, it will be useful to consider some examples: Putnam's 'jade' and Kripke's 'gold.'⁷¹ First, 'jade.' At some point, it was discovered that speakers had been applying 'jade' to two different substances, viz. jadeite and nephrite. Whenever this happens, speakers can make a *semantic decision*. One option is to continue to use the term in question for both things. This reveals that naturalness is not very central to the reference-determining beliefs. If this decision is made, there is no need to re-interpret old utterances (some as true, some as false). Instead, old utterances will still be considered true and speakers will continue talking as before. This seems to have been the choice when it comes to the term 'jade'.

The other option is to continue using the term in question for only *one* of the two substances that speakers used to call a certain name. If this decision is made, old utterances will be re-interpreted as referring to the newly discovered kind that is

⁷¹ Putnam 1975, Kripke 1972.

considered the best candidate for being the referent. Some will be true, some false. Speakers *could* have considered past utterances involving 'jade' that were made in the presence of jadeite as true, and ones that were made in the presence of nephrite as false. With terms where the associated naturalness requirements are non-negotiable, this seems to be the proper response. But 'jade' was not such a term.

Naturalness seems to be a matter of degree. Entities and kinds can exemplify it to a smaller or larger extent. The more fundamental a description of the world, the more natural (or 'elite') the extension of the terms used in the description. For instance, it is plausible to suggest that the extension of predicates in psychology are more natural than the extension of predicates in economics; the extension of predicates in physics are more natural than the extension of predicates in biology. Lewis even suggests that "the less elite [properties] are... connected to the most elite [properties] by chains of definability "(1984: 66).

Now, while few philosophers nowadays would endorse such claims of explicit definability, there are weaker theses that are similar in spirit and that can be used to shed light on naturalness or fundamentality. For instance, a global supervenience thesis. On such a thesis, the properties denoted by fundamental predicates are those properties such that: any world that is a minimal duplicate of the actual world in terms of those properties is a duplicate *simpliciter*. Or one could adopt an *a priori* entailment thesis. On such a thesis, a complete description of the world in terms of fundamental predicates (plus a 'that's all'-clause) entails *a priori* a complete description of the

world.⁷² Neither of these theses is committed to explicit definitions.

Consider another example, Kripke's 'gold.' We can imagine that in the past people were not able to tell apart real gold (the element *Au*) from fool's gold (iron pyrite). Suppose that these speakers defined the term 'gold' as something like 'the shiny, yellow metal that can be found in such-and-such places...' Imagine that these speakers would call samples of both *Au* and iron pyrite 'gold.' When they called bits of iron pyrite 'gold,' their utterances were false. Their term 'gold' did not have an extension that included iron pyrite. Does the descriptivist theory have the resources to agree with this intuition?

As suggested earlier, the meaning of the term 'gold' includes that it is a natural collection. The collection of all *Au* is a more natural, and therefore more eligible, reference candidate for the term than the collection of all *Au* and all iron pyrite taken together. But in light of our discussion from the previous paragraphs, what does it mean to say that something is a more natural extension? To ask this question is, in a way, to ask the question of how the term 'natural' gets *its* meaning. How does this work? The theory of descriptivism needs a satisfactory answer to this question.

One way to understand this is in terms of *future theorizing*. Terms whose part of the meaning is that they apply to natural extensions can be considered *deferential terms*. But instead of deferring to experts that exist at the same time, speakers delegate to future speakers in determining the meaning of a term. Our imaginary past speakers

⁷² Chalmers 2012.

who called both *Au* and iron pyrite 'gold,' spoke truly on some occasions, but falsely on other occasions. Exactly when they spoke truly and when they spoke falsely depends on the discoveries about the various shiny, yellow metals that have since then been made.

Says Lewis, "Descriptivism... may be *futuristic*. That is, the term-introducing theory which is supposed to come true on intended interpretations, if such there be, might be not the theory by which the terms actually were introduced, but rather some improved descendant" (1984: 69). Our current concept of gold is the descendant of the past concept of gold. 'Gold''s current meaning differs from its past meaning, because the surrounding theory has changed. Nevertheless, our 'gold' is a descendant because the changes were fairly continuous. In other cases, such changes were not continuous. There currently is no descendant of the term 'phlogiston' (the stuff that was supposed to exist in flammable things). Nor is there a descendant of the term 'Vulcan' (a supposed planet between Mercury and the Sun).

Terms where it is part of their meaning that they have natural extensions are special in a certain way. When speakers gain knowledge of the world, and discover they have been applying these terms to substances with different underlying structures, they will update the meanings of these terms. We are currently imagining that 'gold' to be such a term. There are also terms that work differently. Consider, for instance, a term like 'mousetrap.' It is not a part of the meaning of this word that its extension is a natural collection. Future discoveries about the world will not cause us to update the

meaning of this term.⁷³

In what follows, I will call this notion of naturalness ‘structural naturalness.’ In the next section, I will contrast it with a very different notion of naturalness. Note that structural naturalness is not ‘mind-dependent’ or ‘theory-dependent’ in any way. The way things are arranged in the world determines which extensions exemplify structural naturalness and which ones do not. The way things are arranged in the world—whether we are currently aware of it or not—determines which extensions are eligible to be the referents of our theoretical terms.

(Why would descriptivists view naturalness in this way? Is this understanding of naturalness compulsory? The answer is that descriptivists see *truth* as the central constraint on interpretation. If a theory T at time t_1 cannot be fully interpreted on the basis of merely the truth of T at t_1 , then a plausible suggestion is to make the truth of the *successor* of T at some later time t_2 relevant to the interpretation of the theory T at t_1 . This is what happens here. By employing this strategy, the guiding idea that truth is the central interpretative constraint is upheld. There are no *ad hoc* modifications to the theory and the overarching idea is maintained.)

The question now is the following: Can the structural notion of naturalness be used to solve Putnam’s model-theoretic puzzle? In my opinion, it cannot. This is what Putnam argues in his so-called ‘just more theory’-response. He writes that “the problem is that adding to our... language of science a body of theory called ‘Causal theory of

⁷³ Lewis prefers ‘a description theory of reference of a global and *futuristic* sort’ (1983a: 46; italics added). See also 1984: 69-70.

reference' *is* just adding more *theory*" (1977: 18; italics in original). The same holds for adding a body of theory called 'naturalist theory of reference.' Putnam's trick is to annex the proposed solution and add it to the total theory, which is then faced with the same problem as before.

We have understood naturalness constraints in terms of deference to future theorizing. Is this understanding useful in creating determinacy in interpretation? When Putnam provides his argument, he speaks of ideal theories, of *complete or finished* theories. Now, what does it mean to have a theory that is complete or finished? It cannot mean *true*, for that is exactly the issue under consideration. Putnam's central question is whether complete theories are guaranteed to be true or whether they can be false. Suppose then that we understand 'complete' to mean that no future theorizing exists (difficult as this is to imagine).

This means that there is no future theorizing that speakers can defer to. There are no further potential discoveries about the world that can narrow down the interpretation of the theory in question. Now, it is tricky to conceive of a situation where there is a complete theory for which the question of truth has yet to rise. In the next section, I will return to the question as to whether this is conceivable or possible. But for now, we can see the reasoning. Appealing to structural naturalness, as we have understood it, will not help with the interpretation of a complete theory.

To provide more support for the claim that structural naturalness is useless in

answering Putnam's challenge, let me discuss Goodman's example of *grue*.⁷⁴ Lewis says, "Think of the man who, for no special reason, expects unexamined emeralds to be *grue*" (1986: 38-9; 1984: 65). The extension of 'grue' consists of all green things before 2050 and all blue things after 2050. Suppose we assign our predicate 'green' the meaning *grue*. Intuitively, this is an unnatural assignment of meaning. But can this assignment be ruled out using the structural notion of naturalness? Let me explain in some detail why it cannot.

Consider the sentence 'Emeralds are green.' For the sake of the argument, let us interpret this sentence as meaning that emeralds are *grue*. Now, this sentence will be uttered by current subjects who exist before 2050, but also by *future* subjects who will exist after 2050. However, if the sentence is assigned the meaning that emeralds are *grue*, then these future subjects will come out saying *falsehoods*, because the extension of their term 'emerald' will not be part of the extension of their term 'green' (at that time). In effect, they will be attributing blueness to emeralds. Yet, emeralds will still be green.

However, the truth of these subjects' utterances can be preserved by re-interpreting 'emerald'. We can interpret 'emerald' as meaning: *emerire*, where the extension of 'emerire' consists of all the emeralds before 2050 and all the sapphires after 2050. If we interpret the sentence 'Emeralds are green' as meaning that *emerires* are *grue*, then these future subjects will no longer utter falsehoods. The extension of

⁷⁴ Goodman 1965.

their term 'emerald' will be part of the extension of their term 'green'. In effect, they will be attributing blueness to sapphires. This assignment of reference performs equally well in terms of truth.

But what about the property of reflecting light of a certain wavelength? Suppose that both current and future subjects believe that green is the surface property of objects that reflects light with wavelength of 500 nm. It appears, then, that future subjects will have incorrect beliefs about the relation between what they call 'green' and wavelength. However, the same trick as above can be applied. Consider the term 'wengthlave'.⁷⁵ The collection of things that reflect light with a wengthlave of 500 nanometer consists of all things that reflect light with a wavelength of 500 nanometer before 2050, and all things that reflect light with a wavelength of 480 nanometer after 2050. Using the same trick, we can preserve the truth of subjects' beliefs and utterances.

Invoking structural naturalness does nothing to block this assignment of meaning. This type of naturalness is something that is exemplified by the extensions of predicates of current and future fundamental theories—*whatever these extensions might be*. It is part of our fundamental theory that the referents of its term are natural. But this does nothing to narrow down interpretation. Lewis agrees that the notion of structural naturalness is of no help. "The proposed constraint is... *not* just that eligibility-*theory* is to be satisfied somehow" (1984: 66; italics added). Putnam's argument shows

⁷⁵ This term is from Hesse 1969.

that “principles of fit... leave the content of belief radically underdetermined” (1986: 38).

At this point, let me briefly mention an intuitive and common-sense explanation of why 'green' does not mean grue. This explanation will be explored further below. Green things before 2050 simply *look different* from blue things after 2050. Imagine a subject, a few minutes before midnight on December 31st in the year 2049, looking at an emerald and a sapphire on his desk in ideal lighting conditions. When the clock strikes midnight, the green emerald will continue to present him with the same experiences as before. And the same goes for the blue sapphire. It is not outlandish to suggest that such experiences play a role in determining the contents of our thoughts, especially when it comes to color concepts.

4.3 Non-Structural Naturalness and Its Problems

Lewis writes that “*only if we have an independent, objective distinction among properties, and we impose the presumption in favor of eligible content a priori as a constitutive constraint, does the problem of interpretation have any solution at all*” (1983a: 55; italics added). Lewis is not just adding ‘more theory’. Instead, he is proposing extra-theoretical restrictions on eligible models. On Lewis’s view, part of how terms acquire their referents is indeed the fit between the theory and the world. But in addition, the assigned referents to have meet certain naturalness requirements. These are external, not theory-mediated.

Lewis needs his naturalness constraints to be extra-theoretical. He writes, “It should *not* be said... that as a contingent psychological fact we turn out to have states whose contents involve some properties rather than others, and that is what makes it so that the former properties are more natural. (This would be *a psychological* theory of naturalness.)” (1983a: 54; italics added). Here, Lewis can be understood as saying that appealing to a structural notion of naturalness is not enough for answering Putnam’s challenge.

In the Lewisian framework, there is a type of naturalness that is all-or-nothing. The so-called ‘perfectly natural’ properties form a group of *elite* properties. They are fundamental and intrinsic. They are fundamental in that two worlds that agree in their perfectly natural properties and their distribution, agree objectively in all respects. They are intrinsic: not disjunctive or determinable, not structural or conjunctive. But what is most important about perfect naturalness is that it is *extra-theoretical*.

Perfectly natural properties are sometimes said to cut at the ‘joints of nature.’ On this view, the world contains properties that relate their instances (in terms of similarity or the sharing of universals), and this is a matter wholly independent of human theorizing. That these elite, fundamental properties relate their instances is *brute* fact, a *primitive* fact. ‘Primitive,’ in this context, does not mean that there are no concepts in terms of which this notion can be explained (after all, it is quite a sophisticated philosophical concept). Instead, it means that it is independent of the shape of our theories.

Perfect naturalness only exists at the fundamental level. But the identity of

complex, structural properties constructed⁷⁶ out of perfectly natural properties *depends upon* the identity of these natural ones. So perfect naturalness, while starting out at the bottom level, ‘spreads upwards’ throughout other properties in the world (Lewis 2009). At the fundamental level, there exist only properties with all-or-nothing naturalness. But less fundamental, ‘higher-up’ properties can inherit this type of naturalness to a smaller or larger degree.

‘Reference magnetism’ is the view that the content of our thoughts is *anchored*—in a way to be explained—to natural extensions.⁷⁷ Lewisian reference magnetism is best regarded as a view that the extensions of the terms that occur in the belief ascriptions of a regimented psychology must meet certain naturalness requirements. As suggested earlier, the totality of a subject’s beliefs can be represented as a belief in a modified Ramsey theory. This attribution will look like this: S believes that there is a unique n-tuple $x_1 \dots x_n$ $T(x_1, \dots, x_n)$. Reference magnetism is the claim that, when extensions are assigned to these variables, truth is not the only constraint: naturalness plays a role as well.

The theory of reference magnetism can be spelled out in a number of different ways. On some versions, naturalness constraints can trump structural fit. That is, a more natural collection can be the object of a belief, even though it performs worse in terms of structural fit, in terms of making the subject’s beliefs come out true. On other

⁷⁶ Some readers may not like the use of phrases like ‘constructed out of’ in this context. They can replace the phrase by ‘supervenes upon’ or perhaps ‘is entailed by’.

⁷⁷ The term ‘reference magnet’ occurs in Lewis 2004, as Schwarz (forthcoming) points out.

versions, naturalness merely functions as a tie-breaker. That is, if there are two mappings to extensions that perform equally well in terms of structural fit, the more natural mapping of the two amounts to reference.

Perfect naturalness does not exist in infinitely complex worlds. Lewis writes that in a “world of ‘structures all the way down’... no perfectly natural properties are instantiated” (2009: fn 4; also 1983a: 12). If there is no ‘bottom-level’ in our world, then perfectly natural properties are nowhere to be found. What is more, in such worlds, perfect naturalness does not ‘spread upwards’ throughout other properties in the world. In infinitely complex worlds, there is no all-or-nothing perfect naturalness at the bottom level, but also no objective naturalness insofar as this is inherited by complex properties.⁷⁸

My first argument against magnetism is the following. Reference magnetism implies that, if our world is an infinitely complex world, then our thoughts have no determinate content. However, we know that our thoughts have determinate content, but not whether or not our world is an infinitely complex world. Therefore, reference magnetism is false. We are way more confident in the fact that our thoughts have determinate contents than we are of the fact that our world has a fundamental, most-basic level. This appears to be a major weakness in the magnetist approach.

Now, a critic could argue that in infinitely complex worlds, there are *other* properties that are suitable replacements for Lewisian perfectly natural properties. For

⁷⁸ See Schaffer 2003 for a discussion of the view that there is no fundamental level.

instance, structural properties that are mentioned in theories of physics. The more fundamental the theory, the more natural the property. But note that this is to return to the degree-based, structural notion of naturalness discussed previously. I argued that no appeal to structural naturalness can be of help in generating determinacy in meaning. It is of no help, because structural naturalness is to be understood in terms of future theorizing. For finished theories, there is no future theorizing.

Exactly why is magnetism not compatible with infinite complexity? Suppose there is no bottom level and that the world consists of structural properties, going down indefinitely. Consider a fundamental theory at some time t . How is the reference of its terms determined? A possible answer is: Not just by the structural fit between the theory and the world, but also by how natural the assigned extensions are. But notice that whether these extensions are natural is a further structural fact about the world and, therefore, uncoverable by future theorizing. This means that the extensions of the terms of the current theory depend upon the fit between the future descendant of the current theory and the world. Magnetism has not played any role.

Now, suppose that our world is *not* one of 'structures all the way down.' Does reference magnetism at least in that case answer Putnam's model-theoretic challenge satisfactorily? Is magnetism at least in that case a satisfactory solution? In what follows, I argue that it is not. I claim that reference magnetism requires a certain view of fundamental properties. What is more, this view of fundamental properties is implausible because it is committed to the existence of *incorrigible error* about the world. I argue below that such incorrigible error does not exist.

Causal structuralism is the view that fundamental properties are individuated by their causal profiles or their nomological roles. It has been defended by Sydney Shoemaker among other philosophers.⁷⁹ The denial of causal structuralism is called ‘quidditism’. On the quidditist view, fundamental properties are individuated primitively, not by their causal profiles or nomological roles. The identity of fundamental properties is basic. Lewis is a quidditist.⁸⁰

The causal structuralist maintains that the Ramsey sentence of our total theory *individuates* fundamental properties. Consider again a total theory, $T(t_1, \dots, t_n)$. The modified Ramsey sentence of this sentence will have the form: $\exists!x_1 \dots \exists!x_n T(x_1, \dots, x_n)$. According to the causal structuralist, removing the existential quantifier ‘ $\exists!x_1$ ’ turns this into an open sentence that identifies the property referred to by theoretical term ‘ t_1 .’ (Usually not every single part of the theory is part of the meaning of ‘ t_1 ’—just a significant enough part. See our discussion in Section 1.3.)

For the causal structuralist, perfect naturalness is simply a *limiting case* of the structural notion of naturalness discussed in the previous section. We understood structural naturalness in terms of future theorizing. Extensions possess structural naturalness insofar as they are the referents of terms from future theories (including descendants of our current theories). So extensions possess perfect structural naturalness if they are the referents of the terms from the finished theory. This is the

⁷⁹ Shoemaker 1980.

⁸⁰ Lewis 2001.

structuralist understanding of perfect naturalness. It would appear then, that magnetism is not compatible with causal structuralism. Magnetism requires extra-theoretical naturalness; but structuralism employs no such notion.

(A quick aside: What about the terms '*causes*' or '*instantiates*'? If these terms occur in the Ramsey sentence, how is *their* interpretation determined? Let me briefly address both issues. First, '*causes*.' The key question here is the question of *Humean supervenience* (Lewis 1986: 14). Lewis holds the view that causal facts supervene on more basic facts. For people who accept Humean supervenience, the term '*causes*' drops out of the Ramsey sentence of the fundamental theory, because it is not a fundamental term. Problem solved.

Others believe that causal facts are basic and do not supervene. For them, the key question is whether '*causes*' is, using George Bealer's terminology, a '*semantically stable*' term or not.⁸¹ Semantically stable terms are terms that have the same meaning for language users in qualitatively identical epistemic situations. Bealer argues that many philosophical terms are semantically stable. If '*causes*' were both basic and semantically stable, it would perhaps resemble a term like '*is identical to*,' which is arguably a semantically stable term for a basic relation. Furthermore, it is plausible to suppose that the meanings of terms from this category are exhausted by their conceptual role.

The remaining theoretical option is the following: '*causes*' is a semantically

⁸¹ Bealer 1996.

unstable term for a basic relation. This would be to treat ‘causes’ in the same way as quidditists treat certain fundamental terms from the Ramsey sentence. Below, I will criticize such a form of quidditism. I will argue that it is *definitive* of quidditism that there are semantically unstable terms for basic relations. I will present an argument against the view that there are such terms. So let us postpone discussion of this combination of views until then.

Next up, ‘instantiates.’ This term, unlike ‘causes,’ cannot disappear from the Ramsey sentence. Instantiation facts are not made true by other, more basic facts; the basic facts *are* instantiation facts. So what is the correct semantics for ‘instantiates’? The most plausible approach, it appears, is to assimilate ‘instantiates’ to semantically stable terms for basic relations, such as ‘is identical to.’ One could argue that its meaning is exhausted by its conceptual role. On our approach, ‘instantiates’ is a term from the language of the hypothetical psychologist who attributes propositional attitudes to thinkers. So this psychology would implicitly define the concept of instantiation.)

Now, let us return to structuralism and magnetism. If the causal structuralist’s notion of perfect naturalness is just a limiting case of the structural notion of naturalness, then it is not compatible with reference magnetism. If this is right, then reference magnetism, in order to be a feasible approach, requires two things: first, it must be the case that the world has a fundamental level, second, causal structuralism must be false. Only if these two requirements are met, can reference magnetism do the job of answering Putnam’s model-theoretic argument. This is what I will assume in the arguments that follow.

Reference magnetism, then, must be combined with the denial of causal structuralism—quidditism. But how should we understand quidditism? How can appealing to ‘quiddities’ be useful for the magnetist? The situation is somewhat confusing here, because some authors understand quidditism (and also haecceitism) merely as theories about the *de re* modal characteristics of properties (or individuals). For instance, Lewis formulates quidditism as a theory of *de re* modality.⁸² David Chalmers and other writers have also understood quidditism as a theory of *de re* modality and properties.⁸³

On this understanding, quidditism is the view that the nomological role of a fundamental property is not necessary to being that property. Supposing positive charge to be a fundamental property, the view says that it is possible for positive charge to play a different nomological role. Differently put, this type of quidditism holds that property identity across possible worlds is not grounded in identity in nomological role, but is *primitive*. Usually, this type of quidditism also holds that the nomological role of a fundamental property is not sufficient for being that property. This means that it is possible for a property other than positive charge to play the positive charge-role. On this type of quidditism, in other words, fundamental theoretical descriptions are *non-rigid* (i.e. such descriptions pick out different properties in different possible worlds).

A natural question at this point is: In virtue of what do fundamental properties

⁸² Lewis 2009.

⁸³ Chalmers 2012, Locke 2012.

have these *de re* modal characteristics? If one does not want to be committed to primitive modality, there should be an informative answer to this question. Lewis, earlier in this career, seemed to be less willing to accept primitive modality than later on. In an article from 1980 article on the philosophy of mind, he writes, “Nonrigidity might begin at home... If a non-rigid name applies to different states in different possible cases, *it should be no surprise if it also applies to different states in different actual cases.* Nonrigidity is to logical space as other relativities are to ordinary space” (1980a: 126; italics added).

One can see how this works. If a term applies to different things in actual cases, then the role is distinct from the occupant, so to speak. For instance, the term ‘US president’ currently co-refers with the term ‘Obama,’ but they did not co-refer at other times. The role of president is distinct from its occupant. On a certain view, principles of recombination allow distinct things to be recombined in various configurations, and so ‘generate’ different possibilities. The quidditist (or haecceitist) differs from the causal structuralist (or anti-haecceitist) in which items they see available for recombination. For the causal structuralist, being a fundamental property and playing a causal role amount to the same thing. For the quidditist, they do not.

Some philosophers understand quidditism in this way. They see it as a view on how to individuate or type property instances in the actual world. John Hawthorne, for instance, allows two different fundamental properties within the same world to play the

exact same causal role.⁸⁴ Two different fundamental properties can be governed by identical laws of nature. Quidditism understood this way is a view on *intra*-world property individuation. A causal structuralist does not countenance such symmetric situations where there are structurally indiscernible fundamental properties.⁸⁵ In what follows, I will understand quidditism as committed to such scenarios being possible.

For purposes of illustration, consider again the case of gold and fool's gold from the previous section. Suppose there was a time when people were not able to distinguish the chemical element Au from iron pyrite. Imagine that these speakers would occasionally call samples of iron pyrite 'gold.' According to quidditism, a similar situation is possible at the fundamental level. Hawthorne's account implies exactly this. He writes that "a pair of properties may have equivalent causal profiles and yet asymmetrical patterns of instantiation, the latter affording a perfectly good basis for unique reference to one of the pair" (2001: 364).

Let me now present an argument against such a view. How instances of fundamental properties affect us is a matter of their causal profile. This means that *what we call* these instances is a matter of their causal profile. But on the form of quidditism under consideration, what property these instances are instances *of*, is not just a matter of their causal profile. This means that a subject could say, 'These are instances of positive charge,' yet be mistaken. They might be referring to 'fool's positive

⁸⁴ Hawthorne 2001.

⁸⁵ See also Schaffer 2005.

charge.’ What is more, this is a type of error that is *irremediable* by the acquisition of new knowledge about the world.⁸⁶ But, I claim, there is no such type of error. Therefore, quidditism is false.

The argument is not just that subjects can be in error, of course. Subjects who call iron pyrite ‘gold’ are also mistaken. However, in this case, the error can be corrected by learning new things about the environment. Once such speakers acquire more knowledge about the world, they will re-interpret their old utterances involving ‘gold’ as being about the chemical element Au, and consider some of their old utterances as false (viz. the ones made in the presence of iron pyrite). They will also start talking differently; they will stop calling iron pyrite ‘gold. This is the corrigible type of error we are familiar with.

However, in the case of quidditism, the situation is different. There is no amount of learning that can remedy the error that speakers make when they say, ‘These are instances of positive charge.’ The problem with quidditism is that subjects can have false beliefs that cannot be corrected by acquiring new knowledge about their environment. This strikes me as an implausible position. There are no such incorrigible false beliefs. I also deny that the view that there is no such error is simply *wishful thinking*. It might go too far to suggest that it is part of the concept of error that it is always corrigible. But

⁸⁶ David Chalmers and Wolfgang Schwarz criticize forms of reference magnetism where naturalness constraints can overrule speakers’ judgments (Chalmer 2012, Schwarz forthcoming). Chalmers considers a form of ‘ultra-strong reference magnetism’, according to which “there is not just a theory-external role for naturalism, but one that can trump speakers’ judgments about cases” (Chalmer 2012, excursus 20). The view we are considering, the combination of reference magnetism with quidditism, is different. Quiddities are such that speakers cannot have judgments or intentions about them.

still, this type of incorrigible false belief (if it did exist) would be very different from the type of false belief we are all too well familiar with.

Note also that this kind of error is not the same as *ignorance*. In his paper on quidditism, Lewis argues that we cannot know the nature of quiddities. He writes, “to the extent that we know of the properties of things only as role-occupants, we have not yet identified those properties” (2009: X). He concludes, “We are irremediably ignorant about the identities of the fundamental properties that figure in the actual realization of the true final theory” (2009: X). But my argument is not concerned with mere ignorance. My argument centers on the claim that subjects can have false beliefs about the world that cannot be corrected by any future inquiry.

In the end, I think Lewis’s view runs into trouble because it contains an irresolvable tension. On the one hand, he writes that, “The furniture of the *Lebenswelt* which presents us with our problems of decision and learning consists, in the first instance, of objects given qua objects of acquaintance, and individuated by acquaintance” (1994: 322). Exactly so. But he also writes, “What we say and think not only doesn’t settle what we refer to; it doesn’t even settle the prior question of *how* it is to be settled what we refer to. Meanings—as the saying goes—just ain’t in the head” (1984: 64). Lewis cannot have it both ways. Either what we mean by our words depends on us, or it does not. I maintain that it does.⁸⁷

⁸⁷ It will not help to suggest that Lewis’ two quotes can be reconciled because in the first he is talking about mental content, while in the second he is talking about linguistic meaning. Remember that by discussing the model-theoretic argument in terms of language, Lewis is merely playing along with Putnam (see 1984: 58, 1983a: 49). The model-theoretic argument is about the assignment of content to mental states. So Lewis’ two quotes concern the exact same thing: thought.

4.4 An Appeal to Experience

Is there a better solution to Putnam's model-theoretic argument? Or is there perhaps a view for which Putnam's puzzle does not even arise? As hinted at earlier, I think there is. Determinacy in meaning and content exists in virtue of a connection to *experience*. Such an approach can be called 'empiricist.' The empiricist outlook has been defended, in various forms, by many different philosophers over the last century, including Bertrand Russell. Formulating such a view in detail would be too big of a project to accomplish here, but let me provide a few sources of support.

At the end of section 4.2, I discussed the meaning of color terms such as 'red' and 'green.' On the functionalist account, such a term refers to the property that satisfies an associated collection of beliefs. A common-sensical view is that it is part of the meaning of terms like 'red' and 'green' that they refer to properties that make objects look a certain way. It may not be possible for speakers to put this part of the theory into words, but it nevertheless is included in the theory.

Consider 'grue' again. Earlier, I considered the question of why 'green' does not mean *grue*, why it does not have as its extension: green things before 2050 and blue things afterwards. According to the view that invokes a connection with experience, the reason is that green things before 2050 simply *look different* from blue things after 2050. If 'green' really meant *grue*, then a subject looking at an emerald at midnight on December 31st, 2049 would experience a change. At midnight, the subject would notice

that the gem stone that used to look green now looks *blue*. He would exclaim, “Hey, this gem changed its color!” Since this will not happen, ‘green’ does not mean grue.

Now, some may suggest that one could interpret ‘having the same color as’ as: having the same *shmolor* as. Blue things before 2050 have the same *shmolor* as green things after 2050. Some authors even imagine observers whose discriminatory apparatus is sensitive to *shmolor*, not color. J.S. Ullian considers such an observer.⁸⁸ In the presence of an emerald that remains green throughout the transition from 2049 to 2050, this observer will surely be surprised. Says Ullian, “[W]e can expect him to look aghast and mumble (for surely he will be too shocked to speak audibly), “It changed its *shmolor*, it changed its *shmolor*, it isn’t grue anymore”” (1961: 388). (Ullian’s subject uses different words from ours, but we can replace them with ‘color’ and ‘green.’)

Now, humans are unlike Ullian’s imagined observer, of course. Our sensory apparatus is sensitive to green, not grue. When humans are confronted with a green gem that turns blue, there will normally be changes in their mental state. It is part of folk-psychology that subjects notice whether or not a change takes places in their experiential state. According to folk-psychology, it is *definitive* of ‘being in a different experiential state’ that subjects are aware of it, that it causes changes in other mental states and potentially behavior.

But what if an idealized, God-like interpreter would interpret *our* thoughts as being about things that he calls ‘grue’? What if such an outside interpreter would say,

⁸⁸ Ullian 1961.

“Hey, look at these humans! How odd, their perceptual apparatus is sensitive to grue, not to green!” Then, I submit, our green-thoughts are really about ‘grue’ things. But here, ‘grue’ is a word in the interpreter’s language, not our human language. Our ‘green’ will *still* not mean: grue. This is all we need to prevent Putnam’s puzzle from arising, so the problem simply vanishes.

Let me now return to the account of narrow content sketched in Chapter 1, section 2. There, I argued for the view that a regimented psychology captures what people really believe and desire. Such a psychology describes environmental impact as inputs at the subject’s sensory surface, and the subject’s actions as outputs going in the direction of the environment. On this approach, the totality of a subject’s belief can be represented as a belief in a modified Ramsified theory. Our regimented psychology will attribute to subjects something like: S believes that $\exists!x_1 \dots \exists!x_n T(x_1, \dots, x_n)$.

On Lewis’s view, the content clause of this belief attribution will contain no ‘O-terms.’ Presumably, the only non-logical term is ‘instantiates.’ Now, is Lewis right in thinking that there are no O-terms? In order to examine this issue, we have to consider the types of covering generalizations used in psychological theorizing. We must carefully look at how folk-psychology explains and predicts. Only then can we figure out what the attitude ascriptions of a regimented, systematized version of folk-psychology will look like. And only then can we find out exactly what kinds of terms will occur in these attitude ascriptions.

Michael McDermott has in a series of articles⁸⁹ convincingly argued that the key psychological explanations ascribe *de re* attitudes. He writes, “When we turn specifically to psychological explanation, the subsuming of particular cases under covering generalizations, common-sense psychology is *de re* through and through” (2009: 264). It seems to me that if we take his lead, we can get a better understanding of the nature of mental content. A bonus of the view is that Putnam’s puzzle does not arise, because the attitude ascriptions indeed contain something resembling ‘O-terms.’

To consider a simple case, imagine a subject looking at a red ball in good lighting conditions. The layman would report the subject's belief with, ‘S believes that the ball is red.’ This belief is inferred from the common-sensical principle that people can tell the color of things by looking at them. Such a principle can be more precisely formulated as follows: For any person x, thing y and color z, if x looks at y and y instantiates z, then x believes of y and z that the one instantiates the other. This principle, with a premise describing an observation about what the subject is looking at, can then be instantiated to: S believes *of* the ball and redness that the one instantiates the other.

In support of the view that such belief attributions are not *de dicto*, McDermott imagines the subject performing a color-matching task. S not only believes the ball to be red, he believes it to be a certain *shade* of red. He is asked to pick out this shade from a chart of samples. He picks out sample #123. Now, consider S’s belief right before he selects the matching sample. There is no *de dicto* belief ascription that can explain S’s

⁸⁹ McDermott 1986, 1988, 2001, 2009.

successful color-matching behavior. The statement 'S believes that the ball is the color of sample #123' is not true yet. The statement 'S believes that the ball is the color of the ball' ascribes an empty belief. Since no *de dicto* clause is available, the belief must be *de re*: S believes *of* the ball and *of* the color of sample #123 that the one instantiates the other.

A properly regimented version of our folk-psychology, then, will not only describe inputs and outputs as events at the subject's interface with the world, it will also ascribe to such subjects *de re* attitudes *about* these inputs and outputs. In the words of McDermott, "[the] content sentences [of the best psychology] will contain no names or descriptions except ones referring to kinds of input and output; and they will be *de re* ascriptions, so as not to imply that the subject thinks of his in- and outputs under the same descriptions as the psychologist does" (1988: 236).

On this picture, subjects have direct and unmediated attitudes about objects of acquaintance—the *means by which* they recognize and act on their environment. This is an updated version of Russell's view on logically proper names. But with important differences. The view is not that the subject's language contains terms for these in- and outputs. The Ramsey sentence that represents the totality of a subject's beliefs contains O-terms—but they are the *interpreter's* O-terms, not the *interpretee's*. Nevertheless, these O-terms refer to entities that the subject has *de re* beliefs about. These are special *de re* beliefs that capture psychological content (they differ therefore fundamentally from the type of *de re* beliefs discussed in Section 2.2).

On such an empiricist account of mental content, these O-terms are where

mental content is anchored, where it 'bottoms out.' The type identities of subjects' in- and output states partly determine the type identities of their mental contents. So the attitudes ascriptions from the best psychology *do* contain names, after all. But they turn out to be very special names, not of environmental entities or kinds, but rather of objects of acquaintance. These objects of acquaintance are the things by means of which subjects perceive and influence the world around them.

Summing up, McDermott writes, "[W]hat you believe is a fully Ramsified theory of the environmental causes and effects of your in- and outputs" (1988: 233). These attitude attributions from a regimented psychology are, of course, miles away from the belief ascriptions we produce in our day-to-day explaining and predicting of people's internal states and behavior. But, they capture people's real propositional attitudes. These attitudes ascriptions capture people's beliefs and desires insofar as these are independent of how their environment happens to be constituted.

What is more, such a view is not susceptible to Putnam's model-theoretic argument. Determinacy in content is due to a connection with experience and action. Subjects' propositional attitudes are about environmental entities and kinds that are related in various ways *only* insofar as this makes a difference to their experience and behavior. This is an account of content that does not suffer from Putnam's threat of indeterminacy. In virtue of connections to experience and action, thoughts determinately represent the world around us as being a certain way.

(It also strikes me that this view has a dialectical advantage. Earlier, we considered the issue of whether our world has a fundamental level or not. Perhaps it

does, perhaps it does not. Whatever the answer to that question ends up being, the current account succeeds. But what is more, the question whether our world has a fundamental level may be a *bad question*; a question where no philosophical progress can be made. If that is the case, the current account also outperforms Lewis's view. It does not need the puzzle of fundamentality to even make sense.)

Now, it is likely that some philosophers will group this view together with certain obsolete or long-discredited theories. These critics may use the labels 'empiricism,' 'anti-realism,' 'verificationism,' or even 'phenomenalism.' They may also accuse the view of being incompatible with materialism. In my opinion, some of these categorizations are incorrect. Others should not be regarded as pejorative. Rather, some of these labels indicate that the view *takes a side* on an issue that philosophers have long fought over. Taking a side in a long-standing philosophical debate, on which the jury is still very much out, is not a negative. It is a positive.

First off, the theory in question is not a form of phenomenalism. While the view counts as a form of *indirect* realism, subjects do not merely have beliefs about past, current and future experiences. Subject's thoughts are about entities in the world and the relations between them to the extent that this makes a difference to their experience and behavior. Whether the view counts as 'anti-realism' is tough to say, given how notoriously difficult it is to define that term. The same applies to 'verificationism' and 'empiricism.' In sufficiently weak senses, the view may qualify. But this should not be considered a negative.

Does the view conflict with materialism? Lewis would probably say that it does.

He considers the doctrine of 'revelation' on which color experiences are associated with "a simple, ineffable, unique essence that is instantly revealed to anyone who has that experience" (1997: 352). He continues, "[I]t is false by materialist lights—and we have pledged ourselves non-negotiably to materialism" (353). The problem with this point is that it reverses the order of priority of two philosophical investigations. The philosophical investigation into the notions of belief and meaning is prior, it would seem, to the question of materialism. A theory of the former can help shed light on the question what it means to be a materialist. But teasing out these difficult matters is a story for a different occasion.

LITERATURE

Anderson, David 1993. What Is the Model-Theoretic Argument? *Journal of Philosophy* 90, 6: 311-22.

Bach, Kent 1987. *Thought and Language*. Oxford: Oxford University Press.

Bach, Kent 2002. Giordione Was So-called Because of His Name. *Philosophical Perspectives* 16: 73-103.

Bach, Kent 2004. Descriptions: Points of Reference. In Marga Reimer and Anne Bezuidenhout (eds.) 2004. *Descriptions and Beyond*. Oxford: Clarendon Press.

Bays, Timothy 2008. Two Arguments Against Realism. *Philosophical Quarterly* Vol 58, #231: 193-213.

Bealer, George 1996. On the Possibility of Philosophical Knowledge. In *Nous, Vol. 30 Supplement, Philosophical Perspectives 10: Metaphysics*: 1-34.

Borg, Emma 2004. *Minimalist Semantics*. Oxford: Oxford University Press.

Burge, Tyler 1973. Reference and Proper Names. *The Journal of Philosophy* Vol. 70, 14: 425-39.

Burge, Tyler 1977. Belief *De Re*. *Journal of Philosophy* 74: 338-362.

Cappelen, Herman and Ernest Lepore 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Blackwell Publishers.

Carnap, Rudolph 1947. *Meaning And Necessity: A Study In Semantics And Modal Logic*. Chicago: University of Chicago Press.

Chalmers, David and Frank Jackson 2001. Conceptual Analysis and Reductive Explanation. *Philosophical Review* 110: 315-61.

Chalmers, David 2004. The Foundations of Two-Dimensional Semantics. In M. Garcia-Caprintero and J. Marcia (eds.) 2004. *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press.

Chalmers, David 2012. *Constructing the World*. Oxford: Oxford University Press

Chisholm, Roderick 1976. *Person and Object*. London: G. Allen & Unwin.

- Devitt, Michael 1981. *Designation*. New York: Columbia University Press.
- Devitt, Michael and Kim Sterelny 1987/1999. *Language and Reality: An Introduction to the Philosophy of Language* (2nd ed.) Cambridge: MIT Press.
- Donnellan, Keith 1966. Reference and Definite Descriptions. *Philosophical Review* 75: 281-304.
- Donnellan, Keith 1970. Proper Names and Identifying Descriptions. *Synthese* 21: 335-58.
- Evans, Gareth 1973. The Causal Theory of Names. *Aristotelian Society: Supplementary Volume* 47: 187-208.
- Dretske, Fred 1981. *Knowledge And The Flow Of Information*. Cambridge: MIT Press.
- Evans, Gareth 1979. Reference and Contingency. *The Monist* 62: 161-89.
- Frege, Gottlob 1891. Function and Concept. In Michael Beaney (ed.) 1997. *The Frege Reader*. Oxford: Blackwell.
- Frege, Gottlob 1948 (original 1892). Sense and Reference. *Philosophical Review* Vol 57, #3: 209-30.
- Frege, Gottlob 1956 (original 1918). The Thought: A Logical Inquiry. *Mind* Vol. 65, no. 256: 289-311.
- Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge, Massachusetts: Harvard University Press.
- Fodor, Jerry A. 1987. *Psychosemantics*. Cambridge: MIT Press.
- Grice, H.P. 1957. Meaning. *Philosophical Review* Vol 66, #3: 377-88.
- Grice, H.P. 1967. Logic and Conversation. In H.P. Grice 1989. *Studies in The Way of Words*. Cambridge: Harvard University Press.
- Goodman, Nelson 1965. *Fact, Fiction and Forecast*. Indiana: Bobbs-Merrill.
- Hawthorne, John 2001. Causal Structuralism. *Philosophical Perspectives* 15: 361-78.
- Hawthorne, John and David Manley 2012. *The Reference Book*. Oxford: Oxford University Press.

- Hesse, Mary 1969. Ramifications of 'Grue'. *British Journal for the Philosophy of Science* 20, 1: 13-25.
- Horwich, Paul 1998. *Meaning*. Oxford: Clarendon Press.
- Jackson, Frank 1998. Reference and Description Revisited. *Philosophical Perspectives* 12 Language, Mind and Ontology: 201-218.
- Kaplan, David 1969. Quantifying In. *Synthese* XIX, 2: 178-214.
- Kaplan, David 1977. Demonstratives. In Joseph Almog, Howard Wettstein and John Perry (eds.) 1989. *Themes From Kaplan*. Oxford: Oxford University Press.
- Kaplan, David 1978. Dthat. In French, Peter A., Theodore E. Uehling, Jr., Howard K. Wettstein (eds.) 1979. *Contemporary Perspectives In The Philosophy of Language*. Minneapolis: University of Minnesota Press.
- Kaplan, David 1989. Afterthoughts. In Joseph Almog, John Perry and Howard Wettstein 1989. *Themes from Kaplan*. Oxford: Oxford University Press.
- Kripke, Saul 1972. *Naming and Necessity*. Cambridge: Harvard University Press.
- Kripke, Saul 1977. Speaker's Reference and Semantic Reference. *Midwest Studies in Philosophy, Volume II: Studies in the Philosophy of Language*: 255-276.
- Kripke, Saul 1979. A Puzzle About Belief. In *Meaning and Use*. Margalit, Avishai (ed.) 1979. Dordrecht: Reidel.
- Kroon, Frederick 1987. Causal Descriptivism. *Australasian Journal of Philosophy* 65: 1-17.
- Lewis, David 1968. Counterpart Theory and Quantified Modal Logic. In David Lewis 1983. *Philosophical Papers Vol. 1*. New York: Oxford University Press.
- Lewis, David 1970. How to Define Theoretical Terms. In Lewis 1983. *Philosophical Papers Vol. 1*. New York: Oxford University Press.
- Lewis, David 1972. Psychophysical and Theoretical Identifications. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Lewis, David 1975. Language and Languages. In Lewis 1983. *Philosophical Papers Vol. 1*. New York: Oxford University Press.
- Lewis, David 1979. Attitudes *De Dicto* and *De Se*. In David Lewis 1983. *Philosophical Papers, Vol. 1*. Oxford: Oxford University Press.

Lewis, David 1980a. Mad Pain and Martian Pain. In Lewis 1983. *Philosophical Papers Vol. 1*. New York: Oxford University Press.

Lewis, David 1980b. Index, Context, and Content. In Lewis, David 1998. *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.

Lewis, David 1981. What Puzzling Pierre Does Not Believe. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David 1983a. New Work for a Theory of Universals. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David 1983b. Individuation By Acquaintance And By Stipulation. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David 1984. Putnam's Paradox. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David 1986. *The Plurality of Worlds*. Oxford: Blackwell.

Lewis, David 1994. Reduction of Mind. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David 1997. Naming the Colours. In David Lewis 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewis, David 2004. How Many Lives Has Schrodinger's Cat? In Frank Jackson and Graham Priest (ed.) 2004. *Lewisian Themes: The Philosophy of David K. Lewis*. Clarendon Press: Oxford.

Lewis, David 2009. Ramseyan Humility. In David Braddon-Mitchell and Robert Nola (eds.) 2009. *Conceptual Analysis and Philosophical Naturalism*. MIT Press.

Loar, Brian 1976. The Semantics of Singular Terms. *Philosophical Studies* 30: 353-78.

Loar, Brian 1980. Names and Descriptions: A Reply to Michael Devitt. *Philosophical Studies* 38: 85-9.

Loar, Brian 1988. Social Content and Psychological Content. In Grimm, Robert and Daniel Merrill 1988. *Thought and Content*. Arizona: University of Arizona Press.

Locke, Dustin 2012. Quidditism without Quiddities. *Philosophical Studies* Vol 160, #3:

345-63.

Martinich, A.P. 1977. The Attributive Use of Proper Names. *Analysis* 37: 159-63.

McDermott, Michael 1986. Narrow Content. *Australasian Journal of Philosophy* 64, 3: 277-88.

McDermott, Michael 1988. The Narrow Semantics of Names. *Mind* 97: 224-37.

McDermott, Michael 2001. Quine's Holism and Functionalist Holism. *Mind* 110: 977-1025.

McDermott, Michael 2009. A Science of Intention. *Philosophical Quarterly* Vol 59, #235: 252-73.

Merrill, G. E. 1980. The Model-Theoretic Argument Against Realism. *Philosophy of Science* 47: 69-81.

Mill, John Stuart Mill 1949 (orig. 1843). *A System of Logic*. London: Longmans, Green & Co.

Neale, Stephen 1990. *Descriptions*. Cambridge: MIT Press.

Newman, M.H.A. 1928. Mr. Russell's 'Causal Theory of Perception.' *Mind* 37: 137-48.

Nunberg, Geoffrey 2004. Descriptive Indexicals and Indexical Descriptions. In Marga Reimer and Anne Bezuidenhout (eds.) 2004. *Descriptions: Semantic and Pragmatic Perspectives*. Oxford: Oxford University Press.

Ostertag, Gary (ed.) 1998. *Definite Descriptions: A Reader*. Cambridge: MIT Press.

Perry, John 1977. Frege on Demonstratives. In John Perry 2000. *The Problem Of The Essential Indexical And Other Essays*. Expanded Edition. Stanford: CSLI Publications.

Perry, John 1988. Cognitive Significance and New Theories of Reference. In John Perry 2000. *The Problem of the Essential Indexical and Other Essays*. Expanded Edition. Stanford: CSLI Publications.

Perry, John 1997. Indexicals and Demonstratives. In Hale, Bob and Crispin Wright (eds.) 1997. *A Companion To The Philosophy Of Language*. Oxford: Blackwell Publishing.

Putnam, Hilary 1975. The Meaning of 'Meaning'. In Keith Gunderson (ed.) 1975. *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.

- Putnam, Hilary 1977. Realism and Reason. In Hilary Putnam 1978. *Meaning and the Moral Sciences*. New York: Routledge and Kegan Paul.
- Putnam, Hilary 1983. Models and Reality. In Hilary Putnam 1983. *Realism and Reason: Philosophical Papers Vol. III*. Cambridge: Cambridge University Press.
- Quine, W.V.O. 1956. Quantifiers and Propositional Attitudes. *The Journal of Philosophy* LIII, 5: 177-87.
- Quine, William V. O. 1960. *Word and Object*. Cambridge: MIT Press.
- Quine, W.V.O. 1977. Intensions Revisited. In Quine, W.V.O. 1981. *Theories and Things*. Cambridge: Harvard University Press.
- Reimer, Marga 1991. Demonstratives, Demonstrations, and Demonstrata. *Philosophical Studies* 63: 187-202.
- Russell, Bertrand 1905. On Denoting. In Bertrand Russell 1956. *Logic and Knowledge*. New York: Capricorn Books.
- Russell, Bertrand 1910. Knowledge by Acquaintance and Knowledge by Description. In B. Russell 1929. *Mysticism and Logic*. New York: W.W. Norton.
- Russell, Bertrand 1918. *The Philosophy of Logical Atomism*. In B. Russell 1956. *Logic and Knowledge*. New York: Capricorn Books.
- Russell, Bertrand and A.N. Whitehead 1927. *Principia Mathematica, Vol. I*, 2nd ed. Excerpts in Ostertag, Gary (ed.) 1998.
- Searle, John 1958. Proper Names. *Mind* 67: 166-73.
- Searle, John 1969. *Speech Acts*. Cambridge: Cambridge University Press.
- Schiffer, Stephen 1972. *Meaning*. Oxford: Oxford University Press.
- Schiffer, Stephen 1978. The Basis of Reference. *Erkenntnis* 13: 171-206.
- Schiffer, Stephen 2005. Russell's Theory of Definite Descriptions. *Mind*, Vol. 114, #456: 1135-84.
- Sosa, Ernest 1995. Fregean Reference Defended. *Philosophical Issues*, Vol. 6: Content: 91-99.
- Schaffer, Jonathan 2003. Is There A Fundamental Level? *Nous* 37:3: 498-517.

- Schaffer, Jonathan 2005. Quiddistic Knowledge. *Philosophical Studies* 123: 1-32.
- Schwarz, Wolfgang (forthcoming). Against Magnetism. *Australasian Journal of Philosophy*.
- Sellars, Wilfrid 1956. Empiricism and the Philosophy of Mind. In Herbert Feigl and Michael Scriven (eds.) 1956. *Minnesota Studies in the Philosophy of Science*, vol. I. Minneapolis: University of Minnesota Press.
- Shoemaker, Sydney 1984. Causality and Properties. In Sydney Shoemaker 1984. *Identity, Cause, and Mind*. Cambridge: Cambridge University Press.
- Siegel, Susanna 2002. The Role of Perception in Demonstrative Reference. *Philosophers' Imprint* Vol 2, #1: 1-21.
- Soames, Scott 2002. *Beyond Rigidity: The Unfinished Semantic Agenda Of Naming And Necessity*. Oxford: Oxford University Press.
- Stanley, Jason and Zoltán Gendler Szabó 2000. On Quantifier Domain Restriction. *Mind and Language* Vol 15, #2/3: 219-61.
- Strawson, Peter 1950. On Referring. *Mind* 59, #235: 320-44.
- Strawson, Peter 1959. *Individuals*. London: Methuen.
- Ullian, J.S. 1961. More on "Grue" and Grue. *Philosophical Review* 70, #3: 386-89.
- Ziff, Paul. 1960. *Semantic Analysis*. Ithaca: Cornell University Press.