INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reconciling A Naturalistic Theory of Meaning with Intentions: The Causal/Conceptual Approach

by

Nancy Adrienne Salay

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August, 1998

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Reconciling A Naturalistic Theory of Meaning with Intentions: The Causal/Conceptual Approach"

by Nancy Salay

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: August 25, 1998

External Examiner

Research Supervisor

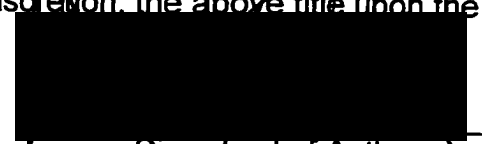
Examining Committee

Copyright Agreement Form

Dalhousie University

Date: August 27th, 1998
Author: Nancy Adrienne Salay
Title: Reconciling A Naturalistic Theory of Meaning with Intentions: The Causal/Conceptual Approach
Department: Philosophy
Degree: Ph.D.
Convocation: Fall
Year: 1998

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.



Signature of Author

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrights material appearing in this thesis (other than brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

TABLE OF CONTENTS

Abstract

The main aim of this thesis is to bring into the open what I see as a serious confusion in contemporary naturalistic approaches to meaning, where the programme is to explicate our semantic and intentional concepts in non-semantic and non-intentional terms. The problem is this: energies are currently being spent on naturalizing intentionality by appealing to a non-intentional relation with the world, namely a causal one, when that relation is really an intentional one. It *is* generally recognized that there are two issues that need dealing with; that of explaining why a particular state is an intentional one (and further how it gets individuated from other intentional states) and that of explaining what makes certain systems capable of *having* intentional states. I argue, however, that because answering the first sort of question necessarily entails an appeal to the intentionality of the system having the mental state, these two questions *must* be answered together, rather than treated as separate projects as they are today. Since no non-intentional answer to the first question can be given unless one is provided in an answer to the second question, a causal account of intentionality that is not underwritten by a non-intentional account of what it is to *be* an intentional system will not have succeeded in providing a naturalistic answer to our problem.

In the final chapter, I suggest what a causal account that embraces these conclusions might look like. On this approach, linguistic meaning reduces to speaker meaning, and speaker meaning is explained by appeal to what I call a causal/conceptual relation. The first step in explicating this relation is to distinguish between thoughts and concepts. In the first three sections of this chapter, I discuss what concepts are and what role they play in our mental lives. Then I discuss what thoughts are and how they are related to concepts by providing an answer to three central questions of intentionality: 1) *How* do we get the stuff in our heads? 2) What makes the stuff in our heads *about* stuff? 3) *What* out there is the stuff in our heads about? In the final section, I explore how this new view can be used to answer two traditional problems in the philosophy of language, namely, explicating attributive uses of definite descriptions and meeting Frege's challenge to views that identify meaning with reference.

Acknowledgements

I wish to thank the following people for helping me put this thesis together: Dr. Peter Schotch, my supervisor, for his unfailing confidence in me and for giving me the boost I needed whenever I was self-doubting, frustrated, or just plain glum; Dr. Richmond Campbell, for his insightful comments on (among other things) naturalism and realism; Dr. Duncan Macintosh, for his exacting scrutiny of each and every draft and for the consequent invaluable constructive criticisms; Dr. Melinda Hogan, for helping me see more clearly countless issues in the philosophy of language and for many hours of edifying and fun conversation; Dr. Robert Martin, for all of his counter-examples and for his willingness to help me work past them; each of the above again, for tirelessly working through countless drafts with me; Dr. Steven Burns, for taking over all bureaucratic burdens and for calming my frazzled nerves when things seemed to go awry; Dr. Tom Vinci, for helping me with my anti-foundationalist position (notwithstanding his own foundationalist one); Judith Fox, for her constant willingness to help and for her always cheering smile and positive word; Susan Macleod, for her patience; and countless others around the department who have helped in ways too numerous to mention here.

Outside of the department I would like to thank the following: Rick Salay, my brother, for inducing me to become a philosopher and then for fostering my interest in philosophy of mind and for being my compadre in more philosophical discussions than I can remember; my parents, for their unflagging confidence in me; all of my friends, who offered me words of encouragement when I was overwhelmed by my task; and, finally I would like to thank the person who struggled through this thesis almost as much as I did, Morgan Wade, for being endlessly patient with all of my misgivings, frustrations, break-throughs, and let-downs and for offering his love and support when I needed them most.

Introduction

This is a book about meaning. Perhaps surprisingly, much of it is spent talking about mental states, the problem of intentionality, cognitive structures, and the like. As we proceed, the reason for this approach will become increasingly clear. My main aim is to bring into the open what I see as a serious confusion in contemporary naturalistic approaches to meaning, where the programme is to explicate our semantic and intentional concepts in non-semantic and non-intentional terms. The problem is this: energies are currently being spent on naturalizing intentionality by appealing to a non-intentional relation with the world, namely a causal one, when that relation is really an intentional one. This mistake arises, I suspect, from trying to explicate the reference *relation* while ignoring the special natures of the *relata*, namely, the people, the intentional systems, that engage in this relation with the world. Thus while this question – what makes this or that brain state an intentional state? – is one that surely needs answering, I want to suggest that we are moving off track if we try to answer it in isolation from another, more fundamental question: what makes certain systems capable of having intentional states and others not?

Now many naturalistic theorists, or at any rate the good ones, *are* attempting to give an answer to both of these questions. To answer the first sort of question about intentionality, they provide some kind of causal story that

explains that a certain brain state is an intentional state because it is causally related to the world in the requisite way. To answer the second sort of question about intentionality, they provide some kind of functionalist story (maybe teleological) about why we would want to call one system an intentional one and another one not. This is not to describe what everyone in the field is doing; for example, some try to answer questions of the first type with functional/teleological answers[1] and others want to avoid functional/teleological explanations altogether.[2] Nevertheless, there is some sense that there are two issues that need dealing with: that of explaining why a particular state is an intentional one (and further how it gets individuated from other intentional states) and that of explaining what makes certain systems capable of *having* intentional states. My main focus here will be to show that these two questions *must* be answered together, rather than treated as separate projects as they are today. This is because, as I shall argue, answering the first sort of question necessarily entails an appeal to the intentionality of the system having the mental state in question; that is, no non-intentional answer to the first question can be given unless one is provided in an answer to the second question. Thus, a causal account of intentionality that is not underwritten by a non-intentional account of what it is to *be* an intentional system will not have succeeded in providing a naturalistic answer to our problem. Now a few words about motivation.

My interest in this project stems from the desire to bring together the following intuitions. It seems right to suppose that we can (and indeed ought if we are interested in giving an *explanation)* give a non-semantic account of meaning and give a non-intentional account of mental states. This is what naturalistic theorists are attempting to do. But the following intuition seems right too. The way in which we see the world is determined not just by the world, but also by the way we sort the world; consequently, none of our observations are objective in the sense of reporting the way the world really is. Further (and this is really a third intuition but it tends to go hand in hand with the second), the idea that there is a way in which the world really is, independently of the way we sort the world, is incoherent.

Now these two intuitions, as stated, are not in obvious tension; nevertheless, those who are convinced by the first tend not to hold the second, and those who agree with the second tend not to hold the first. To see this we need only note that the vast majority of those who are attempting to provide a non-semantic account of meaning and a non-intentional account of intentionality are doing so in the hope of naturalizing these notions, of giving them scientific respectability. But the judgement that an account is scientific is generally taken to entail the judgement that it provides an account of the way one aspect of the world really is. Hilary Putnam makes this point with respect to Jerry Fodor's naturalistic account of meaning (a theory we will be looking at in some detail): "...

genuine sciences, Fodor assumes, tell us what we have to assume to be there mind independently." (Putnam, 1992, 66) Putnam, in contrast, thinks that the second intuition, the one rejecting the objectivity of our observations, is correct: "Our language cannot be divided up into two parts, a part that describes the world 'as it is anyway' and a part that describes our conceptual contribution. This does not mean that reality is hidden or noumenal; it simply means that you can't describe the world without describing it." (Putnam, 1992, 123) He is sceptical, however, of the naturalizing project: "What I have been suggesting is that the philosophy of language is in a bind because it is hell-bent on eliminating the normative in favour of something else, however problematic that something else may be." (Putnam, 1992, 79)

I think that we can have it both ways; we can proceed with the project of providing a non-semantic account of meaning and a non-intentional account of intentionality and at the same time accommodate the second intuition and let go of the idea that our observations and hence our theories of the world capture "the way the world really is". In order to see how this is going to work, we first have to get clearer on the naturalistic agenda.

Many philosophers nowadays purport to be engaging in inquiry that is naturalistic; unfortunately, it is, more often than not, unclear exactly what is contributing to the naturalism of a given enquiry. In particular with respect to the programme of naturalizing intentionality (our main focus of interest), theorists

who claim to be naturalists are notoriously vague about what exactly their naturalism amounts to. As Steven Wagner observes, "... while professed naturalists abound, they tend hardly to explain themselves. Naturalism is widely taken to be understood and accepted from the start." (Wagner, 212) In the course of arguing that naturalism is doomed to failure, Wagner distils from the literature a working target. While I disagree with his conclusions, his work in the area of clarifying the domain on naturalism is valuable, so I will use his characterization to situate the discussion.

According to Wagner, in order for naturalism to constitute a genuine approach, it must satisfy two criteria: "First, it must have critical and methodological import. It must somehow guide inquiry and place restrictions on our theories and attitudes. Second, it must be a distinctive position." (Wagner, 222) After rejecting various possible candidates (for example, Wagner rejects the suggestion that naturalism is a rejection of vitalism and natural theology because this would not serve to distinguish naturalism from empiricism in general), Wagner concludes that there are two possible naturalistic stances, namely, hard and soft naturalism. Hard naturalism is the view that only the natural sciences – physics, chemistry, biology, etc. – constitute acceptable theories about the world. The results of the non-natural sciences – anthropology, philosophy, psychology, astrology, etc. – are not considered to be genuine contributions to our knowledge of the world. Soft naturalism, on the other hand,

is the view that some of the non-natural sciences ought to be taken more seriously than others; thus, while astrology ought not to be taken seriously at all, because predictively it fails miserably, psychology, since it performs quite well in its rough and ready form, could be elevated to the level of the natural sciences if intentionality, the central element in psychology, could be explained in non-intentional terms. So while the hard naturalist thinks that the non-natural sciences will eventually be eliminated (when we can scientifically explain all of the phenomena that are the subject of the non-natural sciences), the soft naturalist thinks that some of the non-natural sciences may eventually take their place alongside the natural sciences. The naturalistic accounts of meaning we will be looking at are thus naturalistic in this soft sense, since the motivation for naturalizing meaning is the belief that such a procedure will help psychology on the path to becoming a natural science. Hereafter, 'naturalism' will be used in the soft naturalism sense.

So the programme of naturalizing intentionality is that of providing a non-intentional explanation of it. But why is providing a *non-intentional* explanation of intentionality the key requirement to elevating the status of psychology to that of a natural science? The answer is that natural sciences, it is generally supposed, paint a picture of the way the world is because they are not infected with particular ways of thinking about the world, that is, because they are stated in non-intentional terms. If a theory were to be stated in intentional terms, its

objective status would be in jeopardy, since its claims would depend upon a particular way of thinking about the world. Thus, in order for a theory to count as a theory of natural science, it must be stated in non-intentional terms.

Now while I agree with the general programme, that we need to give a non-semantic account of meaning and a non-intentional account of intentionality, I disagree with this particular line of reasoning. We do not want to naturalize these notions in order to ensure the objective status of the resulting theory -- no theory has that status (intuition number two); rather, we want to naturalize these notions because only then will we have given an *explanation* of them (intuition number one).

But, one might suppose, accepting intuition number two just means giving up on intuition number one for the following reason. If none of our observation claims are objective because all of our observations are determined, in part, by how we think about the world, then *all* of our observation claims are intentional and hence no non-intentional explanations are even possible. But this does not follow, although I think that this kind of reasoning is the motivation for rejecting intuition number one once intuition number two has been accepted. There is a difference between acknowledging 1) that an explanation is intentional in the sense that the observations appealed to in the course of the explanation were arrived at in an intentional way, namely through perception; and, 2) that an explanation is intentional in the sense that it appeals to intentional notions in its

explanation. Every explanation we give will be intentional in the first sense but not every explanation need be intentional in the second sense. In order to naturalize intentionality, we need to avoid the latter sort of explanation. And this is what contemporary accounts do not manage to do I will argue. They fail because in their accounts they appeal to a relation with the world, a causal one, that they suppose is not intentional, when it is.

In chapter one, I aim to present the thought processes that have led to current understanding, at least in some circles, of what the central problems of meaning are. To this end, I discuss general questions of meaning in the Frege/Russell tradition, Wittgenstein's suggestion that we view meaning as use, the consequent trend toward pragmatic theories of meaning, and finally, the causal accounts that serve as the naturalistic underpinnings of the pragmatic approach.

In chapter two, I argue that the naturalism of causal theories that appeal to an account of non-intentional perception is undermined by the fact that none of our perceptions are non-intentional. Since we cannot naturalistically explain our intentional states by appealing to our non-intentional interaction with the world, we need to give a naturalistic explanation of the fact that this interaction with the world is always intentional; that is, we need to explain what it is to *be* an intentional system. I look at one suggestion to this end, namely, Allan Gibbard's functionalist account of representation. In the final section of this chapter, I fill out

the picture of intentional perception I have been advocating by developing connections between it and John McDowell's illuminating philosophical discussion of our cognitive relation with the world. In the course of this discussion, I elaborate on some metaphysical considerations relevant to the matter at hand, namely, the issue of whether any of our observations are objective. According to this view of intentional perception, none of them are. This result should not be seen as a failing of the view, however, because, as I argue, there are independent reasons for supposing that the notion of an objective observation cannot be made clear anyway.

In chapter three, I suggest what a causal account that embraces the conclusions of chapter two might look like. On this approach, linguistic meaning reduces to speaker meaning, and speaker meaning is explained by appeal to what I call a causal/conceptual relation. The first step in explicating this relation is to distinguish between thoughts and concepts. In the first three sections of this chapter, I discuss what concepts are and what role they play in our mental lives. Then I discuss what thoughts are and how they are related to concepts by providing an answer to three central questions of intentionality: 1) *How* do we get the stuff in our heads? 2) What makes the stuff in our heads *about* stuff? 3) *What* out there is the stuff in our heads about? In the final section, I explore how this new view can be used to answer two traditional problems in the philosophy

of language, namely, explicating attributive uses of definite descriptions and meeting Frege's challenge to views that identify meaning with reference.

Chapter One ••• Theories of Meaning

Introduction

In this chapter I present what I take to be the progression of thoughts, with respect to questions about meaning specifically, that have led to the current trend in naturalistic theories of meaning. What I say here is by no means inclusive of everything important that has occurred in the philosophy of language; nevertheless, the material covered under the four general headings, namely, description theories of meaning, meaning as use theories of meaning, pragmatic theories of meaning, and causal theories of meaning, will prepare the reader for the discussions in chapters two and three.

1.1 Description Theories of Meaning

Questions about language often begin with questions about what function language serves: in some, if not in most, cases, language serves as a tool for communicating our thoughts about objects in the world. This being so, it seems that words must enable us to refer to the things in the world we are interested in talking about. A moment's reflection will reveal that not all words play this role: to what objects do operators such as 'and' and 'or' or articles such as 'a' and 'the' refer? But many words and many collections of words do serve to pick out specific objects in the world. For example, 'Aristotle' and 'the man who founded

the Lyceum' pick out one individual, namely, Aristotle. Frege terms all such words and expressions *Eigennamens* or proper names. (Frege, 1892, 57) Thus, what we generally take to be names, such as 'Fred', 'Sally', etc. as well as descriptive expressions that single out individuals, such as 'the boy in the red coat', are all proper names for Frege.

But we might ask, is reference all there is to proper names? That is, once we have identified the object that is picked out by a word or description, have we said all there is to say about the meaning of that word or description?

Frege's Account of Meaning

Frege thought the answer must be no.[1] His reasoning can be seen as a reductio of the following sort.

Suppose that the answer is yes – one has completely explained the meaning of a word or description when one gives its reference. Then, in cases where two different signs refer to the same object, observing that a=b (where 'a' and 'b' stand for proper names) will be no more informative than observing that a=a. For since, by hypothesis, when one has identified the object of reference one has exhausted the term's meaning, and since there is just one object of reference, then what one is saying by uttering 'a=b' is the same as what one is saying when one utters 'a=a'. But anything of the form a=a is known to be true *a priori* (and is tautologous as well), while statements of the form a=b, if true, are known to be true only upon investigation. To see this, consider Frege's example:

'the morning star' and 'the evening star' are proper names that take the same reference, but it was many hundreds of years before humans discovered this fact. Thus, in giving just the reference of a term, one has not explained everything about its meaning.

With this argument, Frege presents a strong case for treating proper names as having a meaning distinct from reference. This meaning Frege terms a proper name's sense.

What is a sense? Sense, Frege thought, is the mode of an object's presentation. That is, the what-it-is that is grasped when one understands a name of an individual or object. Now we have to be careful here not to confuse senses with ideas. Senses, while they are not completely objective in the same way that actual objects are – objects exist independently of observers – are also not completely subjective in the way that ideas are – ideas are constructions of minds. Arguably, everyone's ideas are different. This is a product of the fact that we all have different interests and different experiences and, consequently, different perspectives on things and happenings in the world. But senses are commonly grasped; otherwise, we would not be able to communicate at all. Mode of presentation is fixed by the object itself and by its relation to the being to which it is being presented. Frege's moon example illustrates this point nicely:

> Somebody observes the Moon through a telescope. I compare the moon itself to the reference; it is the object of the observation, mediated by the real image projected by the

> object glass in the interior of the telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the idea or experience. The optical image in the telescope is indeed one-sided and dependent upon the standpoint of observation; but it is still objective, inasmuch as it can be used by several observers. At any rate it could be arranged for several to use it simultaneously. But each one would have his own retinal image. (Frege, 1892, 60)

Thus, while senses are not objective in the way in which objects, references, are, they are not completely subjective either; they are determined in part by a certain standpoint of observation but, just as each of us can look at the image in a telescope, that standpoint is one which we can all share. It is unclear from Frege's discussions just where this more subjective element of sense derives from, although it is clear that its objective element is determined, at least partially, by the object, the referent, in question. There must, then, be a relation between sense and reference. The relationship, Frege thought, could be characterized in this way: a proper name expresses its sense and designates its reference. (Frege, 1892, 61) Of course, some proper names have only a sense and no reference, as in the case of proper names used in fiction and poetry, for example. This, Frege speculates, is because in natural languages, as opposed to ideal languages, we are not always interested in discovering and communicating what is true or false about the world; sometimes we are more interested in creating moods and evoking feelings: "In hearing an epic poem, for instance, apart from the euphony of the language we are interested only in the sense of the

sentences and the images and feelings thereby aroused." (Frege, 1892, 62) When we are interested in truth, however, the reference becomes important in establishing the connection between our words and the world; with only a sense we can communicate thoughts, but we need a reference if we are to make judgements about the world. Now clearly a judgement about the world will involve more than just a proper name; it will involve making assertions. Thus we move to a consideration of sentences.

Just as a proper name has a corresponding sense and reference, a sentence has a corresponding proposition and reference. A proposition is the thought expressed by a sentence, while the reference is the truth value of the sentence, that is, "the circumstance of its being true or false." (Frege, 1892, 63) The thought expressed by a sentence is objective in the way that the sense of a proper name is. And just as some proper names have a sense but no reference, so, too, some sentences have a sense, that is, express a proposition, but have no reference, no truth value. Examples of such sentences are ones, such as the following, that are constituted by proper names that have no reference: 'Sherlock Holmes has a keen mind' is neither true nor false (has no reference) since the proper name 'Sherlock Holmes' has only a sense and no reference.

So far I have said nothing about the connection between sense and reference or between proposition and truth value. Exactly what this connection is has been notoriously unclear since Frege himself never develops his account of

it, giving us only this general statement: "The regular connexion between a sign, its sense, and its reference is of such a kind that to the sign there corresponds a definite sense and to that in turn a definite reference, while to a given reference (an object) there does not belong only a single sign." (Frege, 1892, 58) Indeed, that this relation cannot be made clear is precisely Russell's point against Frege. But we can, at least, make the matter somewhat clearer.

Let us consider first the relation between sense and reference with respect to proper names (since Frege takes it that the reference of whole sentences depends upon what is going on in their constituent clauses). Now if a sense is given in "the mode of presentation" of an object and is objective in the same sense that the image on the lens of a telescope is, then we might say that the referent of a proper name is partially "responsible" for the sense of the proper name. Of course, cases where there is no reference, for example in the case of 'unicorn', pose a problem; but perhaps in those cases we can understand the sense to be the complex expression of a number of different referents. I will not go into detail working this out here because I think that there are some deep problems with this division between sense and reference; however, Frege might have had something like the following in mind. We know what the sense of 'horse' is because, among other things, we have had experience with horses. We know what the sense of 'horn' is because, among other things, we have had experience with horns. And so on. We

put together the sense of 'horse' with the sense of 'horn' with other relevant senses to get an understanding of the sense of 'unicorn'.

Now I have been alluding to "other factors", implying that reference is only partially responsible for sense. Frege does not talk much about what other factors besides reference enter into the determination of sense; however, that there must *be* other factors is clear from Frege's insistence that sense is different from reference but is also different from subjective ideas we have about objects and concepts. The subjectivity of sense comes from its being the way in which objects are presented to *us*; while the objectivity of sense comes from it being *an object* that is being presented to us. But the previous is not enough to distinguish senses from images since much the same can be said of the latter as well. What Frege needs is an explanation for why or how there is a level at which objects are presented to us all *in the same way.* Because such an account is missing from Frege's discussion, it is difficult to pin down precisely what sense is for him.[2] For the moment, however, let us continue with what we have, namely, some connection between sense and reference: the reference is the object of the presentation while the sense is (or is given in) the mode of its presentation. The reference is partially responsible for the sense, then, and the sense (having been determined, in part, by the reference) picks out, or designates, its reference. Now to sentences.

The sense of a sentence is, for Frege, the proposition or thought the sentence expresses. This proposition in turn is a composite of the senses and propositions

of the sentence's constituent proper names and clauses. Thus, if the sense of a proper name is connected to its reference, and if the proposition of a sentence is a compound of its constituent senses, then the reference of a sentence, its truth value, is connected to the proposition expressed by the sentence via its constituents' connections.

One might, at this juncture, object that I have equivocated on the term 'reference' in Frege; that Frege uses the term to mean one thing with respect to proper names – namely, the object being named – and a completely different sort of thing with respect to sentences – namely, either of the two abstract objects, the True or the False. First let me say that I am not attempting here a scholarly defence of Frege's account; rather, I am interested in explicating his view in the context of a general discussion of the main points of interest in theories of meaning, ultimately in order to position my own contribution in this area. Since, in chapter three, I want to reject the distinction between sense and reference, whether or not we can explicate the connection between the two is not crucial. However, there is much that, I suspect, is right in what Frege says and, for this reason, I will digress here in order to suggest why the previous discussion is compatible with what Frege says about reference.

In "On Sense and Reference", Frege says the following about sentential truth value:

> By the truth value of a sentence I understand the circumstance that it is true or false. There are no further truth values. For brevity I call the one the True, the other the False. Every declarative sentence concerned with the reference of its words is therefore to be regarded as a proper name, and its reference, if it has one, is either the True or the False. (Frege, 1892, 63)

Frege takes the True and the False to be objects to which sentences refer. A natural question is whether they are unrelated to the individual references of the proper names that constitute sentences. If so, then according to this view, the True and the False are two independent objects to which sentences refer. Thus, under this interpretation, the world is filled with tables, chairs, cats, etc. and the True and the False. I think that this way of understanding Frege is extremely misguided. First, it is hard to see what would be the relevance, on such a picture, of the references of individual proper names to the references of the sentences they constitute; and yet from Frege's discussions it is equally clear that he thinks that sentential meaning and reference *are* compositionally determined. Second, from the excerpt above where Frege introduces the terms, it seems clear that he is using the term 'object' in a special way; otherwise, he would not feel the need to qualify by saying "What I mean by an object can be more exactly discussed only in connexion with concept and relation. I will reserve this for another article." (Frege, 1892, 64) Third, Frege says, "For brevity I call the one the True, the other the False." (Frege, 1892, 63) We should take him seriously when he says this (why would we not after all?), and look to the

previous explanation to see what it is that he is abbreviating in this manner. There we see that by 'truth value' he means "the circumstance that it [the sentence] is true or false." (Frege, 1892, 63) Now a circumstance is not an object in the same sense that we take tables and chairs to be objects; however, in certain contexts we often find it useful to treat events, for example, as objects. Of course circumstances are different from events: we would want to call *the cat being on the mat* a circumstance rather than an event (especially if the cat had been on the mat for the past twelve hours) because an event is the having of a property by an individual at a time and does not necessarily involve a change; it is precisely this change that marks whatever it is that we are singling out *as* an event. Nevertheless, it is at least plausible to understand the objects the True and the False as the circumstance that makes the sentence true or false respectively. If our sentence is 'the cat is on the mat', then to say that it refers to the True is shorthand for saying that the cat, referred to by the proper name 'the cat', stands in the relation (expressed by 'is on') of *being on* to the mat, referred to by the proper name 'the mat'. It is this reading of the reference of sentences that I made use of above in my brief description of how sentence reference is determined by constituent proper name reference, and how, consequently, sense might be said to determine reference. Now back to our discussion of the importance of the connection between sense and reference.

Since the reference of a sentence is the circumstance of it being true or false, then, the connection between sense and reference ensures the connection between meaning and the world. But, of course, things are not this straightforward. Let us look just at the question of reference for a moment.

Consider that, given the above discussion, two sentences having constituent clauses that take the same reference must have the same truth value (although they can have different senses or propositions associated with them). Some sets of sentences, however, do not obey this principle of substitutivity. Consider the following: 1) 'Fred believes that the woman in the red dress is wonderful'; 2) 'Fred believes that the CEO of the company is not wonderful.' As it happens, the woman in the red dress *is* the CEO of the company. Given the principle of substitutivity, we ought to be able to exchange these expressions, *salva veritate*. But we cannot. The sentence 'Fred believes that the CEO of the company is not wonderful' is true while the sentence 'Fred believes that the woman in the red dress is not wonderful' is false. Now, why is this a problem? Well, it is a problem for two reasons.

1) Consider Leibniz's law of identity: if two things are identical, then they share all of their properties. By hypothesis we stated that the woman in the red dress and the CEO of the company are one and the same individual. But the statements seem to contradict the principle: the woman in the red dress has the property *being believed by Fred to be wonderful* while the CEO of the

company has the property *being believed by Fred to be not wonderful*. Thus, it must be that the woman in the red dress is not the CEO of the company. But she is. So the problem becomes one of retaining Leibniz's law, while at the same time explaining how the woman in the red dress can be the CEO of the company.

2) Now we return to the compositionality problem alluded to above. If we take the reference of sentences to be determined compositionally in the way that Frege thinks, then we are going to have a problem explaining why, as in the case above, two sentences that are composed of clauses and proper names that take the same reference, can still differ in truth value. That is, if the truth value of sentences depended upon the reference of their parts, then, in the case above, the truth value (reference) of the true sentence 'Fred believes the woman in the red dress is wonderful' must depend upon the reference of its constituent clauses and proper names. But when we substitute a proper name with the same reference, namely, 'the CEO of the company', the truth value of the sentence changes – it is now false. Prima facie, this is a counter-example to our principle that the reference of a sentence is determined by the reference of its parts, or, in other words, that the truth value of a sentence is determined by the reference of its parts. We now have to decide what we are willing to give up: the principle or the example?

Frege opts for the second alternative by arguing that, in certain contexts, descriptions and names often refer to things other than to what they normally refer. These special contexts are called oblique contexts; in them, proper names and clauses refer to something other than their ordinary referents. In the case we have been discussing, because of the description's position in the sentence inside of the oblique context 'believes that ...', the problem is solved by treating the reference of 'the CEO of the company' as the sense ordinarily expressed by 'the CEO of the company'. Since the sense of 'the CEO of the company' is different from the woman in the red dress or even from the sense of 'the woman in the red dress', the proper names do not have the same reference and hence our inability to substitute *salva veritate* is not a threat to the principle of substitutivity.

This solution also answers our quandary with salvaging Leibniz's law. Since it turns out that the proper names do not co-refer in this case, it is not the case that one and the same object has different properties. The CEO of the company does not have the property *being believed by Fred to be not wonderful*; rather, the sense of 'the CEO of the company' does. It may also be the case that the woman in the red dress does not have the property *being believed by Fred to be wonderful*; it may be that the sense of 'the woman in the red dress' does.

But Russell was dissatisfied with Frege's approach. Russell thought that in order for senses to play the role that Frege's theory demanded of them, namely

that of determining reference, there must be a logical connection between sense and reference. For if the connection were only linguistic, that is, if the connection were only of the following sort – if Fred is the reference of 'Fred', then the sense of 'Fred' is **Fred**[3] – then it is unclear how sense can determine reference. As Simon Blackburn and Alan Code point out, in order for the sense of 'Aristotle', for example, to determine Aristotle, there must be some connection between that sense and that object of reference. Otherwise we will be unable to explain the difference between, for example, the following two sets of sentences:

1) Aristotle taught philosophy.
2) Jones believes that Aristotle taught philosophy.

and the pair;

3) Aristotle, the magnate, married Mrs. Kennedy.
4) Aristotle, the philosopher, wrote books. (Blackburn & Code, 1978, 71)

While there is a logical connection between the first pair of sentences, in virtue of the connection between the reference of 'Aristotle' in 1) and the reference of 'Aristotle' in 2) – they imply, for example, that Jones believes something true – there is no such connection between the second pair of sentences. Our explication of the relation between sense and reference ought to explain this difference.

But, Russell argues, there is no way of providing such a non-linguistic definition of sense and its relation to reference. If sense can be defined,

presumably we should be able to refer to the sense of an expression. But we cannot. On the one hand, simply making up a word to designate the sense of a word in order to *mention* the sense, as in explaining the sense of 'Aristotle' by appealing to **Aristotle**, provides only a linguistic definition because in doing this we have not gone beyond our words to the object and its related sense; while, on the other hand, if we *use* the sense expressed by 'Aristotle', we end up talking about Aristotle when we wanted to talk about the sense of the proper name 'Aristotle'. If we try harder with the phrase 'the sense of Aristotle', then we end up talking about Aristotle's common sense. (Russell, 1905, 167-168) So in the one case the definition proves to be merely linguistic, while in the other we are unable to get at the thing that needs defining. Further, Blackburn and Code point out, one cannot explain sense by appealing to its reference because while senses determine references, references do not determine senses. As Russell notes, "... there is no backward road from denotations to meanings, because every object can be denoted by an infinite number of different denoting phrases" (Russell, 1905, 168).

Of course, here, Russell seems to be equivocating between denoting phrases and senses. For Frege, as we have just seen, the two are very different since the denoting phrase is what expresses a sense; it is not to be identified as the sense. Once we see this, we also see that while a single reference may be "denoted by an infinite number of different denoting phrases," we recognize that

this says nothing about the relation between references and senses: it is possible that an infinite number of denoting phrases may express identical senses. Indeed, given Frege's discussion on sense, that the object of presentation is partially responsible for the sense that is given in presentation, this would have to be the case. Of course, it *may* also be the case that a single object of reference can have many different senses, and this *may* cause problems. Saul Kripke, in his "Puzzle About Belief" (1979) raises just such a difficulty. We will digress for just a moment here to introduce that problem.

Kripke points out that Frege, while emphasizing that senses are different from ideas (the former are shared while the latter are private), acknowledges that for any given object of reference, a number of different senses may be associated with it. This is because an object may be given in presentation in different ways. Each of these senses, however, is objective in that it is there to be "apprehended" by us whenever we turn our attention in its direction. I will quote Frege at length in order to make this point clear as it will become important in the discussion to follow.

> Accordingly, with a proper name, it depends on how whatever it refers to is presented. This can happen in different ways and every such way corresponds with a particular sense of a sentence containing a proper name. The different thoughts which thus result from the sentence correspond in their truth-value, of course; that is to say, if one is true then all are true, and if one is false then all are false. Nevertheless their distinctness must be recognized. So it must really be demanded that a single way in which whatever is referred to is presented be associated

> with every proper name. It is often unimportant that this demand should be fulfilled but not always. (Frege, 1918, 524)

Thus, in natural languages, according to Frege, it may, perhaps often, be the case that one individual is associated with many different senses and that two people, referring to the same individual, understand something very different by the proper name they use to refer to that individual. Kripke, in thinking about just such a situation, noticed an undesirable consequence of this. Suppose that someone, who becomes acquainted with an object of reference under two different descriptions, is unaware of this fact and forms two contradictory beliefs about the object. In Kripke's example, it is monolingual French-speaking Pierre who forms the belief that Londres est belle, while the same Pierre, although at a later date and now bilingually English and French speaking, forms the belief that London is ugly. Assuming that Pierre is sane we can say, in Fregean terminology, that Pierre associates two different senses with the city London; one evoked by the proper name 'Londres' and the other by the proper name 'London'. The problem, however, is that we cannot say what Pierre believes about the city London; does he believe that it is ugly or that it is pretty? And more disturbingly still, no matter what Pierre *does* ultimately believe, it is impossible for us to *say* what he means when he says 'Londres est belle' for example. If we say that, by this, he means what we mean when we say 'London is pretty' we run into trouble because Pierre explicitly denies the statement 'London is pretty'. But suppose

we hold that Pierre associates the same sense with 'Londres' that we associate with 'London' and so we express his statement 'Londres est belle' with 'London is pretty'. But then how do we describe what he means by 'London is ugly'? We cannot take 'London' as we understand it, because this, we already decided, was associated with the sense of Pierre's 'Londres', and Pierre believes that what he calls 'Londres' is pretty, not ugly. Thus, we cannot express both of Pierre's beliefs. That is, we cannot say what Pierre means by both of his statements. It is due to problems such as these, then, that Russell's objection that "there is no backward road from denotations to meanings, because every object can be denoted by an infinite number of different denoting phrases" (Russell, 1905, 168) stands. And any theory that makes this split between meaning and denotation, or sense and reference, will be faced with these sorts of problems, Russell claims: "the relation of the meaning to the denotation involves certain rather curious difficulties, which seem in themselves sufficient to prove that the theory which leads to such difficulties must be wrong." (Russell, 1905, 166)

Russell's Account of Meaning

But perhaps there is a way of explaining meaning without bringing in this mysterious notion of sense. Russell's idea was that names and definite descriptions do not really refer to particular objects; rather they denote particular objects via the properties to which they refer. The difference can be thought of something like this. Up until Russell presented his alternate way of viewing what

are commonly thought of as names, philosophers thought that names functioned to designate or pick out specific individuals in the world. That is just what it means, people thought, to be a name *of* something. But then there was this problem. What about those names and descriptions (proper names, for Frege) for which there is no corresponding object? Under this category fall fictional names, mythological names, and names and descriptions of people and events that do not exist. Consider statements about Santa Claus for example: 'Little Johnny believes Santa Claus lives at the north pole.'[4] How do we evaluate this statement and at the same time retain a compositional view of sentences and their truth values? 'Santa Claus' has no reference because Santa Claus does not exist. But clearly the statement has meaning and, consequently, we might want to say (Russell certainly did) that it has a truth value as well. Now if we follow Frege and introduce the notion of sense, then we can explain the meaning of this statement by stating that while 'Santa Claus' cannot refer to Santa Claus, since Santa Claus does not exist, we can still understand the sense of 'Santa Claus' and so the statement has meaning. Whether the statement is true or false, of course, depends upon what we take 'Santa Claus' to be referring to in this case; but, for Russell, who does not want to appeal to sense, there remains the problem of explaining how it is that the name can remain meaningful even though it has no reference. A nice and quick way to see the ultimate difference between Frege's and Russell's approach is to observe that while for Frege a

name could be meaningful and yet have no reference, for Russell, a name is meaningful only in so far as it has a reference (although not, necessarily, a denotation – 'reference' applies to properties while 'denotation' applies to individuals); consequently, for Russell, all meaningful sentences have determinate truth values, while for Frege, some meaningful sentences lack truth value (have no reference).

Russell's solution to the problem of meaningful names that lack reference is to treat so-called names such as 'Santa Claus' as really general statements about properties. Consider the following two statements:

1) Fred is hungry.
2) Someone is hungry.

We can treat both of these statements as essentially of the same form, namely, as though a predicate is being ascribed to a particular individual. Under this interpretation, we take 'Fred' and 'someone' as individual-designating names and the rest of the sentence as what is being predicated of those individuals. But there is a difference between the two. While it seems to be the case that in 1) an individual, namely Fred, is being designated, it is not clear what is being designated in 2). To whom exactly does 'someone' refer? It seems, then, that 1) and 2) are not of the same form at all. The following translations illustrate this point:

1) $\text{Hungry}_{\text{Fred}}$
2) $(\exists x)\ (\text{Hungry}_x)$

In 1) the name 'Fred' designates a particular individual while in 2) 'someone' indicates that a certain general state of affairs obtains, namely, that there exists an individual having a certain attribute. Russell, however, thought that perhaps the two statements *are* of the same form after all.[5] However, unlike as was previously thought – that the two statements are statements about particulars of the form of 1) – Russell surmised that the two statements are really both general statements of the form of 2). On this view, so-called names like 'Fred' are really disguised definite descriptions. Thus, the proper translation of 1) looks like this:

$$(\exists x)(Fred_x \wedge (\forall y)(\ Fred_y \supset x=y) \wedge Hungry_x)$$

On this translation 'Fred' does not designate an individual, rather 'Fred' is a description of a general state of affairs, one that describes relationships between property holders. On this view, then, the sentence is really like the one in 2), 'there exists an individual having the property *being hungry'*, because it breaks down into the following: There exists only one individual having the property *being Fred* and that individual has the property *being hungry*.

Some of Russell's terminology needs to be clarified here before we can continue. Predicates are property attributions. For example, *being lazy* and *being smart* are predicates. A predicate's extension is the set of things of which the predicate is true. So, if the predicate in question is *being lazy,* then all beings who are lazy will be a part of its extension. Descriptions such as 'Santa Claus',

'the woman in the red dress', etc. have extensions as well: the things in the extensions of descriptions are those things that are denoted by the description, or, in Russell's terminology, the description's denotatum. In the above two examples, the denotatum of 'Santa Claus' is empty because Santa Claus does not exist and the denotatum of 'the woman in the red dress' is the woman in the red dress.

Let us consider our problem cases once again. First, we need a way of explaining how it is that empty names contribute to the meanings of sentences. Second, we need to explain, without giving up the relationship between truth and reference, why, in some cases, descriptions and names are not intersubstitutable *salva veritate*. And third, we need to show, in those same cases, how we can retain Leibniz's law of identity.

The problem of empty names gets resolved in this way. Let us take the sentence 'the king of France is bald'. This sentence is meaningful even though the description 'the king of France' is empty, that is, has no reference. Russell's treatment of "names" as definite descriptions solves the problem in the following way. 'The king of France' does not refer to an individual; rather, it refers to a set of properties and denotes that individual who has these properties. In order to determine the truth value of the sentence, we first have to find out the denotata of its constituent descriptions. In this case we find that the extension of the predicate *being king of France* is empty. So the statement is false because there

is nothing that has both the property *being the king of France* and *being bald.* But the sentence has meaning in virtue of the properties to which the description refers, namely, kingship of France and baldness.

Cases where intersubstitutivity *salva veritate* is blocked are explained in a similar fashion. Here we see that descriptions that we thought co-referred do not really co-refer after all. Take our example one more time. 'The CEO of the company' and 'the woman in the red dress' do not refer to the individual that fits both of those descriptions; rather, the descriptions denote that individual but *refer* to the properties *being a unique CEO of the company* and *being a unique woman in the red dress* respectively. Since these are clearly different properties, the descriptions do not co-refer. These kinds of 'oblique' contexts are special only insofar as mistakes are often made in these situations because it is generally less obvious in these contexts than in others where reference lies.

And Leibniz's law is retained of course: the sentences do not refer to one individual with incompatible properties, rather they refer to the different properties that one individual happens to have.

Having described Russell's view to this extent, we can now ask of Russell's treatment the following: has he successfully answered the question of sentence meaning without appealing to a notion of sense? On one construal he seems to have – sentence meaning comes in via reference after all, but it is via reference to properties, not individuals. But what are properties? More important, what

does Russell take properties to be? There are many ways to answer this question – here we will discuss two.

One way is to take Russell as saying that properties are identical with their extensions; that is, a property just is the set of all things having that property. But if this is the case, then Russell's solution fails because the distinction between denoted individual and referred to property collapses: if properties just are their extensions, then co-extensional predicates like 'the CEO of the company' and 'the woman in the red dress' ought to be intersubstitutable *salva veritate* in all situations. But we saw that they are not. Russell's solution is to note that while the predicates are co-extensional, the properties referred to by the predicates differ. Thus, for this distinction to have any force, properties must be distinct from the individuals exemplifying those properties.

This brings us to the second way of seeing Russell's understanding of properties, namely, as something over and above extensions. On this view, membership in the extension of a predicate does not exhaust the meaning of 'property' because it is possible for two properties to be co-extensive and yet not be identical. Cordate and renate are examples of such properties; every individual that has the property *being a cordate*, that is, *being a thing that has a heart*, also has the property *being a renate*, that is, *being a thing that has a kidney*. Nevertheless, cordate and renate are different properties. Russell makes the distinction between these properties by appealing to the idea that they

stand for different universals. Given that Russell thinks of universals as mind independent existing entities, something I will get to in a minute, this makes Russell a realist about properties.

Note that it makes no difference to Russell's commitment to the existence of properties whether we take him to be saying that property ascriptions like 'x has the property *being a king*' *refer* to properties or merely *express* properties. If Russell takes them to refer to properties, then clearly he is committed to property existence since one cannot refer to something that does not exist – this was one of the puzzles that motivated his account in the first place. On the other hand, if Russell takes them to express properties, then he is equally committed to property existence because, on his own view, one cannot express something that does not exist – it is precisely for this reason that Russell takes 'the king of France is bald' to be expressing (or referring to) the properties of kingship of France and baldness rather than to be expressing anything about the King of France. Either way, property realism is essential to Russell's view. And he freely admits to this position.

In *The Problems of Philosophy,* Russell describes a universal as "anything which may be shared by many particulars ...." (Russell, 1912, 53) Falling under this category are properties and relations. He offers two arguments there in favour of his position, the first to establish that universals must exist, and the

second to establish that this "existence" is mind independent. I will present these now, somewhat formally, for ease of reference.

The Existence of Universals Argument

1. Suppose the empiricists are correct in their denial of the existence of universals. Then the following characterization of universals must be wrong: i) for any universal ϕ, anything that is ϕ is so in virtue of the fact that it has the quality of ϕ. And the following explanation of properties and relations must be correct: ii) when we understand that thing in virtue of which all individuals having property ϕ have something in common, we do not understand the property ϕ as abstracted from all particulars; rather, we think of particular objects that have that property and then we "reason concerning this particular, taking care not to deduce anything concerning it which we cannot see to be equally true of any other [ϕ] thing." (Russell, 1912, 55)

2. But in order to apply the method espoused in ii) above, we need to understand and use the concept of resemblance; that is, when we choose our sample, we reason about only those things that bear the correct sort of resemblance relation to it.

3. The resemblance needed will have to be a universal – since the relation will have to hold between many different pairs of things (and this is just what we mean by 'universal').

4. Since we must have this one universal, there is no longer reason to reject others.

The Mind-Independence of Universals Argument

1. If universals were mind-dependent, then the various natures of individuals, i.e. their properties and the relations that hold between them, would be influenced by our thoughts about them.

2. But relations and hence properties are not influenced by our thoughts and knowledge about them. Russell provides the following example in support:

> When we come to know that Edinburgh is north of London, we come to know something which has to do only with Edinburgh and London: we do not cause the truth of the proposition by coming to know it, on the contrary we merely apprehend a fact which was there before we knew it. The part of the earth's surface where Edinburgh stands would be north of the part where London stands, even if there were no human being to know about north and south, and even if there were no minds at all in the universe. (Russell, 1912, 56)

3. Therefore, universals are mind-independent.

Now this realist assumption, while helping Russell deal with one problem, that of ensuring a referent for all definite descriptions, does not help at all in answering the question of how our words and sentences hook up with the world. This is somewhat surprising since, as we have just seen, Russell's unease with Frege's characterization of the relation between sense and reference, namely as a determining relation, stemmed from the suspicion that this relation was merely a linguistic one rather than a logical one that holds between senses and objects of reference. For if the meanings of our words and sentences are going to have any relation at all to the world, then they must do so at the point of reference; that is, the sense or meaning of a term must have *something* to do with what gets to be the term's reference. But Russell suspected that there was no way of characterizing the determining relation such that, given a proper name's sense, it uniquely picks out its related reference. The challenge to Fregeans, then, is this: unless Fregeans can give an account of this vital connection, the Fregean view

will not be able to explain how our words communicate things about the world. Now we can ask, is Russell's view any more successful in explaining this connection? I suspect that it is not.

Russell thinks that he is avoiding Frege's problem by not making a distinction between sense and reference. He deals with the morning star/evening star problem by making a distinction between reference and denotation; descriptions and proper names refer to general property terms while they denote individuals. Thus, descriptions and names that fail to denote are neither nonsense nor do the sentences which they constitute lack a truth value: they are false if an individual is not denoted and either true or false otherwise. Russell, then, reduces reference-to-individuals to reference-to-properties and denotation of individuals. But in doing this, Russell fails to explain the connection between reference and denotation. To explain this we need to know in virtue of what a thing has a property. Russell assumes, for example in premiss two of The Mind-Independence of Universals Argument above, that property attribution and identification is somehow unproblematic and can be done outside of language. But this is not to offer any *explanation* of how our language hooks onto the world.

Of course Russell, as a Platonist about universals, might answer the question in this way: an individual has a property in virtue of the universal that it partakes in. But this is a deeply unsatisfying answer for a number of reasons.

First, there is the host of reasons why Platonism is unpalatable.[6] Second, even were we to accept the framework that Russell is working within, we may still ask – but how do we identify *this* property as *that* universal? Do we not already have to *know* what the universal is; that is, do we not already have to have a *language* for talking and thinking about universals? But if this is the case, then any account of meaning that relies on the existence of properties but remains silent about how properties and language are related, does not address the central issue of how our language hooks onto the world – because it assumes that we already have a language, a language of universals, in order to explain the rest of language. So Russell's account will not do either because, while he has a story to tell about what in the world our words and sentences our hooked onto, he does not explain how they are hooked up.

1.2 Meaning as Use Theories of Meaning

And there were other problems with Russell's account as well. Most of these problems can be characterized as stemming from the following circumstance: while Russell was trying to give an account of the way in which sentences and expressions function in an ideal language, questions about how less than ideal languages, namely ordinary languages, work, remained unanswered. People began to notice that, for example, there is a difference between a sentence and a use of that sentence;[7] that definite descriptions can be used in different ways;[8] that

not all sentences are used to make statements;[9] that the meaning of the use of an expression or sentence is importantly related to what the speaker intends to mean by her usage – that the former cannot be understood until the latter is first understood;[10] and so on.

And somewhere around this time, Wittgenstein recanted on some of what he had propounded in *The Tractatus* by suggesting a new approach to the question of meaning. In his *Philosophical Investigations,* Wittgenstein rejects any philosophy of language enterprise that tries to answer the following sorts of questions. What meaning does this word or expression or sentence have? What is the role of a word or expression or sentence in a language?[11] What exactly is the nature of that which we understand when we come to understand what a word or expression or sentence means? And so on. The framework for this kind of approach to questions about language was set by Frege by his distinction between meaning and reference. And the Fregean idea, that when we understand a word or expression we grasp its sense, is a compelling one. Why give up on such a powerful notion then? Well, one reason is that the concept of sense may be incoherent: there is no way, as Kripke pointed out, to say what the sense of a given word or expression is; and, if we cannot *say* what the sense of an expression is, how can we possibly *grasp* it, that is, say it to ourselves? Wittgenstein's response was to abandon the entire sense/reference framework.

First, he recognized that words and expressions and sentences only have meanings within contexts. We cannot understand what is meant by a statement, question, command, etc. until we know the context within which it is being uttered. For example, the question 'how do names get their meanings?' cannot be answered until we specify that it is this or that language-game that we are talking about. For while in one context names may stand for individuals, in another they may stand for descriptions of qualities that a certain individual has. And there may be different games played within a larger game within which names function differently again. Put another way, it is misleading to search for *the one way* in which various kinds of words mean or function within a language because words function differently in different contexts (§14).

And it is equally misleading to search for the meaning of words in these different contexts or games. That would be to fall once again into the Fregean trap of supposing that there is some identifiable thing that is grasped when we understand a word or expression. Do not ask for what the meaning of a word is, Wittgenstein famously said, ask instead for its use. On this approach, sameness of sense, to use the old terminology, becomes sameness of use within a context, while difference in sense becomes difference in use within a context. What does it mean to say that I understand what this word means? It means that I understand how to use it. What does it mean to say that this word has a meaning? It means that this word can be used in a certain way.

One way of making the conception of meaning as use clearer is to make the distinction between *knowing how* and *knowing that.* On the traditional approach to language, understanding the meaning of a word is seen as a form of knowing *that* such and so, where knowing that such and so has been variously interpreted as, for example, knowing *that* a particular word is associated with a specifiable cluster of properties, knowing *that* a particular set of individuals constitutes the extension of a particular word, knowing *that* a specifiable causal relationship between an utterance of a word and an object or property in the world determines the word's meaning, knowing *that* a particular set of conditions constitutes the state of affairs under which propositions involving this word are true, and so on. But on the meaning as use view, understanding the meaning of a word is more akin to a form of knowing how; knowing *how* to use the word or expression. And any attempt at providing a propositional account of this kind of knowledge would fail for the same reason that any attempt at providing a propositional account of knowledge of tying one's shoes, for example, would fail: the knowledge is practical rather than theoretical and, consequently, cannot be expressed propositionally.

The meaning of a word, then, is its correct application in various circumstances. Since circumstances change, the meanings of words change as well. And understanding this meaning is understanding how to use the word in those various circumstances. Whether or not a particular use is correct in a given

circumstance is determined by the rules that define the language-game within which the context arose.

One well known objection to the meaning as use view is presented by Kripke (1982) in his book *Wittgenstein on Rules and Private Language*. His starting point is the uncontentious observation that predicates that are not coextensive cannot be synonymous. Since it is possible for two predicates that are not coextensive to have precisely the same use, and he uses the infamous 'plus' and 'quus' predicates to illustrate this, it follows that use cannot determine extension. But then, given the assumption, it follows that use cannot constitute meaning either.

More formally,

1. Predicates that are not coextensive cannot be synonymous. (assumption)
2. Determining extension is a necessary condition of meaning. (from 1)
3. It is possible for two predicates that are not co-extensive to have precisely the same use. (from example)
4. Use cannot determine extension. (from 3)
5. Therefore, use cannot constitute meaning. (from 2,4)

Horwich rejects Kripke's reasoning by claiming that the example he uses to establish premiss three above begs the question against the meaning as use view; "namely, that a word whose use is *exactly* the same as the use of our word 'plus' might nonetheless have a different extension." (Horwich, 1995, 364) This, Horwich thinks, is just implausible. To understand his reasoning, we must make a

distinction between complex expressions and primitive terms: an expression is complex relative to a term if the former is defined in terms of the latter. For example, 'rooster' is a complex expression relative to 'chicken' if we define 'rooster' as 'male chicken'. It is possible, Horwich agrees, for the extension of a complex expression to diverge from the extension of the primitive term to which it is related. In our example, the extension of 'rooster' diverges from that of 'chicken'; the former includes only male chickens while the latter includes both male and female chickens. But if the extension of a complex expression diverges from that of the related primitive term, then, Horwich maintains, the use of the former must diverge from the use of the latter as well: "... no complex expression – since it will inevitably bear certain use-relations to its constituents – can have exactly the same use as a primitive expression." (Horwich, 364) If 'quus' is a complex expression, definable in terms of 'plus', then, the counterexample will not go through since the fact that it is a complex expression ensures that it differs in use as well. In order for the example to hold against the meaning as use view, Horwich maintains, "Kripke must assume that the term 'quus' satisfies the following three conditions: (1) that it be a *primitive* term (on a par with 'plus'); (2) that it be co-extensive with one of those possible, complex expressions whose extension diverges very slightly and remotely from that of 'plus'; and (3) that its use be identical to our use of 'plus'." (Horwich, 364-65). But, Horwich argues, 'quus' would acquire such an extension only if i) it were itself a complex expression or ii) it were a primitive term that was "applied in some *definite*

way to certain triples that, given our limitations, are beyond our definite range of application of 'plus'." (Horwich, 365) In either case, Horwich wants to claim, the term is not used in precisely the same way that 'plus' is used.

More formally,

1. The extension of a complex expression may diverge from the extension of its related primitive term.

2. If the extension of a complex expression diverges from the extension of its related primitive term, then their respective uses diverge as well.

3. 'Quus' must then be a primitive term (else its use would diverge from that of 'plus') whose extension diverges slightly from that of 'plus' but whose use is identical to that of 'plus'.

4. But there are only two ways for 'quus' to acquire an extension that diverges slightly from that of 'plus': 1) if it is a complex expression related to 'plus', in which case the conditions in 3 have not been met; or, 2) if it is a primitive term and its use diverges from that of 'plus', in which case the conditions in 3 have not been met.

5. No term meets the requirements set forth in 3.

6. Therefore, Kripke's argument is unsound. (since premiss 3 of Kripke's argument is demonstrably false).

Now this reply seems unsatisfactory. How do we know that the two terms have different uses when they are always used in exactly the same way? Indeed, what can it mean to say that two words that are used in the same way have different uses? Horwich acknowledges that his argument might appear to some as question-begging; he defends the apparent circularity by appealing to other reasons why the meaning as use view is attractive as well as by shifting the burden of proof

back onto Kripke. I am not interested in resolving here the issue of whether Horwich has made a successful reply or not. Rather, I want to highlight a point that Kripke's objection to the meaning as use view brings out nicely; namely, that how we *intend* to use a word and how we *actually* use a word do not always perfectly overlap To see that the meaning as use theory has trouble accommodating this fact, let us look a bit more closely at what exactly Horwich's understanding of that view entails.

On Horwich's view, words have meaning properties; it is the job of our theory of meaning to give a naturalistic explanation of what those meaning properties are. Very roughly, he supposes that meaning properties, that is linguistic semantic properties, have "non-intensional underlying natures". Giving a characterization of these underlying natures will amount to giving a naturalistic explanation of semantic properties. These non-intensional properties, Horwich thinks, are "basic regularities of use, explanatorily fundamental generalizations about the circumstances in which words occur." (Horwich, 356)

Now we can begin to see why the 'quus' example is so telling: it brings out the fact that Horwich's account explains linguistic meaning at the expense of speaker meaning. His account cannot accommodate the 'quus' case because the example makes sense only if one has a way of *characterizing* speaker meaning; 'quus' and 'plus' are plausible predicates to us because we can understand what it is to *intend* to use a term in a certain way, even though we never actually *use* it in that way.

Horwich's account cannot explain what it is for a speaker to intend something by the words she uses. He can explain what the command 'shut the door' means by appealing to regularities of use, perhaps, but he cannot explain, on this account, *why* someone would make the command in the first place. The answer to that question would involve an appeal to the speaker's intentional states: her desire to have the door shut, her belief that shutting the door will keep intruders out, her fear of intruders, and so on. But surely our interest in questions about meaning extend to these waters; we want to know not only what it is for a word to *mean* something, but we also want to know what it is for a speaker to mean something. Indeed, we think that what it is for a speaker to mean something must be connected in some way to the words she uses to express her meaning. Since the meaning as use view does not provide an answer to this question, we need to look to pragmatic theories of meaning, accounts that attempt to explicate this connection. Note that although I am distinguishing between use theories and pragmatic theories here, it is possible for someone to provide a pragmatic account of meaning, one that tries to explain the connection between speaker meaning and linguistic meaning, and still give a meaning as use account of linguistic meaning. I distinguish between the two because one need not do so; one might be an eliminativist about intentional states, for example, and, consequently, see no reason to bring up the issue of intentional states (seeing that, on this view, there are no such states). But, if we think that

eliminativism is implausible, a meaning as use theory is not enough. We turn to pragmatic approaches next.

1.3 Pragmatic Theories of Meaning

As we have just noted, an important turning point in the philosophy of language came about when philosophers began to suspect that the traditional approach to linguistic philosophy (which consisted of investigations of ideal rather than of natural languages) was wrong headed. Strawson, in his 1950 paper "On Referring", hinted at this when he made the distinction between expressions and uses of expressions: sentences or expressions, he argued, are mere tools for making statements; sentences *qua* tools have no meaning, sentences *qua* utterances do. If meaning lies in utterances of words and sentences rather than in the words and sentences themselves, then theories of meaning ought to centre around the question of how words and sentences are used to mean things rather than around the question of how words and sentences function as part of an ideal language. Wittgenstein's advocacy of viewing the problem of meaning as a question of how words are used (and, thereby, his rejection of positivistic approaches to meaning theory), along with Austin's categorizations of the different ways in which speakers can "do things with words", were the first steps towards a pragmatic theory of language.

With this new approach came the realization that there are (at least) two fundamental aspects of meaning that need explaining: 1) how do words mean (and

refer to) things? and; 2) how do we mean (and refer to) things using words? The answer to the first question, many philosophers began to suspect, looks something like this: words mean things in virtue of people using them to mean things. The consequent focus on the *use* of words and on the linguistic practices within communities of speakers, among other things, in order to answer this question, spawned theories of language that we will call pragmatic theories. Answering the second question, how we mean things using words is, of course, also the business of pragmatics. On some accounts, though, a further reduction, to, say, a causal account, is provided on the grounds that without it one does not have a fully naturalistic theory of meaning. We shall explore how pragmatic and causal theories fit together later in this chapter.

Before examining one particular pragmatic approach, Grice's, let us investigate a little further what is entailed by this claim that linguistic meaning reduces to speaker meaning in some way or other.

One, possibly contentious, consequence of this claim is that we cannot understand and explain linguistic meaning *until* we have understood and explained speaker meaning – for to say that linguistic meaning reduces to speaker meaning is just to say that the former is explained in terms of the latter. Thus, the onus is on the pragmatic theorist to explain not only *how* the reduction is to take place, but also *why* the one reduces to the other. The answer to the *how* question will vary, of course, from theory to theory, but the answer to the *why* question serves as the

motivation for pragmatic approaches in general. Let us take a few moments now to investigate *why* some philosophers think that we cannot understand linguistic meaning until we have understood speaker meaning. In order to do this, let us first distinguish those concerns that are specifically semantic from those that are pragmatic.

Semantic Questions

1) How do words refer to things in the world or, perhaps better, how do words hook onto reality? For example, what makes it the case that 'Fred' is Fred's name?
2) What is it to say that a word or sentence *means* something? For example, what do we mean when we say 'Fred is interesting'?

Pragmatic Questions

1) How do people connect their thoughts with the world using words? For example, how do I manage to refer to Fred when I say 'Fred'?
2) What is it to say that a speaker *means* something by her utterances? For example, what do we mean when we say 'Fred *meant* that he was hungry when he uttered 'I am hungry"?

The driving assumption of pragmatic theories is that answers to these and other pragmatic questions will provide us with answers to semantic questions. So, for example, explaining what it is to say that a speaker means something by her utterances will provide us with an explanation of word or sentence meaning. But

why say this? Well, one possible reason, provided by Dennis Stampe, is that semantic relations are subsumed by pragmatic relations:

> The semantic relation is one holding between two terms – a piece of language (a name) and a piece of the world (an object) – to wit, the relationship whereby the former is what the latter is called. This is designation. The pragmatic relation, with which this is identified, is a relationship holding between three terms – between speakers, an object, and a name – to wit, that relationship wherein the first call the second by the third. Assuming, as we may, that when a person calls a thing by its name, he is using the name to do something (perhaps to refer to it, perhaps to describe it), we may regard this three-termed relationship as a genuine pragmatic one. Thus it may appear that the semantic character of a name is nothing other than a certain pragmatic property of it.... (Stampe, 1981, 684)

On this account, the meaning a word has is explained by its being used by speakers to mean something. Thus words have meanings or semantic properties, but, because primacy here is given to pragmatic properties, the semantic properties that words have are, more fundamentally, pragmatic properties. We still do not have an answer to our question, but we now have a way of making our question clearer. What makes pragmatic properties more fundamental than semantic ones? Or, why are semantic properties reduced to pragmatic ones?

There is a reason, of course, for why the identification is made in this way: a naturalistic account of pragmatic relations is far more plausible than a naturalistic account of semantic relations because utterances are actions and actions can be characterized in physicalistic terms while meanings, and this is really the whole point, are elusive. And, ultimately, we (at least some of us) want a naturalistic

account of meaning. We still need an argument, of course, but for now let us leave the matter here and look at Grice's pragmatic account.

Grice's Account

Grice begins by making a distinction between two kinds of meaning: natural and non-natural. In general, if 'x means that p' entails that if x then p, then we have a case of natural meaning; otherwise, we have a case of non-natural meaning. For example, if 'spots mean measles' entails that if spots are present, then measles are present as well, then the usage of 'mean' is natural. Contrast this usage with the following: 'the sounding of the bell means 'the bus is full". On this usage, it is not a logical breach to suppose that while the bell may have sounded, the bus is not yet full[12]; consequently, this is a case of non-natural meaning. Linguistic meaning, Grice supposes, falls into the latter category. We can see this, he suggests, simply by considering some of the linguistic uses of 'meaning' and by contrasting these with other, natural, uses.

Natural meaning involves a "natural" connection between the thing doing the meaning and the thing meant. For example, the connection between spots and measles is a natural one because, roughly, observers or thinkers like us do not have to be aware of or notice this connection for the connection to exist. Non-natural meaning, on the other hand, involves a "non-natural" connection between the thing doing the meaning and the thing meant. The sounding of the bell could not *mean* 'the bus is full' without first some policy or convention establishing that

connection. Thus the *meaning* here is not self-subsistent in the way that natural meaning is.

So much for the distinction. What, in these new terms, does the problem of naturalizing meaning become? Well, since what we are interested in explaining from the beginning is non-natural meaning, it must be 1) explaining what non-natural meaning is; and, 2) explaining how non-natural meaning is a species of natural meaning. In what follows, we will see one approach, Grice's, to doing just this. Note that an important assumption here is that natural meaning is not explanatorily problematic. We will have more to say about this later, in chapter two.

Having established, roughly, the realm of non-natural meaning, Grice makes a further distinction between linguistic meaning and speaker meaning: both, he tells us, are non-natural, although speaker meaning is further reducible to natural meaning.[13] So sequences of letters and words can $mean_{NN}$[14] things – this is linguistic meaning – and sequences of sounds uttered by a speaker can $mean_{NN}$ things – this is speaker meaning. And, as this is a pragmatic theory, linguistic meaning is understood as speaker meaning taken abstractly, out of its context; that is, linguistic meaning is reduced to speaker meaning. Thus, on this view, a certain sequence of words $means_{NN}$ x because, in general, people who use that sequence of words $mean_{NN}$ x. Grice's motivation for making this reduction is familiar. It goes something like this.

There is an unproblematic species of meaning, namely, natural meaning. If we could somehow explain non-natural meaning in terms of natural meaning, we would be well on our way to a naturalistic account of non-natural meaning. Now we can understand two different types of non-natural meaning: speaker meaning and linguistic meaning. The quest for an account of linguistic meaning (in isolation from other considerations) has proven intractable. Trying to understand what a word means in isolation from considerations such as what people mean, or mean to do, by using the word is, many suggest, simply misguided: a word *has* a meaning (and perhaps we see a problem with even talking this way, but for now let us let it go) *only* in so far as people see it as meaningful. To attempt to answer the question of what this word or words mean, or to answer the more general question, what it is to say that a word means something, without considering the role of speakers is to not understand what language *is* fundamentally, namely, a tool for communication. Thus, a naturalistic account of linguistic meaning, on its own, does not seem promising. On the other hand, appealing to the notion of speaker meaning when trying to explain linguistic meaning, and then reducing speaker meaning to natural meaning, seems a likelier route – we just have to consider that speaking is an action and that all actions, at some level, can be explained as natural acts, to see this. If this reduction can be done, then all we are left with is the task of explaining linguistic meaning in terms of speaker meaning. Hence Grice's approach – we can

understand what a sentence means by investigating the relationship of the speaker to her utterance of the sentence.

But what does it mean to say that a speaker *means* something by (uttering) a sentence? After considering some possibilities, Grice lands on this one:

> '*A* meant$_{NN}$ something by *x*' is (roughly) equivalent to '*A* intended the utterance of *x* to produce some effect in an audience by means of the recognition of this intention'; and we may add that to ask what *A* meant is to ask for a specification of the intended effect ... (Grice, 1957, 442)

So 'Fred meant$_{NN}$ *shut the door* by the words 'shut the door" translates into 'Fred intended his utterance to produce some effect in his audience by means of the recognition of his intention, in this case, perhaps, his intention of getting the door shut'. If this is right, then we can explain linguistic meaning in the following way. To say '*x* (where *x* stands for a sentence utterance) meant$_{NN}$ something' is, according to Grice, "(roughly) equivalent to 'somebody meant$_{NN}$ something by x'." (Grice, 1957, 442) And to say 'x means$_{NN}$ that so-and-so' is to say that people in general[15] mean so-and-so by x.

We will talk about how successful this reduction of linguistic meaning to speaker meaning is in a moment, but for now let us look at how the next step, the reduction of speaker meaning to natural meaning, is effected.

Intending to do something, as in intending an utterance to produce some effect, Grice argues, falls under the category of natural meanings, because what a person *intends* to do is, roughly, what a person *means* to do, and what a person

means to do is captured by the natural sense of 'meaning'. Grice seems to take this categorization as uncontentious. He says, "I propose, for convenience, also to include under the head of natural senses of 'mean' such senses of 'mean' as may be exemplified in sentences of the pattern 'A means (meant) *to do* so-and-so (by x),' where A is a human agent." (Grice, 1957, 437)[16] Since we can explain linguistic meaning in terms of speaker meaning, and since we can understand speaker meaning in terms of speakers' intentions, where these intentions are a kind of natural meaning, the naturalistic reduction of linguistic meaning appears to be complete.

Now to the question of how successful this reduction of linguistic meaning to speaker meaning is. Grice's account thus far has provided the following answers to our questions. To the question 'what is it to say that a word or sentence means something?' or 'what is linguistic meaning?' (semantic question 2 above) he answers that (non-natural) linguistic meaning is speaker meaning. To the question 'what is it to say that a speaker means something by her utterances?' or 'what is speaker meaning?' (pragmatic question 2 above) he answers that (non-natural) speaker meaning is a species of natural meaning because speaker meaning is reducible to intentions which are, in turn, a kind of natural meaning. Aside from the problem of treating intentions under the heading of natural meaning, I suspect that, as it stands, this account is circular. Here's how.

On Grice's account, 'S means$_{NN}$ so-and-so by x' translates into 'S intended his utterance of x to produce some specific effect, say y, in his audience via the recognition of his intention'. The problem is this: why does S utter *x* in order to achieve effect y? That is, what is it about *x* that makes S believe that uttering x is the best (or, at least, one) way to achieve *y*? Why is x so useful? The answer is this: S chooses to utter x because of what x *means* and because S believes that his audience understands what x *means* as well. But, according to Grice, 'x means that so-and-so' translates into 'speakers, in general, intend their utterances of x to produce some specific effect, say y, in their audience by the recognition of their intention'. This is circular. Why do speakers, in general, utter *x* in order to produce effect y? They do so because speakers, in general, intend their utterances of x to produce some specific effect y, in their audience by recognition of their intention. This does not explain why uttering *x* is so useful. It is useful, of course, because of what x means and because speakers believe that their audience will understand what x means as well. But this is what we were trying to explain. Explaining linguistic meaning in terms of speaker meaning will not help us, then, because we need an account of linguistic meaning in order to make sense of speaker meaning.

But, perhaps some historical account of how non-natural meaning became a species of meaning might help us out of the circle. Melinda Hogan[17] has made some suggestions in this area. The story might go something like this.

Suppose that certain actions have natural meaning. For example, suppose that humans have an instinctive response to snakes such that whenever they see one, they let out a yelp, say 'yikes'. Yelling 'yikes', then, naturally means something like 'I see a snake' because the relation, if a human yells 'yikes', then he sees a snake, holds. Now the natural connection between seeing a snake and yelling 'yikes' holds for all people. Thus, when you hear me yell 'yikes', you come to believe that I am seeing a snake. I recognize that you make these sorts of inductions and use this knowledge to my advantage: when I want you to believe that I am seeing a snake (perhaps to play a practical joke on you), I reason (correctly) that one good way of bringing this belief in you about is by yelling 'yikes'. Since this action is intentional, I can be said to $mean_{NN}$ by my actions that I am seeing a snake. Now that we have one kind of non-natural meaning, it seems a short step to more. We may develop a convention to replace yelling 'yikes' with uttering 'snake' and, in time, signs such as this one may replace all of our natural signs so that, eventually, a system of non-natural meanings will be formed such that the best way (or, at least, one way) for me to induce the belief in you that I am seeing a snake is by saying 'snake'. If such a story is plausible, then it may serve to ground Grice's account – it would no longer be circular because we would be able to explain why certain utterances are so useful via a naturalistic account of non-natural meanings rather than by appeal to linguistic meaning.

I have no theoretical problems with this general approach, although whether or not a Gricean analysis turns out to be sufficient to tell a whole story is a matter that needs to be investigated (and we will not be doing that here). But even before such an investigation is undertaken, we must address an important Gricean presupposition, namely, that a speaker's intention (meaning to do so-and-so) can be explained in terms of natural meaning.[18] I think, in the end, that this is largely correct, that is, I think that it is possible to give a naturalistic explanation of intention; however, I believe that all of the details of that explanation, given our current understanding of our cognitive processes, are still beyond us. In chapter two I say what I think is wrong with contemporary approaches to naturalizing intentionality and, in chapter three, I give a sketch of what a more promising account might look like. Keeping this thought in mind, let us take a look now at another approach to the question of linguistic and speaker meaning.

1.4 Causal Theories of Meaning

I suggested at the beginning of this chapter that causal theories and pragmatic theories often work together. Here we will explore how this is so, but to get started let us backtrack a bit to another turning point in the development of linguistic philosophy.

In and around the early 1980's, Kripke produced a series of arguments[19] that seemed to show that ordinary proper names like 'Feinman' and natural kind terms

like 'water' do not mean anything at all, contra the Fregean view that identified the meaning of a proper name with its sense; they designate (rigidly on his view) objects in the world. But if names serve to rigidly designate objects in the Kripkean way, they could not be disguised descriptions as Russell thought. Very briefly Kripke's idea is this. If names function to pick out individuals, and we all agree that this is the function of names, and if names are really definite descriptions (or clusters of definite descriptions), then, for each name, the corresponding sense or definite description must be one that necessarily and uniquely applies to that individual. If the definite description did not necessarily and uniquely apply to the individual, then, Kripke maintains, the name would not be designating that individual, or, in other words, it would not be acting as a name of that individual. For example, suppose that the definite descriptions we associate with the name 'Aristotle' are 'the man who studied with Plato for twenty years' and 'the man who founded the Lyceum'. This will not suffice to explain the meaning of the name 'Aristotle', Kripke maintains, because we can imagine a possible world in which the individual we are calling Aristotle did not study with Plato for twenty years (perhaps he studied for nineteen) and did not found the Lyceum; yet, when we use the name 'Aristotle' we still know, in those worlds, that it is *that* man we are referring to – even though he does not answer to the properties we had thought the name was really a description of.

This brings us back to our driving question: how does language hook up with reality? How do names *name* objects? Kripke suggested that names *name* objects via the act of people naming objects. This is the beginning of a baptismal/causal approach to meaning. But now we are in the realm of pragmatics: the meaning of a word, expression, or sentence is what people *use* the word, expression, or sentence to mean. To be sure, a story has to be told about how this works. And any such story, as we have just seen, is an attempt to reduce semantics to pragmatics; to reduce a linguistic theory to a theory of speaker meaning. As we pointed out earlier, however, a pragmatic explanation of linguistic meaning is not, by itself, a naturalistic or non-intentional explanation. In order to get such an account, the pragmatic theory needs to be supplemented with, perhaps, a further reduction. Grice, as we have seen, attempts this further reduction by identifying speaker meaning as a kind of intending and seeing intending as a species of natural meaning. Another method is via a causal theoretic approach. Here the hope is that one can give a naturalistic explanation of linguistic meaning, first, by reducing it to speaker meaning and then, second, by explaining the relationship between speaker's utterances or mental states and objects (or properties) in the world causally. The agenda, then, for such a causal approach is to translate 'S means something by X' into causal terms.

The first step towards such an account is the explication of the relation between speakers' utterances and objects in the world. For example, if we explain,

pragmatically, the meaning of 'cat' as 'what a speaker means by her utterance of 'cat'', then we have to explain 'what a speaker means by her utterance of 'cat'' in terms of the relations between the speaker, the speaker's utterance, and the cat. In other words, we have to tell some sort of causal referential story.

What does a causal theory of reference explain?

1) It explains *what* a sign's reference is; and,

2) It explains *how* a sign is linked to the world.

Crudely put, the causal theoretic account is some version of the following: a sign's, say the sign x, reference is the y (where y is variously taken in different theories as properties or objects) that is causally related to tokenings of x in the requisite way. The link relation, then, between speaker, speaker's utterance, and object is causal.

But this only constitutes a causal account of *reference*. In order to have a complete causal account of speaker meaning, we have to explain not only what it means to say that a speaker is referring to something in the world, but we also have to explain what it means to say that a speaker means something by an utterance. For, as we discussed in section 1.1, if reference exhausts the meaning of a name or description, then if it is true that 'a' and 'b' take the same referent, statements of the form 'a=b' will be of the same kind as statements of the form 'a=a'. But they are not the same; consequently, one might conclude, there must be more to meaning than reference.

So in addition to our causal theory of reference we need to provide a story about meaning. One way of doing this causally is by identifying meaning with the causal structures associated with a word. One can see this as a way (loosely) of explicating the Fregean notion of sense. The reference of a term is that which causes (or originally caused, or in ideal circumstances causes) tokenings of the word – its meaning is the network of causal links between the cause of the tokening, the tokening, and other tokenings of other words. So, for example, in the evening star/morning star case, while the names are caused by the same object, they are caused under different circumstances and, consequently, yield different causal structures. Thus, they mean different things; that is, speakers mean one thing when they utter 'morning star' and another thing when they utter 'evening star'. Another way of explicating the meaning relation, of course, is to maintain that meaning just is reference. But then some answer to Frege's puzzle needs to be provided. In section 3.7 we will look at a suggestion to this end. But now let us take Fodor's account as our paradigm case for discussion to see how a causal approach to explicating the subtleties of meaning might work.

Fodor's theory of content is best understood in terms of its treatment of a problem that plagues all reductive accounts of meaning, namely, the disjunction problem – the issue of determining which cause of a tokening is also its referent while still allowing for the possibility of misrepresentation. The problem gets its name in the following way. Suppose our theory tells us that the referent of a

symbol is just that which causes the symbol's tokening. For example, if cats cause 'cat' tokenings, then the referent of *cat* is cats. But, consider that there are many situations in which things besides cats cause 'cat' tokenings, e.g. small dogs in the dark. This account, however, rather than identifying those tokenings caused by dogs in the dark as erroneous ones, takes the extension of *cat* to be the disjunction 'cats or dogs in the dark'. Hence, the disjunction problem. To avoid it, the task, some think, is to refine the account to identify correct and incorrect representations, that is, to allow for something to be the cause of a tokening and yet not necessarily be the referent. Typically, attempts at solving the disjunction problem, Fodor maintains, have distinguished between the following two types of situations.

Type 1: Symbol tokens are caused by what is in the symbol's extension; thus, all tokenings are correct (if a token is caused by more than one thing, then the disjunction of those things must be a part of the symbol's extension). In this type of situation there is no possibility for error. Think of these situations as ideal situations such that if X causes an 'X' tokening, then *X* just is in the extension of X. For example, Fred Dretske takes Type 1 situations to be learning situations: in learning situations the cause of the tokening *determines* what the content will be. (Dretske, 1981, chapter 8)

Type 2: Symbol tokens can be caused by things which are not a part of the symbol's extension (e.g. 'dog on a dark night' may cause a tokening of

'cat'). In this type of situation, there is the possibility for error. Think of these as sub-ideal situations such that if X causes an 'X' tokening, then X may or may not be in the extension of *X*.

By making this distinction, the hope is that one can explain how content is determined while at the same time allowing for the possibility of erroneous tokenings. For example, we can find out whether a tokening in a Type 2 situation, say a tokening of 'cat' in a dark situation, is erroneous or not by identifying the cause of such a tokening in a Type 1 situation, in this case, the cause of 'cat' tokenings in ideal or learning circumstances, or whatever we take Type 1 situations to be: if the causes are the same, i.e. if cats cause 'cat' tokenings in Type 1 situations and a cat caused the tokening in the Type 2 situation, then the tokening is correct; if the causes are different, e.g. if it was a dog that caused the tokening in the Type 2 situation, there has been a misrepresentation. The task then becomes one of giving a naturalistic distinction between Type 1 and Type 2 situations. Unfortunately, such a distinction has been hard to come by. Those that have been proposed, in particular Dretske's (1981) account and teleological/functional accounts, are rejected by Fodor as not successfully dealing with the problem.[20]

Fodor's unique approach arises out of the consideration that the disjunction problem is really about the difference between meaning and information rather than about error. He explicates this difference in the following way. Symbols

carry meaning while tokenings carry information: "the meaning of a symbol is one of the things that all of its tokens have in common" (Fodor, 1990, 90), while the information that a token carries is determined by how that token was caused. Suppose that the sight of your cat walking into the room prompts you to think 'cat' or 'there is a cat'. Call this tokening S1. Later in the evening you see a small dog (that to you appears to be a cat) and you once again think 'there is a cat'. Call this tokening S2. Now S1, having been caused by a cat, carries information about cats, while S2, having been caused by something other than a cat (a small dog in the dark), does not carry information about cats. On the other hand, both tokenings are about cats, that is, they are both tokens of the symbol 'cat' and all 'cat' tokens mean *cat*. You get the disjunction problem, Fodor maintains, only if you erroneously "identify the meaning of a symbol with the information that its tokens carry" (Fodor, 1990, 90); that is, in this case, if you identify the concept CAT with the various causes of the 'cat' tokenings – e.g. if you take the extension of CAT to be 'cat or 'small dog in the dark' etc.. The problem, Fodor maintains, is that meaning is more robust than these accounts allow for; that is, a symbol's content may not have much, directly, to do with the information that many of its tokenings carry.

Suppose I am in a situation where there are no cats but I carry with me a set of photographs and other mementoes, say collars, toys etc., of my cats at home. Every day I am caused to have 'cat' tokenings by one, some, or all of

these items. None of these 'cat' tokenings carry any information about cats, they carry information about cat photographs, collars, and toys; however, the content of the symbol that these are tokenings of is *cat*. Thus, meaning here is not directly related to information. Fodor maintains that once we have a theory of meaning that captures this "robust" character of meaning, once we have put information and meaning in their proper places and determined the relationship they have to one another, the disjunction problem simply will not arise. Now the key to capturing the robustness of meaning, Fodor proposes, is the concept of asymmetric dependence. In a nutshell, it is the asymmetric dependence relations that exist between information-carrying causal relations that determines content. But to understand what these relations are and how they are set up, the story needs to be told from the beginning, not at the level of mental states, but rather at the level of language and our linguistic practices.

Fodor begins by noting the hierarchical nature of our linguistic policies – some practices are parasitic upon others while some are fundamental. For example, the practice of paging (for example, yelling 'Hey, John!' when trying to get John's attention), Fodor points out, is parasitic on the practice of naming: but for the pre-established policy of naming individuals, we could not have a policy of paging individuals. So one of the things that makes it the case that when someone calls out 'John, you're being paged', it will be John who responds to the call, is the fact that John is named 'John'. Similarly, policies about requests,

desires, needs, etc. are parasitic upon naming policies; my request for an item is possible only once a policy for naming that item, for referring to it, is already in place. So far so good, but how is it that 'John' comes to be the name of John? 'John' cannot refer to John just because we intend for 'John' to be the name of John; that is, it cannot just be because of something about us, our mental states etc., that makes it the case that 'John' is the name of John; there has to be something about John as well, that makes it the case that John's name is 'John':

> For 'John' to be John's name, there must be some sort of *real relation* between the name and its bearer ... this is because, of course, intentions are (merely) intentional; you can intend that there be a certain relation between 'John' and John and yet there may be no such relation. (Fodor, 1990, 98)

These real, causal, relations get formed, Fodor supposes, in the following way:

> My pursuing these policies is my being in a certain complex mental state, and my being in that mental state has certain causal consequences: in particular it has the consequence that there is a certain pattern of causal relations between [for example] slabs and my tokenings of 'is a slab'.... (Fodor, 1990, 99)

So information gets into the picture via the causal relation. And, one plausible way to explain our *semantic* information-carrying causal relations is as consequences of the mental states to which pursuit of our linguistic policies gives rise.

But what about meaning? We have already noted how some linguistic policies are more fundamental than others; that is, that the policy of paging, for

example, could not be in place but for the policy of naming. Causally this situation translates into the following: our naming tokenings of 'John' are directly causally related to John while our paging tokenings of 'John' are causally related to John via the causal relation that exists between John and the name 'John'. The paging policy is asymmetrically dependent upon the naming one because were there no naming policy, there would be no paging policy; on the other hand, the naming policy could exist without the paging one. Or, in causal terms, the causal relation between paging John and 'John' tokenings is asymmetrically dependent upon the causal relation between John and 'John' tokenings. The former tokenings carry information about, say, my wanting to speak to John, while the latter carries information about John; but, both tokenings are *about* John. We know they are both *about* John because no 'John' tokenings could occur but for there being 'John' tokenings caused by John. It is in this way that asymmetric dependence determines meaning.

We can now tell a naturalistic story about how content gets determined. In order for a symbol, say 'x', to mean *x*, it must be the case 1) that some 'x' tokens are caused by x's; and, 2) that those 'x' tokens which are not caused by x's be asymmetrically dependent upon the former. To take our example again, both S1 and S2 mean *cat*, but S2 could not mean *cat* unless there were already tokenings of the S1 variety (i.e. those actually caused by cats). The dependence is asymmetric because S1 tokenings would mean *cat* whether or not there were

tokenings of type S2. Finally, asymmetric dependences hold between properties, not individuals (to allow the theory to deal with uninstantiated individuals such as unicorns). So, for example, 'cat' means *cat* just in case there is a nomic relation between the property *being a cat* and the property *being a cause of 'cat' tokens*; and, if there is a nomic relation between, say, the property *being a small dog in the dark* and the property *being a cause of 'cat' tokens*, this relation must be asymmetrically dependent upon the former. (Fodor, 1990, 93) Note that what makes this account a naturalistic one is the reduction of semantic relations to causal ones. The fact that these causal relations are set up as a result of our linguistic policies need play no role in the theory because, in explaining how content is determined, we are ultimately interested in identifying semantical relations and these, according to Fodor, just *are* causal relations. As Fodor states, " it's the linguistic policies of speakers that give rise to the asymmetric causal dependences in terms of which the conditions for robustness are defined; but the conditions for robustness *quantify over* the mediating mechanisms, and so can be stated without referring to the policies... ." (Fodor, 1990, 100)

To summarize, Fodor's account reduces semantic relations to pragmatic relations, i.e. our linguistic policies. And these linguistic policies, in turn, set up causal relations. So, our initial question is answered in the following way. What 'cat' means is reduced to the problem of what a speaker means by her utterance of 'cat'. A speaker means *cat* and refers to cats by her utterance 'cat' because 1)

cats cause speakers to utter (or think) 'cat'; and, 2) nothing else would cause speakers to utter (or think) 'cat' unless cats did.

Now as promising as the asymmetric dependence thesis appears to be (and I shall say what I think is right about it in section 3.6), my contention is that causal accounts such as Fodor's succeed in being naturalistic only because of some starting assumptions. In the next chapter I shall uncover these assumptions and argue that they are unfounded.

Chapter Two ••• Problems

Introduction

Pragmatic theories of meaning elaborate on an intuition elicited by the thought experiments of Saul Kripke (1972), Hilary Putnam (1975), and Tyler Burge (1979). The intuition is that the meanings of words and sentences and the contents of thoughts are fixed externally by social procedures and by the surrounding environment, both social and natural. Thus, one of the morals we are supposed to take from the various thought experiments dealing with these issues is that meaning has more to do with the stuff out there than it does with the stuff inside our heads. This is demonstrated by the twin earth intuition[1] that, even though someone has no idea what the essential property[2] of an object may be, nevertheless, the presence or absence of that essential property is what ultimately determines whether or not she is successful in referring to it. I will begin this chapter by explicating these thought experiments and their intended conclusions so that we may get a better understanding of the way in which the external environment plays a role in meaning.

On Putnam's account we are asked to imagine a possible world which is identical to Earth in every respect with the exception that what Twin Earthlings call 'water' has a different molecular structure, call it XYZ, from the H_2O stuff on Earth; otherwise, Twin Earth water is indistinguishable from Earth water. Now

consider two individuals, Oscar and Twin-Oscar, who are identical to one another in every respect. Both Oscars have certain attitudes toward water and $water_{TE}$ respectively: they believe it to be thirst quenching, think it necessary for survival, can recognize it when they see some, and so on. Neither of the Oscars is aware, however, of the chemical properties of water and $water_{TE}$. Putnam would have us believe that, in spite of the identity of their psychological states, narrowly construed, whenever the Oscars have *water* /$water_{TE}$ thoughts, their mental contents differ; for Oscar has H_2O in his environment, while Twin-Oscar has XYZ in *his* environment. The distinction between mental contents narrowly or broadly construed is just the distinction between mental contents individually or relationally construed; on the former, mental contents are individuated within the head as it were, while, on the latter, mental contents are individuated according to external factors. Thus, in this case, the Oscars share their mental contents if we are taking mental contents narrowly, but they fail to share their mental contents if we individuate them relationally (since they are in different relational states).[3]

The intuition Putnam supposes this case evokes is this: if Oscar and Twin-Oscar are in the same narrow psychological states but, nevertheless, have beliefs about different things, then narrow psychological states cannot determine extensions. Further, since meaning (intension) determines extension, then narrow psychological states must not determine intensions either. In short,

Putnam's story has broken the connection between meaning and mental contents: meaning, it seems, is determined by the physical environment and the linguistic community, not by the individual as was previously thought.

Burge's story differs from Putnam's in an important way: both individuals are physiologically identical but, in this case, the object of their propositional attitudes is the same as well (unlike in the Twin Earth story where the respective waters were actually different substances). The story goes like this. Imagine an individual, call him Fred, who holds certain beliefs about arthritis: he believes that the arthritis in his ankles is painful, that arthritis can spread, that his father also had arthritis, and so on. In addition to these attitudes toward arthritis, he believes that the ailment in his thigh is arthritis as well. Now imagine a counterfactual Fred who is identical to Fred in every detail: he has the same neurological/physiological makeup as Fred, he has had all of the experiences that Fred has had, he behaves in the same ways as Fred does (when his behaviour is non-intentionally described, of course), and so on. They differ in only this respect: in Fred's linguistic community, the concept ARTHRITIS refers only to inflammations in joints while in Counterfactual-Fred's linguistic community, ARTHRITIS also includes certain rheumatoid conditions. Thus, when Fred goes to the doctor and expresses his alarm about the arthritis in his thigh, the doctor explains that what he has in his thigh cannot be arthritis,

because arthritis occurs only in joints. Counterfactual-Fred's fears, however, are confirmed when he visits *his* doctor.

Now Burge wants us to share two intuitions: 1) Fred understands and uses the concept ARTHRITIS in spite of the fact that he was mistaken about its extension – Burge supposes that he has an incomplete understanding of it; and, 2) Counterfactual-Fred, despite the fact that he is neurologically etc. identical to Fred, has no *arthritis*-attitudes: we may call what he has *arthritis*$_2$-attitudes as long as we remember that, despite the similarities to *arthritis*-attitudes, they have nothing at all to do with arthritis.

Here again, then, is a case where there is an identity of brain states but, in virtue of some relational difference, the mental contents of the individuals are differentiated.

The two scenarios can be summarized as follows:

Putnam

1) Oscar believes that water quenches thirst.
2) Twin-Oscar believes that water$_{TE}$ quenches thirst.

Both Oscars have the same beliefs, narrowly construed, about water and water$_{TE}$, but the contents of the corresponding propositions differ because the extension of water is not identical with the extension of water$_{TE}$ (indeed, they do not even overlap). In other words, it is possible for one proposition to be true

while the other is false; their propositional attitudes, consequently, must differ, in which case, of course, their beliefs, widely construed, must differ as well.

Burge

1) Fred believes that the disease in his thigh is arthritis.
2) Counterfactual-Fred believes that the disease in his thigh is arthritis$_2$.

In this case, the individuals have different beliefs about the same thing, viz. the condition in the thigh. The extension of 'arthritis' differs for Fred and Counterfactual-Fred; thus, it is possible for the proposition believed by Counterfactual-Fred to be true while the other proposition – the proposition believed by Fred – to be false (as in the example by Burge). Since they do not have the attitude of belief toward the same proposition, their propositional attitudes must differ as well.

Thus, Burge wants us to conclude, meaning or content is determined by external factors – we need to look to the social community of an individual in order to understand what he means. Putnam's twin-earth case elicits a similar conclusion except, in that case, the determining factor of the contents is not the social environment, but rather the physical environment – water is *that* stuff, the stuff in my physical environment, and *that's* what I mean when I say 'water'.

Enter Kripke and his thesis of rigid designation. What counts as being like the objects in my environment is, according to this view, having the right property or set of properties. Thus, even though I may not know that water=H_2O, if I

encounter a substance that appears to me exactly like water but it is not H_2O, then, while I may think that I am referring to it when I say 'water', I will not be (since what I *really* mean by 'water' is the stuff around my home and *that* stuff has the property of being H_2O).

Now, as with most accounts, there is something right about the conclusions drawn above and something wrong too; while it seems right to suppose that the contents of our mental states and the meanings of the words we use are related in some essential way to our social and physical surroundings, we should not ignore the role of our internal cognitive processes in fixing content and meaning. In chapter three, I will focus on providing a description of a new approach to understanding these questions of meaning, one that takes into account both its external and its internal aspects. In my discussion, I will also address the impulse to conclude, from Kripke's considerations, that rigid designation is the correct account of name-usage and that, consequently, our metaphysics had better include an account of essential properties as well. I shall argue, by providing a metaphysical picture devoid of essential properties that still allows for the possibility of names being used rigidly, that we need not draw these metaphysical conclusions from Kripke's thought experiments.

But before we do this, we need to iron out another fundamental problem. We hinted at it in our discussion of Grice and the obstacles his and other such

accounts face. The problem is this: if words mean things in virtue of people using them to mean things, how do we explain, in particular cases, why, for example, Fred chose those words and not others to convey what he meant? In other words, must we not have an account of word meaning in order to explain word usage? Otherwise we will not be able to explain why Fred uttered the words 'please close the door' when he wanted the door to be closed: Fred may intend all he wants, but until he finds a way of communicating his intentions, his intentions will not be made clear. And the only way he can make his intentions clear is via some mechanism that is capable of communicating those intentions. Hence the problem of explaining literal meaning, in this case, explaining how the words 'please close the door' are capable of communicating the intention to get the door closed. As we saw at the end of the last section, one way of filling in this gap is by providing a causal account that will give a naturalistic explanation of reference. These accounts solve the problem of literal meaning by focusing on explaining how thoughts are connected to the objects they are about and consequently, how individual terms are connected to the objects to which they can be used to refer. In this chapter, I shall explore why such accounts fail in their naturalism. In the course of this investigation, I shall examine the metaphysical underpinnings of causal accounts, show why they are unsatisfactory, and propose a new picture. We begin by investigating the causal theorist's account of perception.

2.1 Why An Account of Perception is Important

We might ask at this point, why begin with perception? Are we not talking about language and meaning here? Well, consider what the problem of meaning *is*. It is the problem of explaining how we *mean* things – how the things we say and think sign beyond themselves. In other words, it is the problem of explaining how our thoughts and words hook up with the things or ideas they are about. The causal theorist wants to say that the reference relation just is a certain causal relation, that our thoughts and words are about those things they are causally related to in the requisite way. Take, for example, the causal account we looked at in chapter one. It made use of an account of perception in the following way: in order for some property, say the property *being a cat*, to be causally related to, say, the property *being a cause of 'cat' tokens*, it must first be perceptually related to the utterer of 'cat' tokens. The causal relationship between things in the world and our utterances or thoughts is dependent upon our perception of those things in the world; the property *being a cat* cannot be nomically related to the property *being a cause of 'cat' tokens*, if the thing in which cat tokens are manifested cannot perceive cats. This is obvious but not as straightforward as causal theorists would have us believe. In sections 2.2-2.6 I shall argue that it is precisely at this point that the causal theorist unwarrantedly slips in the idea that his account is naturalistic.

To get the argument going, we need to recognize that the causal theorist needs some account of perception to tell a whole story. But not just any account of perception will do for naturalistic causal theories. There is a particular account that underlies the theories in general.[4] I refer here to the distinction Dretske draws between epistemic and non-epistemic seeing, that is, between seeing that entails a belief or set of beliefs held by the perceiver and seeing that does not. For the most part, if they get mentioned at all, the controversial theses of this account are typically merely stated rather than justified. Stampe, for example, makes almost explicit use of the notion of non-epistemic seeing that Dretske works out when he says,

> A seer's knowledge, or ignorance, of what he is seeing has nothing whatever to do with the matter. He needn't know that he is seeing a cat, in order to see one, obviously – any more than he needs to know that what he is stepping on is a cat in order to step on one. (Stampe, 1981, 690)

Unfortunate example aside, Stampe is appealing to the same account of perception and underlying ontology as Dretske is when he says, "S's stepping on a bug is similar in this respect. One can step on a bug without believing that one stepped on it ... ." (Dretske, 1969, 5) To give some account of perception is important, as we just noted, but just as important for the naturalistic agenda of these causal theorists is the specific thesis that there is such a thing as non-intentional seeing (and this is what non-epistemic seeing is taken to be). This step is crucial because, if we recall chapter one, one of the driving motivations for the

approach is the belief that, ultimately, we can give a naturalistic, that is, non-intentional, account of meaning. But the causal story will be naturalistic only if it can be shown that the ultimate link in the chain of causal connections between intentional states and their objects is a purely causal link and is thus non-intentional. Since this ultimate link comes in the form of perception – we are caused by some object in the world to enter into a certain intentional state only once we have perceived (consciously or unconsciously) it – then the naturalistic success of the causal story rests on the naturalistic success of an account of perception. Generally when causal accounts of meaning are put forth, this underlying account of non-intentional perception is assumed. Since Dretske's account is so neatly worked out and is, further, cited favourably by other causal theorists, we will deal with it as the paradigm account of perception underlying all causal theories. If it is not plausible, as I shall argue, then the naturalism of any causal theory that relies on it is in jeopardy.

2.2 Dretske's Account of Non-epistemic Perception

Dretske begins (Dretske, 1969, 4-6) by making a distinction between epistemic and non-epistemic seeing (seeing and seeing$_N$), that is, between seeing that entails a specific belief or set of beliefs on the part of the perceiver and seeing that does not. The non-epistemic variety is "a primitive visual ability which is common to a great variety of sentient beings, an ability which we, as human beings,

share with our cocker spaniel and pet cat." (Dretske, 1969, 4) Furthermore, only those acts in which a discrimination is made between the object seen and its background will count as acts of seeing. Thus, if you are looking at a feature of a landscape, say a tree, that is indistinguishable by you from the rest of the landscape, you cannot be said to have seen the tree (as in cases, for example, when one is viewing the tree from very far away).

The feature that distinguishes seeing from seeing$_N$ is belief-content: seeing is a state of affairs that has positive belief content while seeing$_N$ is a state of affairs that involves zero belief content. A situation has positive belief content if a statement expressing the state of affairs, say, "'S ...' [e.g. 'S sees the table', 'S is eating ice cream'] entails that S has a particular belief, or set of beliefs." (Dretske, 1969, 5) A situation has negative belief content if a statement expressing the state of affairs "entails that S does not have some belief, or set of beliefs." (Dretske, 1969, 5) Finally, a situation has zero belief content if there is no belief that S might have or not have that is logically relevant to the truth of the statement describing the situation. So, for example, the statement 'S has blonde hair' is made true by the state of affair of S's having blonde hair. It is logically irrelevant to the truth of this statement whether or not S believes anything at all; thus, S having blond hair is an example of a state of affairs that has zero belief content. On the other hand, the state of wishing for a bowl of ice cream is a situation with positive belief content because if S wishes for ice cream, then S must

also believe that she does not now have ice cream. Another way to say this is to note that the statement 'S wishes for ice cream' entails that S believe something like 'there is no ice cream here'; thus, this state of affairs has positive belief content. With respect to seeing, then, any statement of the form 'S sees D' entails that S has a specific belief, while any statement of the form 'S $sees_N$ D' does not entail specific belief attributions. In order to make his case for the existence of $seeing_N$, Dretske has to show that "there is a way of seeing such that for any proposition, P, the statement 'S $sees_N$ D' does not logically entail the statement 'S believes P'." (Dretske, 1969, 6) For example, he has to show that no specific belief attributions are entailed by a statement such as 'S $sees_N$ the table'.

One of the considerations motivating Dretske's distinction between epistemic and non-epistemic seeing is the thought that just as non-sentient beings and inanimate objects can discriminate visually and/or in other ways[5] and can even represent bits of the world, so too humans have both an epistemic and a non-epistemic ability to do so. In spirit, this distinction between epistemic and non-epistemic perceptions seems to be roughly in line with a cognitive/non-cognitive distinction – noticing that some activities involve cognition at some level and that others do not. For example, the activity of judging that an apple is red seems to be a cognitive activity while breathing, perhaps, is not. What makes it the case that the former act involves cognition while the latter one does not? Well, the former act involves thinking *about* an aspect of the world while the latter act does not. Now of

course there are different ways to understand cognition, and these differences will affect what sorts of things get counted as cognitive acts, but there are at least some clear black and white areas. Believing or judging something to be the case, for example, is, on any understanding of cognition, a cognitive act. Fodor takes the following as a standard working definition of "cognitive level [of the brain]": "... any level at which states of the system are taken to encode properties of the world counts as a *cognitive* level; and no other levels do." (Fodor, 1997, 314) Since we are largely interested here in the underpinnings of accounts such as Fodor's, we will stick with this definition of 'cognitive level'. Following from this definition, then, we say that something is an act of cognition if it involves a state of a system that encodes properties of the world. All cognitive acts are intentional acts, then, since 'encoding properties of the world' is just a fancy way of saying that the state in question is about the world. And, since all intentional acts involve a state of being about the world, they are also all cognitive acts.

Now I said that the spirit of Dretske's making a distinction between epistemic and non-epistemic seeing is that of making a distinction between seeing that entails cognition and seeing that does not. This is so because, as causal theorists suppose, it is via the route of non-epistemic seeing (or perception in general) that we are going to get the first step in a non-intentional account of mental representation. Obviously, the jump from being about something (that is, being an intentional state) to not is going to have to come in somewhere. The suggestion is

that it is going to be at this level because it is this kind of primitive or brute representation that people like Dretske and Fodor want to take as unproblematic instances of representation, unproblematic in the sense that they are not examples of full blown intentionality. To be sure, this distinction is a fuzzy one, but as long as we think it makes sense to make a distinction between the kind of seeing (and consequently brute representation) going on when a frog sees a fly and the kind of seeing (and consequently sophisticated representation) going on when a human sees her cat (as her cat) then we can table the grey areas for the time being. Or so the hope of the naturalist goes.

All epistemic seeing acts, then, are also cognitive acts; since they entail specific beliefs, there must also be a *thinking about* going on. Non-epistemic seeing acts, on the other hand, as they do not entail specific beliefs, are possibly non-cognitive[6] and, consequently, are also the kinds of acts that can be performed by organisms or objects that are not capable of having beliefs. For example, one might say that a thermometer that has discriminated between temperatures has performed a non-cognitive act. Dretske supposes that humans exhibit similarly non-cognitive behaviour when non-epistemically seeing, when our ability merely to see a thing results from our visual biological capabilities and has nothing to do with our reasoning processes or our beliefs. Thus, thanks to my non-cognitive ability to see things, even though I may never have seen a computer before, it is still possible for me to see_N it. As soon as I see it *as* a computer, however, I am

engaged in a cognitive act because in order to do this I must believe certain things (for example, I must believe that there are computers, that computers look like this, and so on); thus, I will be epistemically seeing the computer.

I should make clear before continuing that the rejection of the distinction Dretske makes is not the claim that there is no distinction to be made: Dretske is right that there is something which corresponds to his epistemic versus non-epistemic seeing; we *can* classify visual perceptions into these different levels. And for many discussions, noting this distinction is very useful and important. My contention, however, is that Dretske, or anyone who wishes to naturalize their account by building it on top of this one, is wrong about what this distinction entails. It does not entail, as it is commonly supposed to entail, that 1) there is a non-intentional way in which we gather information about the world,[7] and that 2) consequently, there is an objective way to view the world. The way I have said this is a bit misleading because it makes it sound as though someone might conclude from the fact that there is a non-intentional way of seeing the world that there is an objective way of viewing the world when really, I suspect, it goes the other way around; one claims that there is a non-intentional way of seeing the world because one is convinced of and is interested in preserving the objective status of at least some of our observation claims. The conviction in the objective status of our observations comes, I think, from recognizing that this is the simplest explanation of the fact that, in general, we all make the same observations. The interest in

preserving the objective status of our observation claims stems, at least in part, from a worry that observation claims that do not have this status cannot serve as the basis for scientific theory. Now if all visual perception is intentional, the objective status of our observation claims is in jeopardy, one might suppose, because then *all* visual experiences depend, at least in part, upon how each individual thinks about the world. So if we think there is an objective way to see the world, then we might also think that there is a non-intentional way to see the world. At any rate, something like this train of thought is motivating Dretske's approach. He says this about his project:

> One can freely acknowledge the relativity inherent in *other* ways of seeing, relativity which depends on the conceptual background, past experience, and modes of association of the individual percipient, *without undermining the objectivity and publicity of what we see*. For the objectivity and publicity of this world resides in the fact that we can all, *regardless* of our conceptual background, associative talents, inferential skill, or past experience, see_N the same objects and events. (Dretske, 1969, 77)

In what follows I propose to agree with this and yet still maintain that $seeing_N$ is an intentional activity by arguing that while Dretske's desire for an account that does not undermine the objectivity of what we see is misguided (where 'objective' has the sense of 'the way things really are'), his desire for an account that does not undermine the publicity of what we see is appropriate and needs to be addressed.

2.3 Why Dretske's View is Wrong

My contention is that this distinction between seeing and seeing$_N$ will not work in the way in which Dretske wants it to work because, as I shall argue in this section, the difference between seeing and seeing$_N$ is not wide enough to support the weight of the naturalistic programme. My position is motivated by two general thoughts.

First, treating notions such as discrimination and representation on a continuum is just misguided: thermometers simply do not discriminate; and supposing that they do and that, consequently, humans have a similarly simple level at which we discriminate, is to have gone wildly off track. Of course, the very basic discriminations we make, for example when we discriminate between an animal and its surroundings, *seem* to be non-cognitive because, if one is a human being, one cannot help but make them; however, involuntariness is not enough to establish that a discrimination is non-cognitive. As I shall argue, the very idea of non-cognitive discriminating or representing is incoherent.

Second, I am also sceptical of a distinction between a kind of seeing and a kind of seeing-as when it is coupled with a requirement for what is going to count as any kind of seeing, namely, that a discrimination between the thing seen and its background environment must be made before an individual can be said to have seen the object in question. I will take a moment here to explain what I think is so wrong with this.

If we see a landscape from very far away, say from a plane, then even though our eyes might be directly trained on, say, a tree and are consequently receiving some kind of visual information from the tree, since we are in no position to discriminate between the tree and the grass, we cannot be said to have seen the tree. This seems reasonable enough. But now suppose that we *do* discriminate between the tree and the background (say we are closer to the tree) but do not see it as a tree. This would correspond to a kind of seeing$_N$ on Dretske's account. What makes it different from epistemic seeing? It is that there is nothing we need to believe in order to see it. But what *kind* of beliefs is it not necessary to have? The kind of beliefs Dretske means here are beliefs that are available to individuals in consciousness but are not necessarily in their consciousness at the time. These beliefs, not to be confused with ones that are being distinguished from *unconscious* beliefs, are those beliefs that an individual could revise, give up, etc. simply by reflecting on them; they are distinct from those beliefs that an individual cannot *access* just by thinking (even under hypnosis). The latter sort are typically not called beliefs. I will clarify what these are as we go along, but for now I will speak of them as 'not-conscious' beliefs and refer to the other sort, the kind we typically just call 'beliefs', as 'conscious beliefs'.

The thrust of my point here is that the fact that an act of seeing does not entail any conscious beliefs on the part of the perceiver does not preclude it entailing not-conscious beliefs on the part of the perceiver. But if the causal

theorist is attempting to give a non-intentional account of intentionality by appealing to our causal relationship with the world, then the causal theorist should be worried not just about showing that no conscious beliefs influence causal relations, she should be interested in showing that nothing that functions to represent the world has an influence on these causal relations, that no conscious or not-conscious beliefs influence causal relations; otherwise, there is no impetus for the supposition that causal relations can serve as the non-intentional relation upon which our semantic notions are founded.

Of course, causal theorists never claim that *every* causal relation is non-intentional; obviously some of the intentional states we enter into are caused by other intentional states which were caused by other intentional states and so on. But the naturalistic hope of the theorists I am speaking of here is set on the idea that this chain can ultimately be *grounded* in a non-intentional relation with the world, namely, non-intentional perception. The following characterization by Stampe of the perceptual relation illustrates this motivation plainly,

> The things that it [a sentient organism] perceives are just those objects or states of affairs that cause these perceptual states to arise within it. Jim sees a cat: that is, some cat causes Jim to be in a certain perceptual – visual – state. ... Notice that it is a certain fact about the object itself – the cat – that determines that *it* is the thing being seen; it is the fact, namely, that it is the thing that caused the visual representation. (Stampe, 1981, 689-690)

And Dretske makes his appeal to the non-intentionality of perception even more explicit: "if perception is understood as a creature's *experience* of his surroundings, then, perception itself is cognitively neutral." (Dretske, 1981, 153)

In what follows, we will concede to Dretske that, given his narrow understanding of belief, it is true that we can make a distinction between epistemic and non-epistemic seeing. But this, as I shall show, is a difference that makes no difference to what the causal theorist ultimately wants to claim, since non-epistemic seeing does not constitute non-intentional seeing.

Non-Epistemic Seeing is Not Non-Intentional

What seems to be underlying Dretske's conviction that non-epistemic seeing (that is non-intentional seeing) is possible is a certain view of the human eye. Dretske supposes that eyes are discriminators, that it is the function of the eye to discriminate between the objects in our visual environment. My contention is that the eye is not a discriminator and that supposing that it is, is to conflate the job of receiving information with the job of perception; the information that our eyes are capable of receiving and do receive, on the one hand, and what we see, on the other hand, by Dretske's own criterion, do not completely overlap. Dretske's own examples support his conclusion, that the eye itself is capable of discrimination, because in all of them, if S does not end up seeing P at all, even non-epistemically, it is because S was in no *position* to discriminate between P and its background; that is, S was in a poor viewing situation. Here is one of those examples.

> ... take nine cubes and place them in the form of a square .... dim the lights, retreat to a sufficiently great distance, so that the ensemble of blocks appears to one as a uniform mass without distinguishable parts (i.e. as a square). Can one, under these conditions, see cube #5? Although cube #5 makes a positive contribution to the way the 'square' looks, in the sense that without it the square might appear to have a hole in the center, and in the sense that the light from #5 is stimulating your visual receptors, I do not think we would go so far as to say that one could see cube #5. ... It must, if indeed it is a distinct part of what one sees, appear to one as a more or less differentiated part. (Dretske, 1969, 24)

Now cases such as these show us nothing about the difference between what information our eyes receive and what we see because these are cases in which there *is* no apparent difference between the two. But now let us look at an example in which there is such a difference, one in which Dretske would say that sight does not occur.

Suppose I am looking at a white wall. On this wall, in the top right corner, there is a patch of beige. Suppose that I do not notice this patch of beige even though my eyes are trained directly on the patch. According to Dretske's account, even though my eye is receiving beige patch information, because I have not differentiated the beige patch from the rest of the wall, I cannot be said to have seen it: "The term 'differentiated' is meant to describe a particular way in which D must look to S in order for it to be true that S $sees_n$ S." (Dretske, 1969, 24) I agree. But if this is the case, then the information that our eyes receive does not necessarily correspond to what we end up seeing. That my eye receives

information from the beige patch is certain because, given a bit of prompting from someone beside me of the 'hey, look at that beige patch'—variety, I will suddenly "notice" the beige patch, that is, see it. But since nothing would have changed with respect to the information my eye was receiving (I did not move my head or close my eyes), then my seeing the beige patch must depend upon something besides the information my eyes received. David Hubel puts this point in the following way:

> The receptors, called rods and cones, are nerve cells specialized to emit electrical signals when light hits them. The task of the rest of the retina and of the brain proper is to make sense of these signals, to extract information that is biologically useful to us. The result is the scene as we perceive it .... (Hubel, 1988, 3)

What we perceive, then, and what information we receive are two different things. Where does discriminating come in? Well, it cannot come in at the level of simple information receiving since, if it did, there would be no way of explaining how, in cases such as the one I just described, the information one receives need not change in order for one to make different discriminations. But at the level of the retina, only information receiving is going on. Thus, discrimination must come in at the next level, at the level of the brain, and not at the level of the eye. But we know that at the level of the brain, with respect to visual perception at least, the neural networks that constitute the visual pathway are recurrent networks; that is, information is travelling from cells deep in the brain up to cells closer to the retinal stage as well as from the retinal stage to the cells deep in the brain. But then various aspects of the brain influence the visual information received through the

cells in the retina *before* we ever end up seeing anything; that is, before, the perception becomes available in consciousness. And increasingly the evidence indicates that the influence the brain exerts on this information is cognitive in nature; that is, it involves encoding of properties of the world. If so, then the resulting perception is an intentional act. Before we continue with some more cases to make this point clearer, since so much of our discussion and of Dretske's centres around the makeup of the human eye, it will help to present a picture of how vision might work.

The eye is a complex sensory organ, but seeing is an even more complex sensory activity. Visual perception does not only involve the eye proper, that is, the cells in the retina and the optic nerve; it also involves, at least, the cells in the lateral geniculate body and the cells in the striate or primary visual cortex. At each of these stages, that is, at the retinal stage, at the geniculate stage, and at the cortical stage, there are sometimes many substages of cells through which information gets passed. The retinal stage, for example, contains receptor cells. These are the rods and the cones that are the initial responders to external visual stimuli, as well as two other stages of cells that contain cells such as bipolar cells, horizontal cells, and amacrine cells; the latter are cells that respond to the information received from the receptors and then pass their information along deep in the brain to the retinal ganglion cells. The retinal ganglion cells then send the information further into the brain to the cells in the primary visual cortex, found in the cerebral cortex, where the

information gets passed through many different layers of cells once again and finally on to other visual and non-visual areas in the cerebral cortex. Cell connection is never a one-way street: cells having a similar function generally reside in the same area of the brain and are interconnected via their axons. Thus any given cell will be receiving information from cells within its own stage and from a previous stage of cells as well as from cells deeper in the brain.

The upshot of these observations for our purposes is that the act of actually seeing something in our environment is an act that results from a highly complex set of procedures involving far more than just the cells in our eyes. Hubel puts this point nicely as follows:

> The visual world is thus systematically mapped onto the geniculate and cortex. What was not at all clear in the 1950s was what the mapping might mean. In those days it was not obvious that the brain operates on the information it receives, transforming it in such a way as to make it more useful. People had the feeling that the visual scene had made it to the brain; now the problem for the brain was to make sense of it – or perhaps it was not the brain's problem, but the mind's. The message of the next chapters will be that a structure such as the primary visual cortex does exert profound transformations on the information that it receives. (Hubel, 1988, 61)

But how do we understand these 'profound transformations'? Well, think about familiar aspects of our environment such as the knobs on a drawer, features on a face, pictures on a wall. Our eyes are set up to distinguish between dark and light contrasts and thus easily pick out edges and borders. But not all of the edges and borders in our *visual* environment correspond neatly to the edges and borders

of the *objects* in our environment. A chest of drawers, for example, is itself made up of many edges – the knobs, the drawers, and so on. Our eyes will receive information about all of these aspects of the chest of drawers, not privileging at this stage a subset of the information. If the chest of drawers stands beside a bed, then there are edges between the chest of drawers and the bed, between the bed and the wall, and so on, to contend with. We, of course, see the edges between the bed and the chest of drawers as different from the edges that result from imperfections in the wood etc. because we see the edges between the bed and the chest of drawers as object-defining edges. But *this* information is not passed on from the world to our eyes. Our eyes do not, indeed cannot, discriminate between edges that are object defining and edges that are not precisely because there is nothing intrinsic to an edge that makes it object defining. Thus, if we want to get a map of all of the objects in our environment (not specify what they are, just get a picture of them) we should not look to our eyes. Of course, what we later come to call objects, we are capable of seeing because of the eye. But it is not just thanks to the eye that we see objects (as opposed to just patterns of dark and light and motion); without the cognitive functions of the geniculate and primary visual cortex, lodged deep in the brain and receiving information from many other parts of the brain, we would not see a thing.

Now one might object here that while I may have shown that we do not non-intentionally see objects, since, as our beige patch example showed, which edge is

counted as a thing-to-be-discriminated edge is not determined solely by the world and by the cells in our retina, it may be that we *do* non-intentionally see very basic things like light and dark patches and motion. This may be the case. If some cells in our retina are set up to respond to light and others are set up to respond to dark and we end up seeing dark and light patches as a result, then we non-intentionally see those light and dark patches. But the crucial point is that we never do end up just seeing dark and light patches; we end up seeing objects arranged in a certain manner across the landscape we are viewing. It is this seeing that Dretske supposes is non-intentional and it is this seeing that we are arguing is intentional.

So it is one thing to receive visual information from the world – this does not count as seeing (not even non-epistemic seeing because no discrimination has taken place) – and it is another to see objects – this is one cognitive step along from receiving information and corresponds to Dretske's non-epistemic seeing – and it is yet another to see objects *as* the objects they are – this is one cognitive step along from seeing_N objects and corresponds to Dretske's epistemic seeing. Dretske makes the mistake of conflating the stage at which we receive information with the cognitive stage at which we see_N objects in our environment.

This point relates to the second of the general motivating thoughts I discuss at the beginning of this section, that even non-epistemic seeing entails some beliefs, not-conscious beliefs, on the part of the perceiver. As I note there, what I am calling 'not-conscious beliefs' does not correspond to 'beliefs' as used in the

literature. But if all intentional states are or depend upon belief states (as is generally supposed), then not to treat states of our brains that are intentional as belief states is a grave oversight. Why are these states of the brain intentional? They are intentional because they are cognitive, that is, they serve to 'encode properties of the world'. Which way the language goes from here on in is not that important; we can either broaden the concept of belief to include states of our brain that are not accessible to our consciousness but that are nevertheless intentional states in that they function to present a certain picture of the world to us, or we can concede that brain/mental states that are not belief states and that do not depend upon belief states, can also be intentional states. I prefer the latter course; and I will explain why in section 2.7.

My second reason for supposing that the eye is not a discriminator relates to the first of the general motivating thoughts I presented at the beginning of this section, that putting discriminating or representing on a continuum is misguided.

It is generally accepted that the notion of discrimination can be understood on a continuum so that the kind of discriminations made by a thermometer, say, can be seen as extremely low on the intentional scale (so low that they are not really intentional, or at least their intentionality is really easy to explain) while the kind of discriminations that humans make, typically, can be seen as high on the scale (that is, really intentional and so very difficult to explain). And, the argument goes, if we can show that even humans engage in different degrees of

discrimination and, hence, intentionality (some more like thermometers – intentionality that is really easy to explain – and some not at all like what thermometers do – intentionality that is really hard to explain), then maybe we can explain the latter in terms of the former. That is the hope at any rate. But now, I ask, why should we accept this picture at all? What argument is there for supposing that the discriminations a thermometer makes are of the same kind and only differ in degree from the sorts of discriminations humans make? Indeed, what justification is there for even calling what a thermometer does *discrimination*, if we want to also use that term to describe a standard human activity? This is clearly a burden of proof argument and those sorts of arguments generally are not very positive. Here is why I think the thermometer analogy is so implausible.

The first step is to pick out what it is about a thermometer's activity that makes it seem prima facie plausible to think of it as an act of discriminating. Well, think of the function of a thermometer. We might say that its function is to discriminate between temperatures; when it is hot outside, the mercury in the thermometer rises (and we can then read off the level it is at and thereby get a temperature reading) and when it is cold outside, the mercury in the thermometer drops and so on. Indeed, it is because thermometers discriminate between temperatures, someone might claim, that we find them so useful as a tool for

measuring the temperature. If thermometers did not discriminate between temperatures, they would not be thermometers after all.

But using 'discriminate' to describe the activity of thermometers is just a metaphorical use of that term. The thermometer itself *does* nothing. The mercury inside the thermometer expands and contracts in response to the temperature in its environment. The *mercury* is not doing any discriminating; it is merely responding to its environment in the way in which it is disposed to behave. Of course, one might say, nobody said that the mercury was discriminating – it is the thermometer that is doing the discriminating. But why should we say this? Certainly, the thermometer allows *us* to discriminate between temperatures – that is why we build thermometers. Any discriminatory capacity a thermometer has comes from those who designed it; from beings with an intention. The thermometer itself is just a graduated cylinder containing mercury. The mercury contracts and expands in response to the environment, but there is nothing discriminatory or representational about this activity. The same, of course, goes for any such device that gathers information about the world that we are interested in getting: these devices are not doing any discriminating at all; we are using them in order to aid *our* discriminating.

Well, what about our eyes then (or the eyes of less cognitively developed organisms)? Isn't it just the *function* of the eye to discriminate? And isn't this

basic visual discrimination pretty similar to the kind of activity the thermometer "engages in", that is, in the sense of not being cognitively driven?

There are two ways someone with this sort of position might go. First, she might concede that a thermometer is discriminating only by virtue of its design in the way just argued: on this view the thermometer is not really doing any discriminating at all – it is merely responding to its environment; but we might want to *call* this activity discrimination in virtue of the discriminations made by the designers, those who use the thermometer in order to discriminate. If this is the way we explain thermometer discrimination though then, by parity of reasoning, it seems we need to explain the eye's discriminating ability by design as well. But by whose design? Evolution's? Mother Nature's? If the analogy is to hold, then whatever is doing the designing imbues its design with discriminatory capability by way of intention. Note that appealing to a notion of "design" by natural selection will not work here because using the term in that way is not analogous to the way it is being used here, namely, as a term that carries with it the idea that the thing designed gets its intentionality *by way of its designers* and not by way of its design (we will be looking at this latter suggestion in section 2.6). Now I take it that most people are uncomfortable with a picture of evolution as somehow inherently intentional. Anyway, if one did go this route, one would not have succeeded in explaining intentionality; one would have merely pushed the explanation one level down.

The second route is to reject the design argument and maintain that the eye just is capable of primitive discriminations (primitive representation) and, by parity of reasoning, if we accept this activity as discriminating activity, then there is no reason why we ought not to accept what the thermometer does as discriminating activity as well. The suggestion that this primitive visual discriminatory ability of the eye can serve as the foundation of full-blown discrimination (in the fully intentional sense) is, after all, the driving hope of information causal theories of meaning.

But this will not work for two reasons. First, even if someone could make a convincing case that this activity of the eye is *discriminating* activity, even in an impoverished sense of that term, it still remains unclear how *that* fact can explain the much more sophisticated kind of discrimination and consequently representation we are ultimately interested in explaining. Second, thinking about the eye in this way just seems plain wrong for the same reason that thinking about a thermometer in that way seems plain wrong. The eye, just like mercury, responds to the environment in various ways – in ways that it is set up to respond to the environment. The human eye, for example, has a lot of receptor neurons that get excited when there is motion. In fact, it seems as though visual receptor neurons *only* get excited by motion of varying sorts, so that when nothing is moving in our environment, our eye (imperceptibly to the agent) moves instead; otherwise we would cease to see anything the minute we stared at a motionless

object.[8] Other receptors respond to dark, others respond to various lights, etc.. But all of this is to say that the eye itself is not doing any discriminating per se; like the mercury in the thermometer, its neurons are responding to various aspects of the environment and, in this sense, are receiving information. Our eyes receive all sorts of information from the environment, but not all of it is stuff we eventually see. Of course, our eyes also do not receive information from many aspects of our environment because not everything in our environment is capable of causally influencing our eyes. Now the idea that the causal relation is an information relation is a powerful one that scientists as well as information causal theorists such as Dretske and Fodor utilize: we learn much about the world by studying the causal effects of the environment on rocks, sea beds, bones, etc.. But (I hope) we do not want to say that rocks, sea beds, and bones discriminate between the various aspects of their environment simply because the results of their causal interactions are informative to us and allow *us* to discriminate between various causal sources (in the same way that the results of a bit of mercury's interaction with the environment is informative to us when we stuff it in a graduated glass)? But then why treat the eye any differently? The eye, on its own, is no different from a quantity of mercury (in its discriminating capacity). Of course the eye is not on its own generally. It is connected to our brain. And what information we end up getting from our eyes, that is, what we end up seeing, is the result of a whole lot of cognitive processing.

So eyes are like mercury but the visual system is not like a thermometer. Why? Because our visual system does not get its discriminatory capacity from an external source; it is, in itself, a discriminatory system. I am not advocating here a view that intentionality must, as a result, remain forever mysterious; that there is not *some* physicalistic description of our brain and our interaction with our environment that will explain intentional phenomena. I am saying, however, that it is not going to work in the way that Fodor, Dretske, and others seem to imagine it to work. At the very least, an explanation is going to have to let go of the idea that an account of intentionality in non-intentional terms entails that we can or sometimes do gather information about the world in non-intentional ways.

2.4 Consequences for Naturalistic Causal Theories

Recall that causal accounts explain reference in the following way: say x is a sign; then x's reference is the y (where y is variously taken in different theories as properties or objects) that is causally related to x in the appropriate way (what this way is depends upon which causal theory we are talking about). The relation between speaker, speaker's utterance, and object is causal. How does y cause tokenings of x? Via our perception of y's of course.[9] But remember that the causal theorists are trying to tell a naturalistic story, that is, one that does not appeal to semantic or intentional concepts. Since causal accounts rely upon perception as the vehicle for these causal relations, it is essential that that perception be non-

intentional. But that is not the case, as I have just argued: any perceiving we do is cognitively driven. But if this is the case, then the resulting perceptions are not non-intentional and thus the causal theory has failed in its naturalistic agenda.

2.5 What is at the Root of These Problems?

The problem, I take it, is one of confusing intentional explanations with intentional activities; confusing the idea that we can give a physicalistic explanation of intentional states with the idea that there is a level at which our information gathering behaviour is non-intentional. The latter does not follow from the former.

Once we have let go of this idea, we can begin to investigate why the perceptual relation is an intentional relation. A plausible suggestion is that we just are intentional systems; when we take in information about the world through perception, we shape that information according to the way we think about the world. If this is the case, then one part of the task of a naturalistic theory of meaning must involve providing a non-intentionally described account of what it is to *be* an intentional system. Functionalist accounts of representation are attempts at doing just this. Now functionalism may not, in the end, be the right way to understand intentional systems, but at the very least it gives us an idea of what such an account could look like. Let us briefly take a look at one suggestion for a functionalist account.[10]

2.6 A Functionalist Account of Representation

Since I am primarily interested here in the issue of intentionality, my discussion will centre around examples of functional explanations of discrimination and representation.

It should be clear by now why I take it that these two issues are intimately related: to discriminate between two aspects of one's environment is to represent one's environment in a certain way. And to represent one's environment in a certain way, to be in a representational state, is to be in a certain type of internal state. Since this state is internal, it is difficult, when observing from the outside, to discern whether or not an organism is in such a state. Mere response behaviour, for example, can be seen as discriminatory behaviour when observed from the outside. But if we want to come to an understanding of what being a discriminatory system amounts to, as opposed to a system that is set up to respond to the environment in fixed ways, then we need to be careful to keep these two kinds of behaviour as distinct as our knowledge of representational states will allow us. Now while we cannot as yet group *all* systems according to whether or not they are capable of discriminations, there are some clear cases of each.

Suppose, for example, that a certain kind of tree responds to the proximity of a kind of fungus by secreting a dark green pigment. The tree, then, enters an internal state in reaction to some external state. Suppose also that the tree only

responds in this way to this kind of fungus. Is the tree discriminating between this fungus and the rest of its environment? No, because the tree has no choice but to enter into this internal state when the fungus is present, and because this internal state is the only one possible for the tree in response to the presence of the fungus. In this, the tree is behaving in the same way that the blob of mercury is behaving when it contracts or expands in response to the temperature of the air. When we later cut down the tree, we can determine exactly when this fungus was in the vicinity by dating the dark green ring in the tree's trunk. This internal state, however, is not a representational state (although *we* manage to get information from it), since it did not result from a discrimination.

On the other hand, when Fred thinks 'there's a tree', he *has* entered into an internal state, in response to some external state, that is the product of a discrimination since Fred need not have entered that state (yesterday when Fred walked by this way he did not notice the tree) and Fred might have entered another state in response to the external stimulus (e.g. he could have thought 'there's a forest'). There are other reasons, of course, for why Fred is discriminating here rather than merely responding to his environment; the aim of the functionalist account is to provide these. This is just a first approximation and is intended merely to illuminate the uncontentious claim that there is a difference between how trees, for example, and humans interact with their environment.

So, all things respond to their environment in various ways, e.g. rocks, water, cats, etc., but only some systems are capable of representing their environment. If we want to understand what representing amounts to, then we had better start by examining representational systems.

Allan Gibbard proposes an interesting functionalist account of representation along the lines of that presented by Ruth Millikan and Fred Dretske, one which draws on an account of biological function worked out by Larry Wright.[11] In order to understand the proposal, the first idea to get is that of a natural function. To say that a system's natural function is to do F is to say that doing F is the reason why the system exists where 'the reason why the system exists' is shorthand for an explanation in terms of natural selection. In this way we get an account of how biological organisms are "designed" (function as if they are designed) without presupposing a designer with intentions. In Gibbard's terms, something "is a *natural function* of a system to do F in an organism, let us say, if and only if the system has some of the features it has in that organism because of natural selection to do F." (Gibbard, 109)

This line seems promising because it suggests possibilities for the person who wants to accept that thermometers have *their* functions by design (and consequently can be seen as systems of artificial representation[12]) and yet does not want to suppose that biological systems have *their* functions by design. The analogy between designed systems and natural ones holds, the argument goes,

because natural selection *mimics* design. If one can maintain an analogy between the function of systems such as thermometers and those such as the human visual system in virtue of the fact that they can both be given functional explanations of the sort above, however, then it seems that we are buying into the practice we have just rejected, of looking to non-intentional systems in order to explain ours. But supposing that functionalism might provide a naturalistic explanation of intentionality need not commit us to this position. Here is why not.

Systems that are supposed to be "primitively" intentional (on my view they are not intentional at all because they respond to their environment rather than discriminate between aspects of it) are functionally different in kind, not just in degree, from intentional systems. Gibbard makes the following suggestion toward explicating this difference:

> A *system of natural representation* for a feature S of the world is a system one of whose natural functions is to adjust some feature R of the world to correspond to S – ... by a scheme of correspondence that is somewhat arbitrary. R can then be called a *natural representation* of *subject matter* S. (Gibbard, 109)

Now this is an interesting suggestion. On this account does a thermometer serve as a system of artificial representation? Well, in cases where thermometers are taken to be representing the temperature, it is suggested that the reading on the thermometer at a given time is what is representing the temperature. So this reading would correspond to R above. But is the correspondence between the temperature outside and the reading on the

thermometer an arbitrary one? Our impulse is to say no, of course. But why? This notion of arbitrariness is not as clear as one might want and Gibbard does not help in making it clearer. But some of Millikan's suggestions, although not made with reference to this problem in particular (her discussion focuses on the distinction between systems that learn and those that do not), may shed some light on how we might understand the arbitrariness condition.

In her article "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox", Millikan (1990) makes an interesting distinction between the behaviour of rats and the behaviour of hoverflies. Both a rat's and a hoverfly's behaviour, in responding in various ways to external stimuli, can be characterized functionally as rule-conforming behaviour. A rat and a hoverfly differ, however, in the way in which each conforms to his respective rules. The hoverfly's internal system does not alter as a result of his rule following, while the rat's internal system does:

> When the hoverfly conforms to his rule, nothing in his body undergoes a permanent change, but this is not so in the case of the rat. Suppose, for example, that the rat has just become ill after eating soap. In order to conform to the proximal rat rule, in order to avoid henceforth what tastes like soap, the rat's nervous system must first conform to certain preliminary "rules", rules that dictate that a certain sort of permanent change take place in it. (Millikan, 1990, 338)

So the rat follows his rules by learning. One difference, then, between systems that follow rules by learning and those that do not is this: systems that follow rules by learning change in response to certain external stimuli, whereas those

that do not learn do not change. Changing in response to external stimuli in this fashion, then, can be seen as a necessary condition for learning. Thus, "animals that learn can acquire biological purposes that are peculiar to them as individuals, tailored to their own peculiar circumstances or peculiar histories." (Millikan, 1990, 339) Now this way of capturing the difference between systems that *learn* to accommodate rule following and those that do not seems to capture at least one aspect of the arbitrariness requirement Gibbard suggested is a necessary condition of a functional system being a representational system. The (or one) difference between a system that merely responds to its environment, as thermometers and hoverflies do, and systems that discriminate and hence represent their environment, as humans and rats do, is that in the former case, the internal states of the system can affect other internal states or result in external behaviour in only set ways; while in the latter case, some of the internal states of the system are capable of *changing* the functional roles of other states or of *creating new* internal states with new functional roles.

To see how non-intentional systems do not meet this requirement, let us look at another example, one a little more functionally complex than a thermometer, say a thermostat. Some of a thermostat's internal states are capable of influencing other of its internal states; the dropping of the thermostat's temperature gauge below the temperature at which the room is regulated at, for example, will cause another internal state, namely, the one that activates the

furnace. Now a number of different factors may cause a thermostat to enter into the internal state that activates the furnace; both the normal functioning of the temperature comparator and a broken temperature comparator, for example, may cause the temperature controller to activate the furnace. Does this fact make the correspondence between this internal state and the temperature outside an arbitrary one? No because notice that in none of the thermostat cases we can dream up will one of its internal states affect another of its internal state so as to either change that state's functional role or cause a new internal state to be created: whatever happens, the temperature controller will either turn on the furnace, turn off the furnace, or do nothing. Not so in the case of the rat. Some of the internal states that the rat enters into *will* cause the creation of a new internal state, one with a new functional role; in the example provided, the new functional role is characterized by Millikan as the conforming to the 'do not eat what tastes like soap' rule.

The suggestion here, then, is that a system is a system of representation if one of its natural functions is to adjust one (or many) of its internal states to correspond to one (or many) external states according to an arbitrary scheme. The scheme is arbitrary if the correspondence may result either in a change in the functional role of the internal state or in the creation of a new internal state with a new functional role. Of course much more needs to be said in order to give a functionalist *account* of representation; nevertheless, I think this

suggestion is promising because it does capture a central intuition we have about intentionality – that an intentional system somehow directs its actions, rather than merely responds to its environment.

2.7 Intentional Perception

The conclusion of sections 2.1 to 2.6 is this: one route to providing a naturalistic theory of meaning, a route that is popular among various sorts of contemporary causal theorists, is to argue that our meaning relation with the world is grounded in a non-semantic relation with the world. This non-semantic relation with the world is a causal one (typically) and the vehicle for this causal relation is perception. In the past six sections I have been arguing that this enterprise is misguided because human perception, being an intentional activity, cannot serve as the vehicle for such a non-semantic relation with the world. In the course of this discussion, I have begun to paint a picture of what this intentional perception looks like. In this section I will complete this picture and thereby end this largely negative chapter with a positive view. The view of perception I develop here will, in turn, have ramifications for theories of meaning, in particular for naturalistic theories of meaning. These will be explored in chapter three.

To perceive something is to represent it, that is, re-present it. This distinguishes the kind of causal interactions that occur in perception from those

that occur in, say, chemical reactions. Now right away one might object that in restricting the use of 'perception' to acts that are representations, I am begging the question of whether or not perceiving is an intentional activity; if perceiving just is representing, then of course it is an intentional activity. In this section I am not going to argue that perception is an intentional activity – I have already argued, in the course of examining Dretske's position, that understanding perception as an intentional activity makes the best sense of our philosophical and scientific findings. Instead, I am going to fill out what the picture of perception as an intentional activity looks like and address some of the consequences for this view.

So, to perceive an aspect of our environment is to represent it as something in our environment. The way in which we represent our environment, then, seems to be partly up to us. This is to admit that our perceptions do not yield objective observations. To say that our perceptions do not yield objective observations is to admit that our perceptions are always subjectively coloured. But this does not preclude our perceptions being public. To say that our perceptions are public is to claim that the subjective coloration of our perceptions is a universal condition. Now I said in my introduction that I found the intuition that none of our observations have "objective" status compelling. Since accommodating this idea is one of the motivations for my project, I will spend some time here explaining why.

Descriptions in general occur from a vantage point, a perspective. As such, it is acknowledged that, in describing, they never fully capture whatever it is they are describing. A description is just that, after all: a description; it cannot take the place of the thing itself. Now an "objective" description is one that is had after transcending various points of view and to that extent it is either not really describing any more or else it is describing from a god's eye point of view. But only a god can have a god's eye point of view – this is the nature of that perspective. But we are not gods ... so how do we make sense of this idea?

What could it mean to say there are objects independent of our representations of them as objects? If by this it is meant that there are things in the world that answer to what we mean by 'object' then we have no reason for argument. But this is not what is typically meant when something's objectivity is appealed to. In those cases, the idea is not just that there is something in the world answering to one of our concepts; rather, it is that our concepts capture a way in which the world is really structured. But finding instances of our concepts is not justification for concluding something about the world itself. Thinking that it is is to suppose that finding that there are things in the world that fit the concept KLURG[13] ("portions of surfaces of books plus tables delimited by a chalk line" (Searle, 1995, 160)) is reason to suppose that the world, independent of us, is full of klurgs. Well, in a sense it is. But it would be nonsensical to then go on a hunt for *real* klurgs, things that are klurgs according to nature. What could this

hunt possibly be except the search for the application of yet another concept, namely, the KLURG-ACCORDING-TO-NATURE concept? People who are on the hunt for the list of natural kinds are like these people; they are searching for something that fits their natural-kind concept and when they find it they think that they have found real natural kinds when what they have found are really just things that answer to their NATURAL-KIND-ACCORDING-TO-NATURE concept.

So there are many different ways of describing, say, an aspect of the environment. Since we cannot make sense of a god's eye perspective, because we are stuck with human perspectives, a possible, anti-realist, take on these considerations might be that no sense can be made of a correct description at all (as opposed to the view that while some descriptions may be correct, no sense can be made of *the* correct description). That argument might go something like this. Making any judgements of correctness involves assessing against the background of the way things really are. Now the only way to capture the way things really are is by a complete description. But only a god's eye description is a complete description, and since no such description is possible, the notion of judging relative to the way things are is an incoherent one. Consequently, none of our descriptions can be judged to be either correct or incorrect.

This result is unsatisfying. It does not do justice to the thought that there really do seem to be correct and incorrect ways of understanding situations. On this anti-realist view, for example, the person who is in charge of his faculties and

who is in optimal viewing conditions is no more correct in judging that there is a table in front of him than is the hallucinating individual who judges that there is a tiger in front of her. But surely this is an unhappy result. The reason for this, bluntly speaking, is that there really is a table there, not a tiger. Of course, insisting that there really is a table there is not to hold that all descriptions of the world that are correct must include 'table, here, now'; rather it is to say that from the perspective of the conceptual scheme humans use to interact with and to understand their environment, there is a table, here, now. As Putnam points out, "... why should one suppose that reality can be described independent of our descriptions? And why should the fact that reality cannot be described independent of our descriptions lead us to suppose that there are only the descriptions?" (Putnam, 1992, 122)

Now one way of understanding this idea of a human perspective is via the notion of publicity I mentioned earlier. It is true that our perceptions cannot yield objective observations; but, they can and do yield public observations (indeed this is one of the reasons we often *take* our observations to be objective). A perception yields a public observation if what helped shape the perception are concepts common to all humans (this idea will become clearer in my discussion of concepts in section 3.1). When we are trying to do respectable research, we want our theories to be based on observations that are public in this sense, rather than on ones that are coloured by concepts that are not universally held.

Sorting out when biases are at work and when only the peculiarities of a human perspective are influencing one's observations is a difficult process. But making this distinction in theory can help us to see that while our concepts are *always* influencing even our most basic observations of the world, there is still a distinction to be made between observations that we all make and those that we do not share.

John McDowell (1994) in *Mind and World* takes up this issue and advances a view similar to the account of intentional perception I am presenting here. He does not come to all the same conclusions;[14]however, the philosophical connections he draws to Kant in this enterprise are illuminating and will help in making the explication of intentional perception clearer.

McDowell sets about breaking free from what he calls the dualism of Scheme and the Given, namely the dualism between the range of our concepts and the range of our experiences, by rejuvenating a Kantian picture of our interaction with the world. Two notions are important in understanding how this works: spontaneity and receptivity. According to McDowell, on the traditional dualistic picture (the one that sense data theorists ascribed to, for example) a distinction is made between Scheme and the Given, between the range of our concepts and the world. Scheme is associated with spontaneity: within the range of our concepts we have input in our beliefs and thoughts. If *all* of our thoughts, assertions, judgements, etc. were founded in our scheme, they would be

completely spontaneous, and thus would not be accountable to anything or grounded in anything outside of human spontaneity. Because we think there is something more to our thoughts, assertions, judgements, etc. than this, that is, because we think that our beliefs have content, the realm of the Given must be postulated. Our interaction within this realm is characterized as receptive. Seeing humans as receptive in their interaction with the realm of the Given ensures that the elements of the Given really do serve as constraints upon our thoughts and consequently, can account for the content of our beliefs. Thus when we are interacting with the world, on this view, we are *receiving* input from the external world. Because we make no input at this level, there is some kind of guarantee that we receive a bit of the way the world really is.

But, McDowell claims, this picture will not do what it is meant to do, namely, establish a justification for our beliefs about the world. This is because the brute inputs from the external world cannot serve as justifications so long as they are outside of the realm of the space of our concepts, as they must be if they are in the realm of the Given.[15] McDowell puts this point this way:

> But we cannot really understand the relations in virtue of which a judgement is warranted except as relations within the space of concepts: relations such as implication or probabilification, which hold between potential exercises of conceptual capacities. The attempt to extend the scope of justificatory relations outside the conceptual sphere cannot do what it is supposed to do. (McDowell, 1994, 7)

The solution according to McDowell is to give up on this dualism. But giving up on this dualism need not be seen as giving up on the notion of truth or justified beliefs. Again, invoking a Kantian notion, McDowell suggests that we understand our experiences of the world in the way in which Kant saw them, as intuitions. On this understanding, an experience of the world is both receptive and spontaneous: at every moment during which we are impinged upon by the external world, we are also conceptually engaged:

> This makes room for a different notion of givenness, one that is innocent of the confusion between justification and exculpation. ... When we trace the ground for an empirical judgement, the last step takes us to experiences. Experiences already have conceptual content, so this last step does not take us outside the space of concepts. But it takes us to something in which sensibility – receptivity – is operative, so we need no longer be unnerved by the freedom implicit in the idea that our conceptual capacities belong to a faculty of spontaneity. We need not worry that our picture leaves out the external constraint that is required if exercises of our conceptual capacities are to be recognizable as bearing on the world at all. (McDowell, 1994, 10)

McDowell spends some time pointing out how Davidson, in "A Coherence Theory of Truth", recognizes this problem with the 'myth of the Given' – that so long as the space of reasons extends beyond the space of concepts, it cannot serve as justification for our concepts – but does not see the conclusion that McDowell reaches, namely, that our concepts are involved even at the level of receptivity; Davidson cannot rid himself of the view that experience just is outside the conceptual sphere. Since, by his own argument, this experience cannot

serve as a justification for beliefs, then the only justification for beliefs is other beliefs, not experiences: "Davidson recoils from the Myth of the Given all the way to denying experience any justificatory role, and the coherentist upshot is a version of the conception of spontaneity as frictionless, the very thing that makes the idea of the Given attractive." (McDowell, 1994, 14) Davidson's solution is a problem because, so long as our thoughts remain rationally unconnected to the world, they will remain empty of content as well:

> When Davidson argues that a body of beliefs is sure to be mostly true, he helps himself to the idea of a body of beliefs, a body of states that have content. And that means that, however successfully the argument might work on its own terms, it comes too late to neutralize the real problem for this horn of the dilemma. (McDowell, 1994, 68)

In order to retain this notion of content, then, we must recognize that our experiences are spontaneous as well as receptive; they involve our conceptualization of the world as well as input from the world.

Now we can see how these philosophical conclusions tie in with the view of perception we have been developing thus far. It is through perception that, on the old story, we were supposed to have received a bit of the Given. But we do not receive a bit of the Given through perception; to perceive the world is to experience it, and experiences are intentional acts. To be sure, they are different from intentional activity that only involves our thoughts; for example, when we think about our thoughts. But this is because perceptual experiences are partly

receptive as well; it is the world that we are perceiving after all. So what makes perceptions partly spontaneous?

I argued earlier, in discussing non-epistemic perception, that beliefs that are not-conscious play a part in shaping our perceptions; and I suggested that using a term besides 'belief' is to be preferred. There are good reasons for using the term 'concept' instead of 'belief' to refer to these not-conscious beliefs, even though doing this makes discussion confusing at times since, in general, philosophical discussions about the mind and epistemology are presented in terms of beliefs. But, as Paul Thagard argues, not to distinguish between beliefs and concepts leaves us unable to account for much of our mental activity. [16] Without appeal to concepts, for example, how does one explain the origin of beliefs? How does one come to believe 'the cat is on the mat' unless one already has a concept CAT and MAT? Neither can the representation of our mental life solely in terms of beliefs and belief revisions adequately explain the transformations in thinking that occur when conceptual shifts take place. In addition there is the awkwardness, that we have already encountered, that beliefs are generally taken to be mental entities that are available to consciousness, while concepts may not always be so available. I will follow cognitive psychologists, then, in calling these aspects of our mental lives 'concepts'. Put in more familiar terms, I will follow Thagard in treating "concepts and propositions as mental representations, with concepts corresponding to

predicates and propositions corresponding to sentences." (Thagard, 1990, 258) On this view, concepts are more fundamental than beliefs or, as I will call them from here on in, thoughts, in the sense that while one could not have a certain thought without also having concepts related to it, one could have a concept without having a related thought. For example, one could have the concept CAT but never think about cats (although, of course, the kind of being who has concepts must also be the kind of being who has thoughts).

Tying this back to McDowell's discussion, we see that the spontaneous component of the otherwise receptive activity of perception is the influence of concepts. It is not clear where in the body conceptualization begins, that is, at what point of our interaction with the world we are influencing and shaping the information we are receiving. With respect to vision, the perception we know the most about, it seems that it begins at the outer most stage, at the stage of cells in the retina because even these are connected via recurrent networks to cell structures deeper in the brain. Not all of our causal interactions with the world are causal/conceptual ones of course; when a body is caused to perspire by the rise in temperature, the relation between the air and the skin is a purely causal one. But the causal relation that results in perception is always, on this view, attended by conceptualization and is, thus, a causal/conceptual relation rather than a purely causal one. What the nature of the conceptual side of the relation is and what consequence this has for a theory of meaning we will turn to now.

Chapter Three ••• A New Approach

Introduction

I have argued thus far that the naturalism of causal accounts of meaning is undermined by the fact that the causal relation taken to be the reference relation (and, if meaning is reference, the meaning relation as well) is an intentional one. The problem, I have been suggesting, is the failure to appreciate the significance of the intentional nature of speakers. The relevant causal relation, namely the perceptual relation, is an intentional one because systems that have perceptions are intentional systems; it is because of something about them that the perceptual relation is an intentional one. Thus, if the causal theory is going to appeal to the perceptual relation in its explanation, and it is hard to see how this could be avoided since it is only once a speaker has perceived something that she can come to think about it, then a naturalistic causal theory needs to be wedded to a non-intentional account of what it is to *be* an intentional system. We looked at one suggestion toward this end in section 2.6. Now because the causal story seems to be really right about some things, namely that the content of an intentional state must be related to its object in some non-mysterious way, I will focus, in this chapter, on giving a sketch of what a causal theory that takes into account the intentional nature of the causal relation that determines meaning might look like. In sections 3.1 to 3.3, I will discuss the role

that concepts play in our mental lives in order to make the conceptual aspect of the perceptual relation clearer. In sections 3.4 to 3.6, I provide a sketch of a causal/conceptual answer to the following three central questions of intentionality: 1) How do thoughts come to be? 2) What makes a brain state an intentional state? 3) How are intentional states individuated? And, in the final section, I give two examples of how this new view can be used to answer traditional problems in the philosophy of language. But before we proceed we should situate this approach with respect to the questions raised in chapter one.

I concluded from chapter one that pragmatic accounts seem to be on the right track. So one of the issues surrounding meaning, that of determining what words and sentences mean, is answered pragmatically in the following way: the meaning of a word or sentence is what we intend to communicate by using the word or sentence. By taking this approach, we make room for the context-sensitivity of language and we reduce the problem of linguistic meaning to that of speaker meaning (which is a more likely candidate for naturalization). But now we have two problems: 1) explaining how words become usable in the way they are – that is, explaining literal meaning; and, 2) explaining speaker meaning.

We now solve the first problem in this way. Literal meaning is a matter of convention. Words are signs or names for the distinctions we make in the world. We learn which signs express which distinctions during language acquisition. So much for literal meaning.

Now to the hard part – explaining speaker meaning. Let us look first at what question we need to answer: How does speaker meaning hook up with the world? This is *the* question, the problem of intentionality. To make sure we do not go off track, I will break this problem down into three sub-questions, each of which will be familiar to us: 1) *How* do we get the stuff in our heads? 2) What makes the stuff in our heads *about* the stuff out there? 3) *What* out there is the stuff in our heads about? To answer the first question, we need an account of language acquisition. To answer the second question, we need an account of the reference relation, an account of what makes one thing about something else. And, to answer the third question, we need an account of how the contents of our mental states are individuated. I will deal with each question separately, presenting problems with the ways in which each has been addressed and suggesting a causal/conceptual alternative, that is, one that is in line with the understanding that perception is an intentional activity. Now we turn to a discussion of concepts.

3.1 Language Acquisition

As soon as we begin to tackle the question of how stuff in our heads gets to be about stuff out there, we realize that there are two groups of stuff in our head we have to deal with, namely, thought tokens and concepts. I make a distinction between the two because, on the view of mental structure I am

advocating here, concepts are more fundamental than thoughts in this sense: one cannot have a thought without a related concept, while one can have a concept without a related thought, although, as I mentioned in 2.7, only beings who are capable of having thoughts also have concepts. For example, one can think 'there is a cat before me' only if one has the concept CAT, while one can have the concept CAT and never (with the exception, perhaps, of the time at which the concept is formed) think 'there is a cat before me' (or some such thing). We can think of this relationship in this way: concepts are the non-occurrent states of our brain that make occurrent states, thoughts, possible. But since both concepts and thoughts are mental representations, we must answer our question for both. Since concepts are more fundamental, we will begin with those.

Concepts[1]

What is a concept? At the end of the last section I indicated that I am treating concepts as mental representations that correspond to predicates. This is appropriate since concepts prepare the terrain for belief formation. Working together they form what is often called a conceptual scheme, a way of understanding the world.[2]

For our discussion, we need to recognize the following three characteristics of concepts: the hierarchical relation they stand in with respect to one another; the degree to which they pervade our mental lives; and, the way in which they influence our thoughts. The concept ROOSTER, for example, which

could only exist within a highly complex *conscious* intentional system, is high in the hierarchical structure because it depends upon so many other concepts. We can understand this dependence in this way. If we take a rooster to be a male chicken then the concept ROOSTER includes the concepts MALE and CHICKEN. These concepts are more fundamental, we will say, since you cannot have ROOSTER without MALE and CHICKEN, but you can have MALE and CHICKEN without ROOSTER. Thus, ROOSTER directly depends upon MALE and CHICKEN. Concepts may also indirectly depend upon other concepts. For example, if we think that a chicken is a domestic fowl, then the concept CHICKEN includes and is, consequently, dependent upon the concepts DOMESTIC and FOWL. But dependence is a transitive relation. Thus, if ROOSTER depends upon CHICKEN and CHICKEN depends upon FOWL, then ROOSTER also depends upon FOWL. I will call a concept a secondary level concept if it either directly or indirectly depends upon other concepts in the way just described. By this same method, we can distinguish between secondary level concepts that are higher, lower, or on the same level on the dependence hierarchy, but we will not be going into this kind of detail here.

The degree to which secondary concepts pervade an individual's mental life can vary somewhat from thinker to thinker; however, since all secondary concepts are at a higher hierarchical level than basic concepts – those that do not depend upon any other concepts – secondary concepts are less pervasive

than basic concepts: since secondary concepts are not connected to as many of our concepts as basic concepts are, they play a smaller role in our conceptual lives. The concept MOVING THING, for example, is a much more fundamental concept than ROOSTER (perhaps it is even a basic concept) as it depends upon fewer (if any) concepts. And as a more fundamental concept, it is more pervasive, that is, it influences a wider range of agents' thoughts than the concept ROOSTER in virtue of the greater amount of relations it has to other concepts

Finally, some concepts are available to agents through consciousness and, generally, influence thoughts in this way, that is, through the agent thinking directly about the concept. Other concepts are not so available and influence thoughts outside of consciousness. Our example above, the concept ROOSTER, is one of the former sort: it influences our thoughts via our ability to think directly about it. The concept MOVING THING, on the other hand, may be a concept that exerts influence outside of consciousness (although it *could* exert influence through consciousness if one started to think about the concept directly). If so, it is a concept that systems that are not conscious and yet are intentional could have. This last observation needs some explaining.

Traditionally, concepts have been thought of as the territory of conscious beings; having a concept, people thought, is a highly developed mental capability. Certainly this is true for the majority of the concepts we entertain, for

the majority of the concepts we entertain are of the ROOSTER variety I discussed above; since these sorts of concepts exert their influence through consciousness, a system must be conscious in order to have them. But some thinkers have distinguished between conscious systems those capable of having concepts – not conscious systems – those not capable of having concepts – and systems (like dogs and cats) that may not be conscious, but that nevertheless are capable of having concepts. If, in the latter case, the system in question is not a conscious one, then none of *its* concepts will have what is a central feature of the majority of *our* concepts, namely, availability through consciousness. Consequently, some have suggested, we should distinguish its concepts from ours by calling them 'proto-concepts' (for example, Putnam, 1992, 29). People who ascribe to this picture acknowledge that while proto-concepts are not as highly developed as ours are – "there is probably nothing in the dog's neural architecture which would allow it to think 'this piece of meat came from an animal'" (Putnam, 1992, 29) – they do nevertheless serve to categorize important aspects of the system's environment: "the dog's mental processes involve a 'data structure' which was selected for to do certain things; perhaps the data structure responds in a certain way when the dog sees meat ...." (Putnam, 1992, 27)

Now there certainly seems to be something right about distinguishing systems along these lines; but this manner of stating the difference is unsatisfactory because it fails to explain why, if proto-concepts are so different

from our concepts, we want to call them concepts at all. In order to get a clearer picture, we first have to recognize that not all of our concepts are of the ROOSTER variety either. As we have already noted, some concepts, like the concept MOVING THING perhaps, influence the way we perceive our surroundings without our ever being aware of this influence. Concepts such as these exert their influence outside of consciousness in a way that may be similar to the way in which a non-conscious intentional system's "proto-concepts" influences its perceptions. In order to understand how this might work, let us turn to the many different suggestions as to what, in the brain, can be identified with concepts.

Connectionist models of the brain provide promise in explaining the way in which concepts influence a system outside of consciousness and, consequently, they give us a way of understanding the similarity between some of our concepts and the proto-concepts of non-conscious systems. Very roughly, they claim that a concept is a pattern of neuronal activity, consisting in varying degrees of synaptic responses, where the interconnected neurons are spread throughout different regions in the brain. It is possible that we are born with at least some of these structures; but, even if we are, it is certain that we add on to the existing ones and form others throughout our lifetime. Now the interesting point about these structures is that they are capable of existing in and influencing systems that are not conscious. According to Paul Churchland,

> If our cognitive activities arise from a global theory as a global configuration of synaptic weights, ..., then it is clear that no cognitive activity whatever takes place in the absence of vectors being processed by some specific configuration of weights. That is, no cognitive activity whatever takes place in the absence of some theory or other.
>
> This perspective bids us see even the simplest animals and the youngest infants as possessing theories, since they too process their activation vectors with some configuration of weights or other. The difference between us and them is not that they lack theories. Rather, their theories are just a good deal simpler than ours, in the case of animals; and their theories are much less coherent and organized and informed than ours, in the case of human infants. (Churchland, 1997, 284)

Thus all systems that engage in cognitive activity, conscious or not, also engage in conceptualization. The difference between intentional systems that are conscious and those that are not is that the former can engage in conceptualization through consciousness in addition to conceptualization outside of consciousness, while the latter cannot. These are distinctions between intentional systems, of course; while it may be true that some systems have a more complex system of representations than other systems – in that they can represent more things and are more flexible in adapting their representations to new environments – it still holds that some systems may not be representational at all. For example, trees and oceans and thermometers and perhaps simpler animals, ones that do not engage in cognitive activity, are not representational systems at all. Thinking about them as such merely leads us away from understanding what being a representational system amounts to.

Now I have been suggesting here a way of thinking about what concepts are and how we can differentiate between different kinds of concepts, but I have not yet mentioned the issue of concept acquisition. While it may be that at least some of our concepts are innate, it is generally accepted that the majority of them are acquired, although no good account of concept acquisition is in circulation at the moment. We do not have to decide this question, however, in order to speculate as to the reasons for the acquisition or innateness of our concepts. Doing this will help us understand the role that concepts play in shaping our representations. Since a concept is, roughly, a category that separates off certain aspects of the world from others, the following two general questions will frame our discussion: 1) Why make *any* distinctions? and 2) Why make *these* distinctions?

3.2 Why Make *Any* Distinctions?

Consider what it would be like if we did not make any distinctions. There would be nothing to focus our attention on, nothing to react to. If I did not distinguish between edible and inedible things, I would die in short order. If I did not distinguish between tigers and ants, I would die in short order. If I did not distinguish between my baby and the rest of the world, my line would die off in short order. And so on. In short, the answer to the question of why we make *any* distinctions is 'to survive'.

3.3 Why Make *These* Distinctions?

Now as we have seen in our discussion of perception and our discussion of concepts, given our capabilities and the external environment, there are a number of different distinctions we could possibly make relative to any given aspect of the world. For example, we could sort our environment into objects according to the concepts GREEN and NOT-GREEN or we could sort our environment into a different set of objects according to the concepts HARD and NOT-HARD. Were we to engage in this sorting activity, we would be doing so consciously, by reorganizing our perceptions according to concepts we have in mind. But there seems to be another way in which we go about sorting that is not conscious. When we have a visual perception, we do not see patches of light and dark, for example; we see tables, chairs, and books. But that we see tables, chairs, and books (not *as* tables, chairs, and books – that we just see them, in Dretske's non-epistemic sense) is not the result of something we consciously set out to do. Furthermore, because these are non-epistemic perceptions, even people with different beliefs will share them. Now as I argued in chapter two, even non-epistemic perceptions are intentional, that is, they are influenced by the concepts of the perceiver (indeed, it is the having of concepts that makes it a perception, a representation of the environment, that a system is having rather than a mere response to the environment). But if all humans share these

perceptions, then, plausibly, we share some of our concepts as well. If there is a set of concepts that we share, then we must answer the question, if there is a choice in concept formation, why we *choose* these same concepts.[3] Further, if we also hold that we do not share *all* of our concepts (and this seems obvious), then we must explain why we choose these different concepts and how we communicate these differences. I will use the word 'interest' (with qualifiers) to serve as a general term covering our motivation both for why we choose the same concepts that we do and for why we choose the different concepts that we do.

We should clarify right off the mark what an interest is and what sorts of things are capable of having interests. The way we generally use the term, an individual who is capable of having interests is also one that is capable of thinking about herself, or, in other words, is capable of consciousness. To use the term to describe non-conscious systems is to use the term metaphorically or it is to illegitimately import teleological descriptions where they do not apply. Thus, to say that a frog snaps at all fly-like shapes because it is interested in getting flies or because it is interested in surviving or because it is interested in reproducing etc. is at most a convenient shorthand way of telling a much more complex natural selection story that does not appeal anywhere to the frog's *interests*. On the other hand, when we want to give an explanation for why Fred signed up for the course in metaphysics, citing his interest in metaphysical

questions is not shorthand for anything; Fred really does have interests and these interests are the explanation for his behaviour. Of course, if we do manage to come up with a naturalistic account of intentionality, we could tell the story in non-intentional language, for example in the language of neurons and cortical cell structures. But this story would not change the fact that Fred's behaviour is different in kind, not just in degree, from the frog's because Fred really does have interests and these interests motivate him to act in various ways while the frog really does not have interests. The difference no doubt lies in the fact that Fred is a conscious system while the frog is not.

Now this kind of interest, the kind that seems to require consciousness, is the kind of interest that adult humans have. Probably cats, dolphins, whales, etc. do too. Thermometers do not have interests in this sense. Neither do plants or clams, paramecia, or very young human babies. But adult humans and other conscious systems are not *just* conscious systems; they have interests[4] like those of the frog as well. We do not, generally, eat because we have a conscious desire to do so; rather, we eat because our bodies compel us to search for food. These sorts of interests, the kind we share with all biological organisms, we will call our basic interests.[5]

Our interests fall into two general groups then: those that we all share in virtue of our biology and our general environment, and those that we do not share. I will call the former our basic interests and the latter our secondary

interests. Examples of basic interests[6] may include interests in breathing, in eating foods that are digestible by human bodies, in being sheltered from extreme heat and cold, and so on. Secondary interests may include interests in having a specific individual for a mate, in being sheltered by *that* shelter, in eating sweet things, and so on.

Now one plausible reason why we share a set of basic concepts, namely those that ensure that, by and large, our non-epistemic perceptions are all the same, is that we also share our basic interests. Thus it might be because we share a basic interest in eating nourishing foods that we see the world in terms of medium-sized objects; doing this enables us to pick out poultry, fish, plants, etc.. Of course, seeing the world in terms of medium-sized objects does not entail that we have the concept MEDIUM-SIZED OBJECT; a baby, for example, does not need to have the concept MOTHER in order to see her mother. Most likely, the basic concepts we have that make it the case that our perceptions are *of* medium-sized objects are not available to us through consciousness. Since we are highly developed intentional systems, unlike frogs perhaps, our basic conceptual basis is flexible; in environments where seeing the world in terms of medium-sized objects does not serve our basic interests, we might develop alternate conceptual schemes. Frogs, on the other hand, might just die off. We share, by and large,[7] a set of basic concepts, then, because we share our basic interests (by virtue of our biology) and we live in similar environments.

At another level, of course, our interests diverge. Social factors as well as our beliefs play a role in determining what our secondary interests will be. These secondary interests, in turn, play a role in determining what secondary concepts we have. But we will not go into these details here. What is important for our purposes is that we recognize that it is the shared set of basic concepts that enables us to communicate and to understand the conceptual differences that emerge as a result of our secondary interests.

Now that we have a better understanding of the role that concepts play in perception, we can turn our attention to thoughts. We begin by explaining how thoughts come to be.

3.4 *How* Do We Get the Stuff in Our Heads?

Let us begin by clarifying what objects are. Some objects are physical and some are mental. A table is an example of a physical object. An idea of a table is an example of a mental object. A necessary condition of a thing's being an object is that it be a segment of the environment that is differentiated from the rest of the environment on the basis of one or more distinguishing criteria. This is not a sufficient condition since not all cases of distinguishing pick out objects; one can distinguish between patches of dark and light on a wall, but not take it that either the light or the dark patch is an object. But providing a necessary condition for objecthood is enough for our purpose here; we need to recognize

only that something cannot causally affect our mental lives unless it is distinguished from its background. Another way to say this is to note that something cannot influence our thoughts and beliefs, that is, cause us to have a thought, unless we can pick it out. If I do not distinguish between the drawer and the chest, I cannot think about the drawer even if it is affecting my visual receptors. If I do not distinguish between ethical egoism and psychological egoism, I cannot think about ethical egoism even if I read a paper expounding the view (since I have not really understood the view). Our ability to make these distinctions is our ability to conceptualize.

Now I mentioned that there are two general levels of concepts: those that are motivated by our basic interests and those that are motivated by our secondary interests. The basic level plays a part in determining what is going to count as a public perception (a non-epistemic one in Dretske's terminology), while the secondary level plays a part in determining how one judges that perception (as well as in determining our more complex thoughts). Our public perceptions are shared because what secondary level concepts and beliefs one has plays no part in determining these perceptions. But this does not preclude the possibility of our judgements of our perceptions, for example, being shared as well. Indeed, one of the reasons we communicate is to share these judgements. We are able to do so because, as we saw in our discussion of concepts in 3.1, all of our secondary concepts depend directly or indirectly upon

our basic concepts. Since we share these basic concepts, we should be able, with greater or lesser difficulty, to communicate these secondary concepts and the thoughts we have as a result of them. In addition, since our social environment plays a role in determining what sorts of secondary interests we have, and these in turn play a role in determining what secondary concepts we have, it seems likely that similarity between the experiences and social environments of speakers will translate into a number of shared secondary concepts as well. An example of how a secondary concept might be communicated will make this clearer.

We both see the painting on the wall. I judge it to be a fine example of the French neo-classical period. You think it is nice. You do not have the concept NEO-CLASSICAL ART (or any other history of art concepts for that matter) so you do not judge your perception in this way. Nevertheless, because we have such extensive common ground, namely all of our non-epistemic perceptions as well as many other judgements about perceptions, I can, if I try very hard, communicate this concept to you. If I succeed, we will share a secondary concept as well.

The claim, then, is that there are two general intentional entities we need to explain; thoughts and concepts. Thoughts constitute what we think about the world. Concepts constitute the way we think about the world. Now concepts can, of course, be formulated as thoughts. When I say 'the cat is on the mat', for

example, I am expressing a thought. A number of different concepts helped to determine this thought. For example, the concept CAT, the concept MAT, the concept BEING IN THE RELATION OF SITTING ON, and so on. I do not generally think about these concepts. But I can do so, and when I do, I will be having a thought. I might think, for example, 'a cat is a small, muscular, mainly carnivorous animal, that is agile and has exceptional night vision.' Here, in having this thought, I am expressing a concept. So, while the role of a concept *qua* concept is to enable thinking, we can, because we are conscious intentional systems, think about our concepts as well. The answer to the question of this section, then, begins to look something like this.

We are intentional systems – beings equipped with the ability to think about the world. We are highly flexible intentional systems, which means that our conceptual architecture is adaptable. Nevertheless, it seems as though, given a certain environment, certain concepts get fixed outside of consciousness and so remain impervious to revision through consciousness. We come to have thoughts, then, through our causal/conceptual interaction with the world: our thoughts get to be about the world through our causal interaction with the world, but the result of our causal interactions with the world are *thoughts* as a result of the conceptual architecture of our brains.

Now I have not really answered the how question here and that is because we still do not really know what in the brain corresponds to concepts.

What I have suggested here is that it is due to the conceptual architecture of our brains that some of our brain states end up being mental states; that it is the ability to perceive the world, rather than just respond to it, that is the first condition for the ability to think about the world.

But we still have not explained what makes our thoughts and concepts *about* what they are about; that is, we have not explained what makes them representations rather than something else. We turn to this difficult task now.

3.5 What Makes the Stuff in Our Heads *About* Stuff?

Given the difficulties of this task, it would be extraordinary to hope for a definitive answer here; however, we can learn from others' mistakes and set ourselves in a promising direction. First we need to recognize that to say that a certain state, say S_1, is representational while another, S_2, is not because S_1 one is governed by a certain set of nomic relations while S_2 is not, is not to explain *why* S_1 is representational; rather, it is to explain what being a representation amounts to. Perhaps that is the best we can hope for. Nevertheless, causal accounts such as the one we have been looking at, do not succeed even in explaining what being a representation amounts to.

What Is Wrong With the Causal Story

First, let us recap. Why do causal theorists think that the causal relation could serve as the reference relation in the first place? For some, it is because

they believe that causal relations are information carrying relations. That is, whenever some event or state of affairs is caused by something, say x, that event or state of affairs carries information about x: if a cat caused my 'cat' tokening, then my 'cat' tokening carries information about the cat. The problem of misrepresentation is the problem of explaining how it is that my 'cat' tokening means *cat* in spite of the fact that it was caused by a dog and, consequently, carries information about the dog. Enter the distinctions between Type 1 and Type 2 situations that, as Fodor points out, are unsatisfactory.[8] Or enter Fodor's asymmetric dependence account. The token 'cat' means *cat* even though it carries information about a dog because, if there were no cats around, the dog would not have caused the tokening in the first place; and, cats would cause 'cat' tokenings even if nothing else did. One problem with this idea, however, that causal relations are information relations and that asymmetric dependences between these information relations set up meaning structures, is that it might result in a lot more meaning than Fodor suspects.

For example, consider the following causal relation. The boiling of the water in the pot is caused by the heating up of the element. On this view, the boiling water carries information about the hot element. Now notice that one can also get the water boiling simply by turning on the stove; that is, water boiling is caused by heating up the element *and* by turning on the stove; in the former case, the boiling water carries information about the hot element, and in the latter

case, the boiling water carries information about the turned-on stove. But notice that the latter relationship is asymmetrically dependent upon the former: turning on the stove would not cause the water to boil if heating up the element did not cause the water to boil. Does this mean that the boiling water means *hot element*? (Remember that meaning is determined by asymmetric causal dependence relations: 'x' means *y* just in case nothing would cause a tokening of 'x' unless y's did). Surely water is not the kind of thing that *means* (in the way that our signs do). No it is not, Fodor tells us, because only *robust* asymmetric dependence relations determine meaning. An asymmetric dependence relation is robust just in case it is possible for something besides a y to cause an 'x' tokening, even in the absence of y's. In the case above this does not hold because turning on the stove would never cause the water to boil in the absence of a heated element. As it turns out, Fodor claims, the only kinds of asymmetric causal dependence relations that are robust are the ones that exist between mental states. It all works out then, because meaning is where it should be – in mental states only, not in teacups and pots – while information is everywhere. (Fodor, 1990, 118)

But something about all of this does not seem quite right. Let us look at a non-mental example. The reading on a thermometer can be caused by the temperature of the surrounding air (ceteris paribus). The reading on a thermometer can be caused by shaking the thermometer (ceteris paribus). Now

for the counterfactuals (and here is where we run into problems). Would the thermometer reading be caused by shaking if the temperature-to-reading connection were broken? To figure this out we have to go to the nearest possible world (wherever this may be) and look. Why would the temperature-to-reading connection be broken? Well, perhaps because mercury is no longer sensitive to temperature change. But this would not necessarily affect the shaking-to-reading connection; thus, it is possible for the shaking-to-reading connection to remain while the temperature-to-reading connection is broken. What about the other way around? Would the temperature-to-reading connection be severed if the shaking-to-reading connection were broken? If so, then the temperature-to-reading connection is asymmetrically dependent upon the shaking-to-reading one. Once again, let us look at the possible world in which the shaking-to-reading connection is broken. Shaking would no longer cause a reading if the stuff being shaken is not movable, i.e. not liquid. But if the stuff is not liquid, then the temperature of the surrounding air would not cause it to move either. So it looks like the temperature-to-reading relation is asymmetrically dependent upon the shaking-to-reading relation. Is this relation robust? It seems so because it is possible for the shaking of the thermometer to cause a reading even in the absence of the temperature causing a reading. On Fodor's account, then, temperature readings *mean* thermometer shakings. This is the kind of pansemanticism that he was, reasonably, trying to avoid.[9] Thus,

robust asymmetric causal dependence relations are not enough to explain what it is for a state to be a *representational* state. Presumably we *do not* want to say that a thermometer reading is a representational state – one that means thermometer shakings. But if we do not want to say this, then why do we want to say that, in virtue of its robust asymmetric dependence relations, the mental token 'cat' *is* a representational state – one that means *cat*?

And there are other problems with causal accounts as well. Not all causal accounts are also informational accounts; that is, not all causal theorists hold that the causal relation is the reference relation in virtue of the fact (and others) that causal relations are information carrying relations. Some causal theorists merely contend that "the actual causal relations of a mental state token are what *determine* its content." (Fodor, 1990, 125). Now this is unsatisfactory because either the referential-causal relation is the same as the causal relation referred to in explanations in other sciences (the striking of the match caused it to be lit) or it is different. If it is the same, then it is unclear how the relation can serve as a referential relation in the case of mental tokens when it does not so serve anywhere else (or does the lit match *mean* that it was struck?). If it is different, then it needs explaining – we need to be told how it is different and how it yields intentional states.

Our final, by now familiar, problem with causal accounts affects both informational and non-informational accounts. The causal relationship between

speaker or thinker and the world comes in the form of perception. As we have seen, the perceptual relationship that people have to the world is not a passive. If the causal relation is manifested only when the perceptual relation is present, and if the perceptual relation is representational, then a pre-condition of the referential-causal relation is a representational relationship with the world. But this representational relationship with the world is precisely what the causal account was supposed to explain. Take, for example, the causal relation between cats and 'cat' tokenings. Cats cause 'cat' tokenings via our perception of cats. Our perception of cats is dependent upon a representation or conceptualization of the world that includes cats – in other words, the act of perceiving a cat is the act of representing the world as including cats. Thus, the causal relation between cats and 'cats' depends upon this representational relation between us and the world. But the causal relation was supposed to explain this representational relation, that is, the causal relation was supposed to *be* the representational relation. Something has gone wrong.

I have been arguing here that this sort of causal account is not narrow enough to exclude non-representational states in its explanation and consequently it fails to give a complete characterization of what being a representational state amounts to. The considerations of chapter two indicate that what is missing from the account is the recognition that only intentional systems can have representational states. Thus, baldly stated, the answer to

why the causal asymmetric dependence relation results in a representational state in humans, while it results in a non-representational state in thermometers, is that humans are intentional systems while thermometers are not. But this is precisely what we are trying to explain, one might object: what makes the causal relationship we have with the world yield intentional states while the ones thermometers have with the world yield non-intentional states? Fodor's account is surely on the right track because it tries to account for this difference by appealing to a difference in the level of complexity of the causal relations in question. But this sort of appeal by itself will not do. What we need in addition is the recognition that it is the unique structure of intentional systems themselves that yield unique causal relationships with the world. So the answer to the question 'what makes the stuff in our heads *about* the stuff out there?' will have to include an account not only of what sort of causal relation is necessary for representation (this is the causal part of the relation); it will also have to include an account of the influence our brains exert on the information being carried by these causal relations – this is the conceptual part of the relation. Both parts of the relation need to be present in order for the resultant state to be a representational state.

3.6 *What* Out There is the Stuff in Our Heads About?

Having pointed out what is wrong with Fodor's asymmetric dependence picture, I want to say that there is something right about it as well. Before we can get to that, though, I need to say something about how his account's failure to fully explain what it is for a brain state to be a *representational* state causes it to fall short in explaining how intentional states are individuated as well.

Consider the following example. I want to determine the content of my 'blat' tokenings, that is, I want to find out what *blat* means. Following Fodor, I look to see what properties cause my 'blat' tokenings. I notice that the property *being small and fuzzy* causes 'blat' tokenings, that, in dim lighting, the property *being small and not-fuzzy* causes 'blat' tokenings, that the property *being a certain kind of picture* or *being a certain smell* causes 'blat' tokenings as well. In short, many different properties cause 'blat' tokenings. Now I begin my counterfactual evaluations. Would, in dim lighting, the property *being small and not-fuzzy* cause 'blat' tokenings if the property *being small and fuzzy* did not? I do not know: it might and then again, it might not. All I have to go on is the fact that, in this world, they both cause 'blat' tokenings. I cannot establish a truth value here because I do not know the factor which determines the truth value of the counterfactual, namely, what the reference of *blat* is. And, if I do not know what the primary or fundamental cause of my 'blat' tokenings is, then I also do not know what would or would not cause 'blat' tokenings in various situations.

Fodor manages to evaluate his counterfactuals unproblematically only because, in doing so, he presupposes the very thing he is trying to find out; namely, which collection of properties is causally related to this symbol such that all other causal relations are asymmetrically dependent upon it. I can evaluate the counterfactual 'would the property *being a picture of a cat* cause 'cat' tokenings if the property *being a cat did not*' in the negative *only* if I already know that 'cat' tokenings are about cats. But this is what the account is supposed to help me to find out.

So Fodor's account is only partially explanatory because telling us that tne content of a thought is that thing to which tokenings of the thought bear a certain causal relation would not bring everyone (say non-human intentional systems) closer to understanding what that content is: if we really did not know what the content was, we would not know what 'cat' was picking out in the explanation 'the content of 'cat' is *cat* because cats are causally related to 'cat' tokenings in the requisite way'. The statement *is* explanatory for us, of course, because we do know what 'cat' is picking out (or can be helped along in this by having cats pointed out to us). This is because our basic concepts also play a role in determining *what* the content of our thoughts will be. Meaning *is* reference, we want to agree with Fodor, but reference is not just a world to mind relation; it is also a mind to world relation.

But as I noted above, there seems to be something right about Fodor's asymmetric dependency account as well. We turn again to it to see if it can shed light on the issue of content individuation.

What is Right About Fodor's Asymmetric Account

Fodor's account does seem to be right about one important aspect of the problem of intentionality and that is in specifying the logical structure of the meaning relation. Putting aside for a moment what could stand as the physical manifestation of this relation, it seems right to say that my thought 'x' is about x's just in case, if x's would not cause tokenings of 'x', then nothing would. And, we argued, x's would not cause tokenings of 'x' unless I was capable of perceiving x's, that is, unless I had the concept X. So a thought 'x' is about x's just in case the agent has the concept X and, if x's would not cause tokenings of 'x', then nothing would. Thoughts are individuated, then, according to asymmetric causal/conceptual dependence relations.

But how are concepts individuated? This is trickier. Does the concept CAT depend upon the concept ANIMAL, for example? It is not clear that it has to since one could imagine having all sorts of different concepts corresponding to different animals but no concept that generalizes over these. Does the concept ANIMAL depend upon the concept CAT? Clearly not since there are no doubt a number of animals out there that we do not have concepts for and yet we have the concept ANIMAL. But perhaps there are other concepts that the concept

CAT does depend upon. Now this idea of concept dependence raises the spectre of holism. Holism, however, at least some form of it, need not make an account of concept individuation untenable. We turn to a suggestion of how this might go now.

Concept Individuation

Meaning holism is the view that meanings or contents cannot be specified in isolation from other meanings and contents because what one word means, or what the content of a specific mental state is, is partially determined by the meanings of other words and mental states. For example, what mental state a person is in when he thinks 'a cat is a small, mainly carnivorous animal' depends upon what he means by 'small', 'animal', and so on. Fodor characterizes meaning holism as a commitment to the view that "the intentional content – of a propositional attitude is determined by the *totality* of its epistemic liaisons" (Fodor, 1987, 56) where an epistemic liaison is a connection between intentional contents based on what the intentional agent deems is relevant to understanding one or the other content. This sort of conclusion about meaning is a threat to the programme of naturalization because, on this view, it looks highly unlikely that any two people will ever share a single belief. Any belief an individual has depends upon the various other beliefs that individual has – just which ones is determined by which ones the believer thinks are relevant to the belief in question. But since it is probable that the set of epistemic liaisons of one

individual will not exactly overlap with the set of epistemic liaisons of another, and since mental content is determined by the totality of one's epistemic liaisons, it is probable that no two individuals will ever be in the same intentional state. But if no two individuals are ever in the same intentional state, then we can say good-bye to a science of psychology because, in order to have that, one must be able to generalize over intentional states.

Fodor is a semantic atomist, then, not because he thinks semantic atomism is particularly right, but because he thinks that the only alternative, semantic holism, since it makes psychology impossible, must be hopelessly wrong. This is not just wishful thinking. To suppose that a science of psychology is impossible (in the sense that holism renders it impossible) is to suppose that our ordinary citations of beliefs and desires and our generally fairly successful ability to predict behaviour on this basis is just completely misguided. It is completely misguided if holism is right because unless we are able to subsume our predictions under general principles, they will be nothing more than guesses (uninformed ones at that). And *this* conclusion, that this just seems plain crazy, Fodor must be right about.[10]

Fodor thinks that the solution must be to deny meaning holism. He points out that people have been notoriously sloppy about arguing for meaning holism, taking the route of presupposition instead, and suggests that, intuitions aside, there must be another way. Now it is not hard to see why our intuitions point so

strongly in holism's favour: we can see by simple introspection how entangled our beliefs are with one another. This is why Fodor's atomism is so implausible in spite of the excellent reasons he has for rejecting holism. My suggestion here is that we do not have to "recoil" to either extreme; neither holism, as characterized by Fodor above, nor atomism need be a recourse.

On the approach to understanding meaning I have been advocating here, meaning is holistic; but, because so many of our thoughts depend upon only basic concepts, concepts that we all share, this holism may not pose the threat that Fodor imagines it does. Here is a sketch of a promising direction for a new way of understanding holism.

On the one hand, our ordinary folk psychological predictions are so successful not because each content can be specified atomistically, but because, for a large range of thoughts, the epistemic liaisons relevant to determining them, are ones that all humans share. On the other hand, the reason that our folk psychological predictions are not always successful, and have their limitations, is because some of our thoughts, the ones that depend upon higher level secondary concepts, for example, have epistemic liaisons that are unique to individuals.

Thus we might revise holism to this view: all beliefs are determined partially by the totality of their epistemic liaisons; but what these epistemic liaisons are is not always, and not always completely, up to the agent. Some

liaisons, such as the ones between our basic concepts and our perceptual beliefs, are the same for everyone. Since our perceptual beliefs serve to underwrite a large number of our other beliefs, there is reason to expect that a number of our other beliefs and secondary concepts will be shared as well. With respect to concept individuation, this might mean that all of our concepts are asymmetrically dependent upon a small group of basic concepts, some directly and some indirectly. For example, the concept CAT might be asymmetrically dependent upon basic concepts such as MOVING, ENTIRELY WITHIN FIELD OF VISION,[11] and so on, while the concept ANIMAL might be asymmetrically dependent upon the concepts CAT, DOG, HORSE, and so on, taken together. Thus, while both the concept CAT and the concept ANIMAL are asymmetrically dependent upon a set of basic concepts, we can individuate them according to how directly they are based on those concepts. In this example, which may or may not reflect the actual hierarchy, CAT is directly dependent upon the basic concepts while ANIMAL is indirectly dependent upon them via a set of directly dependent concepts.

Both thoughts and concepts, then, may be individuated according to the asymmetric dependence relations that govern them. We now answer the question of this section – *what* in the world is the stuff in our heads about? – in this way. A thought 'x' is about x's just in case the agent has the concept X and, if x's would not cause tokenings of 'x', then nothing would. An agent's concept X

is basic if it does not depend upon any other concepts. An agent's concept Y is a secondary concept, less fundamental than concept X (which may be either basic or secondary), if it is either directly or indirectly asymmetrically dependent upon X; that is, if the agent could not have concept Y unless it had concept X, but could have concept X even in the absence of concept Y.

3.7 Two Traditional Problems and Their Solutions

Much of what we have been talking about in this chapter has been represented in terms that seem more properly to belong to discussions about mind than to discussions about meaning. This is as it should be, of course, if we agree with the suggestion made at the end of chapter one, that in order to understand meaning in general, we must first come to understand speaker meaning; one cannot explain speaker meaning without talking about speaker thoughts, issues of mind creep in. In this section, I would like to turn to two areas of debate, traditionally formed in philosophy of language, in order to illustrate how the causal/conceptual approach can be applied to these problems.[12] Before I do, though, I want to address a comment I made in my introduction to chapter two. There I claimed that the new approach being developed here would allow us to maintain a thesis of rigid designation while ridding ourselves of the undesirable essentialist metaphysical commitments that are sometimes thought to attend that view. Here is how.

Rigid Designation

In order to use a name to rigidly designate an individual on the causal/conceptual view, one must first be equipped with the conceptual architecture that allows us to pick out that individual in this world and in other possible worlds. Now that conceptual architecture, remember, not only allows us to pick out individuals, it also, in part, determines what is going to count as an individual. If the conceptual architecture involved in determining a particular individual is all basic, then our intuitions about which properties are essential to the individual and which are not will be in general agreement. That is why some examples, such as Kripke's suggestion that origin is an essential property of all human beings, are so compelling. On the other hand, if the conceptual architecture involved in determining a particular individual is not all basic, then our intuitions about which properties are essential to the individual and which are not will be in disagreement. (Consider, for example, the different possible reasonable answers to the question 'what essential properties must a thing have to be a fruit?') Thus, on this view, to say that individuals are distinguished according to certain essential properties they possess is not to make a metaphysical claim; it is merely to say that our concepts are individuated according to those properties.

We now turn to our two issues in the philosophy of language.

Attributive and Referential Uses of Descriptions

We have been distinguishing throughout between thoughts and concepts – concepts are *ways* in which we think about the world (they both enable and limit our thoughts) while thoughts are *what* we think about the world. We have also noted that one source of confusion with respect to understanding our thoughts and the relations between them is that, while concepts underlie our thoughts, they are also expressible *as* thoughts, as in 'people who kill other people are insane.' We talk like this often. Indeed, perhaps the majority of our speech is concept expression. And this is not surprising. When we express and thereby attempt to communicate our concepts in this way, we are attempting to communicate to others the way in which we see things. Keith Donnellan (1966) recognized this use of language when he argued that not all definite descriptions are used in the same way; some are used in what he called a referential sense and others are used in what called an attributive sense.

In referential uses, the definite description is intended to pick out an individual in the world; whether or not the definite description correctly describes the individual in question is not important so long as the description manages to pick out for the interlocutors the right individual. In attributive uses, on the other hand, it is the description or the attributes of the individual that is of importance, not the individual himself.

Now this seems right except that, *prima facie*, it leaves room for situations where, in some attributive cases, there is no reference. This can happen because in attributive cases, the speaker is using a definite description in order to bring attention to the *description* rather than to the thing being described (as in referential cases); thus, if nothing meets the description, there will be a failure to refer as well. For example, the teacher who (attributively) exclaims 'everyone who came in late today will serve a detention after school' when all the kids were on time, has not succeeded in referring. Cases such as this one seem to pose a problem to a theory of meaning that takes meaning to be reference because if there is no reference, there is no meaning either. But this is not a problem for the causal/conceptual account I have been advocating here. To see why not, we have to remember that, on this view, while concepts serve to determine the way we think, concepts can also be thought *about*, that is, they can be expressed as thoughts. In using definite descriptions attributively,[13] then, an individual is expressing a concept and is, thus, referring to that concept. Of course, not all attributive uses of definite descriptions are purely attributive uses; in some cases, a speaker intends to bring attention to the description *and* pick out the individual that meets that description. In such cases, since the speaker is saying two different things with one sentence, there are, correspondingly, two references.

Consider one of Donnellan's examples. Jones is accused of murdering Sam. Fred, a spectator in the courtroom, turns to Sally and whispers, 'Sam's

murderer is insane,' intending by this to convey the thought that Jones is insane (Fred is previously acquainted with Jones). Here Fred is using the definite description referentially; 'Sam's murderer' is referring to Jones.

Now suppose that, in saying 'Sam's murderer is insane,' Fred intends instead to communicate the thought that anyone who could kill Sam is insane. Fred is using the definite description 'Sam's murderer' attributively here because he is not interested in talking about the individual who killed Sam, namely Jones. In this case, Fred is expressing a concept. He is describing the way in which he thinks about the world, namely, in a way that groups people who do things like this into a certain category. The reference of *this* statement then is a concept. Determining just which concept is the referent in question, is determining where it is placed in the conceptual asymmetric dependence hierarchy.

Finally, suppose that, in saying 'Sam's murderer is insane,' Fred intends instead to communicate the thought that anyone who could kill Sam is insane *and* the thought that Jones is insane (perhaps *because* he falls into the category captured by the concept Fred is expressing). In this case, Fred is both referring to a concept and referring to Jones.

Attributive uses of definite descriptions in which there is a seeming failure to refer do not pose a problem for the causal/conceptual approach to meaning, then, because, on this view, all definite descriptions refer – in some cases they refer to objects and in others they refer to concepts.

Sense and Reference Revisited

Since we began our discussion in chapter one by looking at Frege's argument for a distinction between sense and reference, it seems fitting that we end by examining how what we have argued for here meets these Fregean concerns. Since we agree with causal theorists that meaning is reference, it seems that Frege's challenge poses a problem for the causal/conceptual account. To see why it does not, we will present three variations of the problem and provide a solution to each in turn.

First Variant

If reference is all there is to meaning, then the logical structure 'a=a' would be an accurate representation of identity statements between words and expressions that have the same reference. But this is almost never the case. Sentences of the form 'a = a' are tautologous, known a priori, and are pretty uninteresting, while sentences like 'the morning star is the evening star' better fit the form 'a = b' because they are not tautologous, are not known a priori, and some people find them quite interesting. But if 'a=a' represents the reference claim of the statement while still failing to capture its meaning, then there must be more to meaning than reference.

Second Variant

If reference is all there is to meaning, then all terms that have the same reference are synonymous and, consequently, can be used interchangeably

salva veritate. This turns out not always to be the case, however. For example, it is possible for the sentence 'Fred thinks that the morning star is beautiful' to be true and the sentence 'Fred thinks that the evening star is beautiful' to be false. If meaning is reference, then the definite descriptions 'the morning star' and 'the evening star', since they both take the same reference, are synonymous and, consequently, ought to be intersubstitutable *salva veritate*. But they are not. Thus, there must be more to meaning than just reference.

Third Variant[14]

Intentional states are characterized by their mode and content. If two intentional states differ, it must be in one or both of these respects. If Fred believes that Venus is the morning star but does not believe that Venus is the evening star, then believing the former must constitute being in a different intentional state from being in the latter. Since they are both belief states, they do not differ in mode; therefore, they must differ in content. If they differ in content, then the sentences 'Venus is the morning star' and 'Venus is the evening star' express different propositions. If the sentences express different propositions, the descriptions 'the morning star' and 'the evening star' must not be synonymous. Since they are not synonymous and yet have the same reference, there must be more to meaning than reference.

Causal/Conceptual Response

The solution to these problems is to once again recognize, as Donnellan has pointed out, that definite descriptions can be used both referentially and attributively (where attributive uses are to be understood in the way I explained above, as expressions of and, consequently, reference to, concepts).

Consider that the descriptions 'the morning star' and 'the evening star' can be used either referentially or attributively. If Fred says 'the morning star is the second planet from the sun', intending by this to refer to Venus, he is using the description 'the morning star' referentially. In this case, we can substitute 'the evening star' with 'Venus' without changing the meaning of his utterance. On the other hand, if, when Fred says 'the morning star is beautiful', he is talking attributively, then he is expressing the way in which he thinks about the morning sky, namely, as containing a beautiful star. It is not acceptable to substitute 'the evening star' for the description here because 'the evening star' refers to a *different* concept, namely one about a star in the night sky. The confusion arises because sometimes we make statements about the world directly, and at other times, we make statements about the way we think about the world. In cases where we are doing the former, using descriptions referentially, they *are* intersubstitutable *salva veritate*; in cases where we are doing the latter, using descriptions attributively, they *are not* intersubstitutable *salva veritate*.

To make this even clearer, suppose, very simplistically, that the concept VENUS includes the concepts MORNING STAR and EVENING STAR. The concept MORNING STAR might include BEING VISIBLE IN THE MORNING SKY and the concept EVENING STAR might include BEING VISIBLE IN THE EVENING SKY. When we use any one of these descriptions referentially, since they all overlap at one concept, namely the concept VENUS, they each serve to pick out the same aspect of the world; consequently, we can interchange them with one another *salva veritate*. When we are using any one of these descriptions attributively, however, we cannot interchange them *salva veritate* because, when using them attributively, we are expressing a concept (referring to it), and each of these concepts is different from the other. Thus, if Fred attributively says 'the morning star is beautiful', then even if he also holds the concept VENUS, he may still refuse to assent, without inconsistency, to the sentence 'the evening star is beautiful'.

We can deal with each characterization of the problem, specifically, in the following ways.

Solution to First Variant

When the descriptions are being used referentially, the statement 'the morning star is the evening star' exemplifies the form 'a=a'; when the descriptions are being used attributively, the statement exemplifies the form 'a=b'. And our surprise at discovering that 'the evening star' is 'the morning star' is to be

expected. It *is* a discovery. We are discovering that two different concepts serve to pick out one aspect of the world.

Solution to Second Variant

If it is possible for the sentence 'Fred thinks that the morning star is beautiful' to be true and the sentence 'Fred thinks that the evening star is beautiful' to be false, then the descriptions 'the morning star' and 'the evening star' are not synonymous, and hence have different references. The descriptions have different references if one or both are referring to the concepts MORNING STAR and EVENING STAR respectively. If the descriptions have different references, then the fact that they are not intersubstitutable *salva veritate* in these cases does not pose a threat to the view that meaning is reference.

Solution to Third Variant

If Fred believes that Venus is the morning star but does not believe that Venus is the evening star, then it is true that being in the former state constitutes being in a different state from being in the latter state. Since they are both belief states, the two must differ in content. Since they differ in content, the sentences 'Venus is the morning star' and 'Venus is the evening star' must express different propositions. If the sentences express different propositions, then the descriptions 'the morning star' and 'the evening star' must not be synonymous. If they are not synonymous, then they must have different references. They can have different references if either 1) they refer to different concepts (as when 'the

morning star' refers to MORNING STAR and 'the evening star' refers to EVENING STAR), or 2) one refers to Venus and the other refers to a concept (as when 'the morning star' refers to Venus and 'the evening star' refers to a concept). So Fred's believing that the morning star is the evening star might consist in his believing that the object picked out by the concept MORNING STAR is the same object that is picked out by the concept VENUS, while Fred's believing that the evening star is not Venus might consist in his believing that the concept EVENING STAR does not include the concept VENUS. In the first case the content of his belief is an object and in the second case the content of his belief is a concept. This is not a counter-example, then, to the view that meaning is reference.

Conclusion

I said in my introduction that the motivation for this project was to bring together two apparently conflicting intuitions: the first is the observation that the only way to really explain our semantic and intentional notions is by providing an account of them in non-semantic and non-intentional terms; the second is the conviction that all of our observations are influenced by the way in which we see the world and, consequently, are not objective. The causal/conceptual approach to meaning I have been arguing for here accommodates both of these thoughts by presenting a naturalistic account of meaning that acknowledges the intentional nature of our perceptions; it allows us to hold on to the hope for a non-intentional account of intentionality and, at the same time, maintain that our observations are conceptually guided. Affirming the latter does not jeopardize the non-intentionality of the former; holding this view entails only that none of our theories about the world are objective (because they are founded on intentional observations) and *not* that none of our theories are naturalistic (in the sense that they do not appeal to intentional notions). Thus, a non-intentional account of intentionality that never loses sight of the intentional nature of its subject is possible. Such a theory, I have argued, must, in addition to explaining the reference relation and the individuation of mental states by means of a causal/conceptual relation, provide a non-intentional account of what it is to be an intentional system.

In section 3.7, I looked at two ways in which this new approach can be used to explain traditionally vexing problems in the philosophy of language. My suspicion is that a great number of these problems stems from the impulse, when engaging in these discussions, to ignore the role of concepts in determining meaning. No doubt the fact that these discussions are typically presented in terms of sentences, propositions, and beliefs is partly to blame. My hope is that the causal/conceptual approach will present avenues for recasting these old problems in new, perhaps, fruitful ways and will open the door to studying aspects of language, metaphors for example, that have had little attention.

Endnotes

Notes to Introduction

1 For example, see Ruth Millikan's discussion in Millikan (1986).

2 For example, see Jerry Fodor's discussion in Fodor (1990).

Notes to Chapter One

1 Frege discusses the following in the first page or so of Frege (1892).

2 In chapter three, I shall be presenting an account that does give such an explanation. Then it shall become clearer what sense is, but, I hope, it shall also be clear why the distinction between sense and reference is not necessary.

3 I will use the convention of presenting a term in boldface to indicate a sense.

4 This is Terry Tomkow's (1995) example.

5 See his discussion in Russell (1905).

6 Bernard Linsky and Edward Zalta (1995) defend a version of Platonism, what they call Platonized Naturalism, that does not fall prey to the traditional problems encountered by Platonistic positions. The second problem, however, that of explaining how we identify universals outside of language, still remains.

7 Strawson (1950).

8 Donnellan (1966).

9 Austin (1962).

10 Grice (1957).

11 Where the answer expected is context-free and unchanging.

12 Grice (1957).

13 In "Meaning" Grice claims that the term 'means' in the context 'A means something by x' is a non-natural use of the term while in the context 'A means *to do* so-and-so (by x)', 'means' is being used in a natural way. Since, under the Gricean analysis, the former can be explained in terms of the latter, i.e. where 'means *to do* so-and-so' becomes

'intends to produce a certain effect in the audience', speaker meaning (the former) will have been reduced to a form of natural meaning. (Grice, 1957, 378)

14 I will follow Grice's terminology by using 'mean' to indicate natural meaning and 'mean$_{NN}$' to indicate non-natural meaning.

15 Or normally. Grice is vague on this point.

16 I have not been able to find an argument in Grice to support this move.

17 In personal communication. The example I use is thanks to Bob Martin, also in personal communication.

18 Even were we to abandon Gricean analysis, this question would still need to be answered since, unless we are behaviourists or eliminativists, whatever theory we have, we must address the role that speaker intention -- what we mean by our words, what we use them to do, etc. -- plays.

19 In Kripke (1972 and 1979).

20 Here, for example, is why Fodor rejects Dretske's account. Since, in informational accounts, it is both the actual and the counterfactual tokenings in training situations that go toward determining the content of a symbol, then either 1) it is possible to have an erroneous tokening in the learning situation, in which case the content of the ideal tokening will be a disjunction and we will have the disjunction problem to deal with again; or, 2) it is not possible to have an erroneous tokening during the learning situation, in which case one wonders how it is possible to make a mistake at any other time.

Notes to Chapter Two

1 This intuition is not universally held, of course, but it is the one we're *supposed* to have about the story.

2 If you are a conventionalist, insert here 'essential property by convention'.

3 Of course, the very point of this thought experiment is to show that 1) our intuitions take it that mental contents should not be narrowly construed and that 2) if this is the case then the supervenience relationship between mental states and brain states is not as straightforward as one might have supposed.

4 Dretske has worked out this theory of perception in Dretske (1969).

5 I will follow Dretske and focus on vision here but, of course, what is said goes for all of our perceptual abilities.

6 I say only possibly here because, as I shall argue, when humans non-epistemically see we are still cognizing, that is, this is still an intentional act.

[7] Dretske has much to say about the fact that adults almost never just non-epistemically see anything because our background beliefs play such a pervasive role in our lives. (See sections III, V, and VI of Dretske (1969) for his discussion of this and related issues.) But he believes that such perception is *possible* and probably occurs to a great degree early on in life during the language learning stage.

[8] These small movements of the eye are called microsaccades, small jumps. Here is Hubel's description: "Evidently, microsaccades are necessary for us to continue to see stationary objects. It is as if the visual system, after going to the trouble to make movement a powerful stimulus -- wiring up cells so as to be insensitive to stationary objects -- had then to invent microsaccades to make stationary objects visible." (Hubel, 81)

[9] This is not quite right because an object can cause tokenings in a variety of other ways as well, e.g. via our perception of other objects, via our perception of its effects, and so on. But causal relation via direct perception is the simplest instance of the referential relation, and so all accounts of meaning have to get this part of the story right before they can move on to more difficult problems.

[10] Note that while such an account of intentionality could complement the approach to meaning I am espousing here, it could also be incompatible with it if embedded into it is the supposition that intentional systems gather information or discriminate between aspects of their environment in non-intentional ways as well (ways that are functionally described differently of course). This is incompatible with what I have been saying since my claim is that acts such as gathering information or discriminating between aspects of one's environment are intentional acts.

[11] See Wright (1973) for more on this.

[12] Gibbard (1990) suggests that systems that humans use to represent, such as computers, can be seen as systems of artificial representations, p. 108.

[13] This is John Searle's (1995) example.

[14] He develops a view that connects intentionality with consciousness while I hold that a system may be intentional and yet not be conscious, e.g. cats may not be conscious but they are most certainly intentional systems.

[15] Wilfrid Sellars (1963) presents a comprehensive attack on the myth of the Given along these lines.

[16] Specifically in Thagard (1990).

Notes to Chapter Three

1 I shall be talking about the foundation of concepts under this heading. Now, we might think that some concepts are innate, not acquired. But I shall not address the issue of innateness here. This is hairy terrain and, either way, does not affect the conclusions we want to draw here.

2 The scope of a set of concepts is not always this broad of course. A certain set of concepts may only give one a way of understanding a particular situation. But some of our concepts do shape the way in which we understand our perceptual world. It is this set that I will be focussing on.

3 Of course, if we do not choose the basic concepts we see the world through, that is, if they are innate, then we do not need an explanation here. On the other hand, we may still like to have an explanation of why *these* concepts are innate.

4 Needs would probably be a better term here since here 'interest' is really shorthand as I described above; however, for the sake of the distinction to follow, I am going to refer to these as basic interests and the other, real sort, as secondary interests.

5 These interests I call 'basic' because they are the interests we need to serve before any other interests become possible.

6 David Braybrooke (1987) has compiled a list of what he calls 'matters of need'. His list, which includes headings such as 'The need to have a life-supporting relation to the environment' and 'The need for food and water' bears resemblance to the examples I provide; however, I am not interested in providing a list of basic interests here. I suggest what follows only to give an example of the sorts of interests I have in mind. When I do make use of basic interests here I do so only for the sake of demonstrating what role they play; nothing hinges on it being this or that interest.

7 I say 'by and large' because some people seem to be born with an inability to conceptualize, that is perceive the environment, in the way in which most people do. Autism might be an example of this.

8 See my section 1.4 for this discussion.

9 We want to avoid pansemanticism because we are trying to explain what being a representational state amounts to. If the necessary and sufficient conditions for being a representational state are met by non-representational states as well, then our explanation is too broad and we have not succeeded in explaining what we set out to explain.

10 For Fodor's discussion of this see chapter three of Fodor (1987).

[11] I am not suggesting this as a list of basic concepts. As I mentioned earlier, determining *what* that list consists in belongs to a different project altogether. I provide these only to aid in illustrating what the structure might look like.

[12] My treatment of these two issues is intended only as an example of how a causal/conceptual approach to meaning can open up new ways in which we might address these problems, rather than as a complete answer to these quandaries. The job of filling in the sketch I provide here I will leave to another occasion.

[13] I think that calling these uses referential and attributive is misleading actually because it leads one, falsely, to suppose that attributive uses are not referring uses. Since I think that meaning is reference, I hold that all uses of terms are referential. But we need to make *some* distinction, because in some cases, as in attributive ones, the kind of thing being referred to is different; however, because it is familiar to everyone, I will stick with this terminology. (My terminological preference, for the record, would be objectual use as distinguished from conceptual use).

[14] Fodor (1990) presents this variation of the problem in chapter six (specifically, p. 166).

Bibliography

Alston, W. 1963. "Meaning and Use": *Philosophical Quarterly* 51, 107-124.

Austin, J. L. 1962. *How to Do Things With Words.* Cambridge: Harvard University Press.

Bestor, Thomas. 1991 "Naturalizing Semantics: New Insight or Old Folly?": *Inquiry* 34, 285-310.

Blackburn, S. & Code, A. 1978. "The Power of Russell's Criticism of Frege": *Analysis* 38, 65-77.

Braybrooke, David. 1987. *Meeting Needs.* Princeton: Princeton University Press.

Burge, T. 1979. "Individualism and The Mental", 393-446: Harnish, R. (ed.) *Basic Topics in the Philosophy of Language.* Englewood Cliffs: Prentice-Hall.

Carney, J. & Fitch, G. 1979. "Can Russell Avoid Frege's Sense?": *Mind* 88, 384-393.

Churchland, Paul. 1995. *The Engine of Reason, The Seat of The Soul.* Cambridge: MIT Press.

Churchland, Paul. 1997. "On the Nature of Theories: A Neurocomputational Perspective", 251-292: Haugeland, J. (ed.) *Mind Design II.* Cambridge: MIT Press.

Cummins, Robert. 1975. "Functional Analysis": *Journal of Philosophy* 72(20), 741-765.

Cummins, Robert. 1991. "The Role of Mental Meaning": McLaughlin, Brian (ed.) *Psychological Explanation in Dretske and His Critics.* Cambridge: Blackwell Press.

Davidson, Donald. 1967. "Truth and Meaning": *Synthese* 3(17), 304-323.

Davidson, Donald. 1973-74. "On The Very Idea of a Conceptual Scheme": *The Proceedings of the American Philosophical Association* 47, 5-20.

Davidson, Donald. 1977. "The Method of Truth in Metaphysics": *Midwest Studies in Philosophy* 2, 244-254.

Dennett, Daniel. 1990. "The Interpretation of Texts, People and Other Artifacts": *Philosophy and Phenomenological Research*, 177-194.

Dennett, Daniel. 1997. "True Believers: The Intentional Strategy and Why it Works", 57-80: Haugeland, J. (ed.) *Mind Design II.* Cambridge: MIT Press.

Devitt, M. & Sterelny, K. 1989. *Language and Reality: An Introduction to the Philosophy of Language.* Cambridge: MIT press.

Dretske, Fred. 1969. *Seeing and Knowing.* Chicago: The University of Chicago Press.

Dretske, Fred. 1981. *Knowledge and the Flow of Information.* Cambridge: MIT Press.

Dretske, Fred. 1995. "Does Meaning Matter", 5-17: Villaneuva, Enrique (ed.) *Information, Semantics, and Epistemology.* Cambridge: Blackwell.

Donnellan, Keith. 1966. "Reference and Definite Descriptions": *Philosophical Review* 75, 281-304.

Dreyfus, Hubert. 1997. "From Micro-Worlds to Knowledge Representation: AI at an Impasse", 143-182: Haugeland, J. (ed.) *Mind Design II.* Cambridge: MIT Press.

Dummett, Michael. 1974. "The Significance of Quine's Indeterminacy Thesis": *Synthese* 27, 351-397.

Dummett, Michael. 1976. "What is a Theory of Meaning?": Evans, G. & McDowell, J. (eds.) *Truth and Meaning: Essays in Semantics.* Oxford: Clarendon Press.

Egan Frances. 1994. "A World Without Mind: Comments on Terence Horgan's 'Naturalism and Intentionality'", 327-337: *Philosophical Studies* 76(2-3).

Frege, G. 1892. "On Sense and Reference": Geach, P. & Black, M. (trs.) *Translations from the Philosophical Writings of Gottlob Frege.* Oxford: Basil Blackwell, 1960.

Frege, G. 1918. "The Thought: A Logical Inquiry" (A.M & M. Quinton trs.): *Mind* 65, 1956, 289-311.

Fodor, Jerry. 1968. *Psychological Explanation: An Introduction to the Philosophy of Psychology.* New York: Random House.

Fodor, Jerry. 1974. "Special Sciences, or The Disunity of Science as a Working Hypothesis", 120-133: *Synthese* 28.

Fodor, Jerry. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge: MIT Press.

Fodor, Jerry. 1990. *A Theory of Content and Other Essays* Cambridge: MIT Press.

Fodor, J. A. & Pylyshyn, Z. W. 1997. "Connectionism and Cognitive Architecture: A Critical Analysis", 309-350: Haugeland, J. (ed.) *Mind Design II.* Cambridge: MIT Press.

Geach, Peter. 1950. "Russell's Theory of Descriptions": *Analysis* 10, 84-88.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment* Cambridge: Harvard University Press.

Grice, H. P. 1957. "Meaning", 436-443: *Philosophical Review* 66(3), reprinted in Rosenberg, J. & Travis, C. (eds.) *Readings in the Philosophy of Language.* Englewood Cliffs: Prentice-Hall, 1971.

Grice, H. P. 1969. "Utterer's Meaning and Intentions": *Philosophical Review* 78, 147-77.

Hacker, P.M.S. 1996. "On Davidson's Idea of a Conceptual Scheme": *Philosophical Quarterly* 46, 289-307.

Haldane, John. 1992. "Putnam on Intentionality": *Philosophy and Phenomenological Research* 52(3).

Haugeland, John. 1997. *Mind Design II.* Cambridge: MIT Press.

Honderich, Ted. 1969. "On the Theory of Descriptions": *The Proceedings of the Aristotelian Society* 69, 87-100.

Horgan, Terence. 1994. "Naturalism and Intentionality", 301-325: *Philosophical Studies*, 76(2-3).

Horgan, Terence. 1994. "Reply to Egan's 'A World without Mind'", 339-347: *Philosophical Studies*, 76(2-3).

Horwich, Paul. 1995. "Meaning, Use, and Truth": *Mind* 104(414), 355-368.

Hubel, David H. 1988. *Eye, Brain, and Vision.* New York: Scientific American Library.

Kaplan, David. 1975. "How To Russell a Frege-Church": *Journal of Philosophy* 72, 716-729.

Katz, J. & Fodor, J. 1963. "The Structure of a Semantic Theory": *Language* 2(34), 170-210.

Kaye, Lawrence. 1993. "Are Most of Our Concepts Innate?": *Synthese* 95, 187-217.

Kripke, Saul. 1972. *Naming and Necessity.* Cambridge: Harvard University Press.

Kripke, Saul. 1979. "A Puzzle About Belief", 352-392: Harnish, R. (ed.) *Basic Topics in the Philosophy of Language.* Englewood Cliffs: Prentice-Hall.

Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language.* Cambridge: Harvard University Press.

Koethe, John. 1992. "And They Ain't Outside the Head Either": *Synthese* 90(1), 27-53.

Levine, James. 1993. "Putnam, Davidson and the Seventeenth-Century Picture of Mind and World", : *International Journal of Philosophical Studies* 1(2), 193-227.

Linsky, Bernard. 1993. "Why Russell Abandoned Russellian Propositions", 193-209: Irvine, A. & Wedeking, G. (eds.) *Russell and Analytic Philosophy.* Toronto: University of Toronto Press.

Linsky, B. & Zalta, E. 1995. "Naturalized Platonism versus Platonized Naturalism": *The Journal of Philosophy* 92, 525-555.

Martin, R. 1987. *The Meaning of Language.* Cambridge: MIT Press.

McDowell, J. 1994. *Mind and World.* Cambridge: Harvard University Press.

Millikan, Ruth. 1986. *Language, Thought and Other Biological Categories.* Cambridge: MIT Press.

Millikan, Ruth. 1990. "Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox": *Philosophical Review* 99(3), 323-353.

Mill, J.S. 1843. "Of Names": *A System of Logic.*

Minsky, Marvin. 1997. "A Framework for Representing Knowledge", 111-142: Haugeland, J. (ed.) *Mind Design II* Cambridge: MIT Press.

Morrison, James. 1970. "Husserl and Brentano on Intentionality": *Philosophy and Phenomenological Research* 31.

Nelson, R.J. 1984. "Naturalizing Intentions", 173-197: *Synthese* 61.

Putnam, Hilary. 1975. "The Meaning of Meaning": *Mind, Language, and Reality: Philosophical Papers*, Vol 2. Pp. 215-271. Cambridge: Cambridge University Press.

Putnam, Hilary. 1983. *Realism and Reason.* Cambridge: Cambridge University Press.

Putnam, Hilary. 1990. *Realism With a Human Face.* Cambridge: Harvard University Press.

Putnam, Hilary. 1992. *Renewing Philosophy.* Cambridge: Harvard University Press.

Putnam, Hilary. 1994. "Dewey Lectures": *Journal of Philosophy* 91.

Putnam, Hilary. 1995. *Pragmatism: An Open Question.* Cambridge: Basil Blackwell.

Putnam, Hilary. 1995. "Pragmatism": *The Proceedings of the Aristotelian Society,* 291-306.

Quine, W. V. O. 1960. "Translation and Meaning": *Word and Object.* Cambridge: MIT Press.

Quine, W. V. O. 1969. "Ontological Relativity": *Ontological Relativity and Other Essays.* London: Columbia University Press.

Quine, W. V. O. 1970. "On the Reasons for the Indeterminacy of Translation": *Journal of Philosophy* 67, 178-183.

Quine, W. V. O. 1987. "Indeterminacy of Translation Again": *Journal of Philosophy* 84, 5-10.

Rorty, Richard. 1971. "Verificationism and Transcendental Arguments": *Nous* 5, 3-14.

Rorty, Richard. 1993. "Putnam and the Relativist Menace": *Journal of Philosophy* 90, 443-461.

Russell, Bertrand. 1905. "On Denoting": *Mind* 14, 479-93.

Russell, Bertrand. 1912. *The Problems of Philosophy.* London: Oxford University Press.

Russell, Bertand. 1919. "Descriptions": *Introduction to Mathematical Philosophy.* London: Allen and Unwin.

Russell, Bertrand. 1919. "On Propositions, What They Are and How They Mean": Marsh (ed.) *Logic and Knowledge.* London: Allen & Unwin, 1956.

Salmon, N. 1986. "Frege's Puzzle": *Frege's Puzzle* (Cambridge: MIT Press.

Schiffer, Stephen. 1978. "The Basis of Reference": *Erkenntnis* 13, 171-206.

Schiffer, Stephen. 1986. "Kripkenstein Meets the Remnants of Meaning": *Philosophical Studies* 49, 147-162.

Searle, John. 1978. "Literal Meaning": *Erkenntnis* 13, 207-224.

Searle, John. 1995. *The Social Construction of Reality.* New York: Free Press.

Sellars, Wilfrid. 1949. "Acquaintance and Description Again": *Journal of Philosophy* 46, 496-503.

Sellars, Wilfrid. 1963. "Empiricism and the Philosophy of Mind": *Science, Perception and Reality,* pp. 127-196. London: Routledge and Kegan Paul.

Stampe, D. 1981. "Pragmatics and Causal Theoretic Aspects of Semantics": Parret, H., Sbisá, M., Verschueren, J. (eds.), *Possibilities and Limitations of Pragmatics, Proceedings of the Conference on Pragmatics.* Amsterdam: John Benjamins.

Stampe, D. 1986. "Verificationism and a Causal Account of Meaning": *Synthese* 69, 107-137.

Stampe, D. 1990. "Content, Context, and Explanation", 134-152: Villaneuva, Enrique (ed.) *Information, Semantics, and Epistemology.* Cambridge: Blackwell.

Strawson, P. F. 1950. "On Referring": *Mind* 59, 320-44.

Tarski, Alfred. 1944. "The Semantic Conception of Truth": *Philosophy and Phenomenological Research* 11, 341-75.

Thagard, Paul. 1990. "Concepts and Conceptual Change": *Synthese* 82, 255-274.

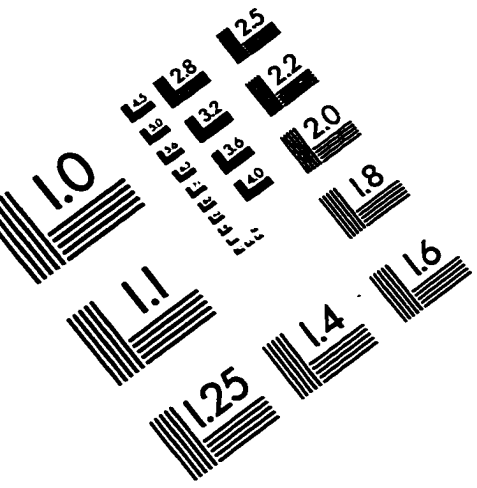
Tomkow, Terrence. 1995. *Against Representation: A Theory of Language and Thought.* Forthcoming.

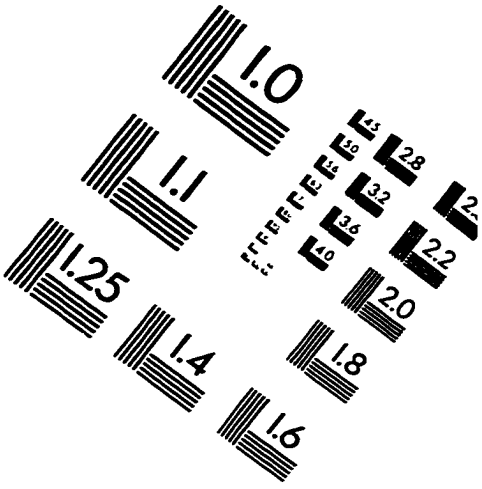
Wagner, Steven J. 1993. "Why Realism Can't Be Naturalized": Wagner, S. & Wagner, R. (eds.) *Naturalism: A Critical Appraisal.* Notre Dame: University Press.

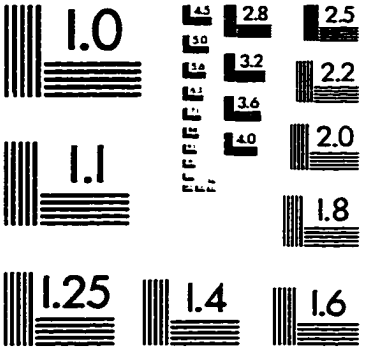
Wittgenstein, L. 1953. *Philosophical Investigations.* New York: Macmillan.

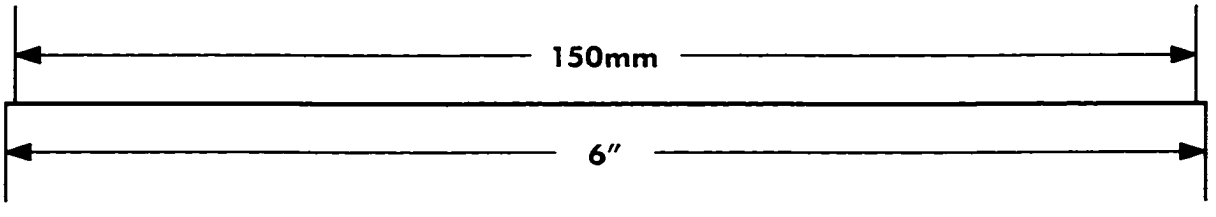
Wright, Larry. 1973. "Functions": *Philosophical Review* 82, 139-168.

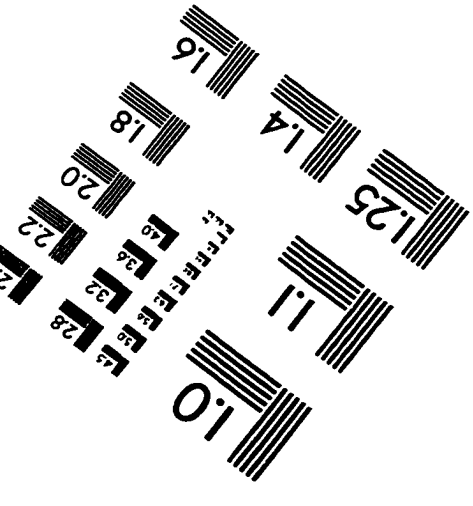
IMAGE EVALUATION TEST TARGET (QA-3)

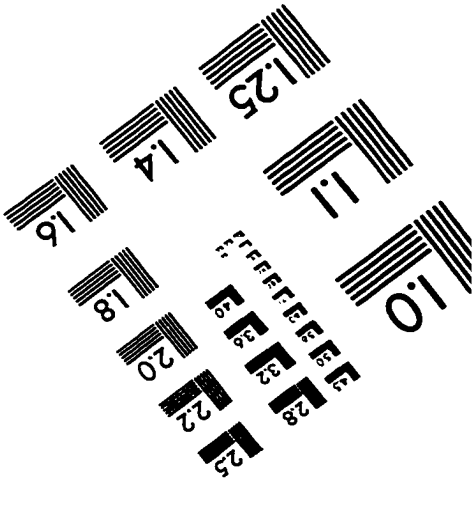












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989