Naturalizing Meaning

Jerry Fodor's Theory of Content

Magnus Stavik Rønning

Thesis Submitted for the Master of Arts Degree,

Department of Philosophy, Classics, History of Arts and Ideas

UNIVERSITETET I OSLO

November 2008

Acknowledgements:

I want to thank my supervisor Carsten Hansen for providing invaluable support and advice. I also want to thank Are Reichelt Føreland, Beate Ravn Øhlckers, Tore Fjetland Øgaard, James Swann and Hallvard Markus Stette for reading an earlier draft and providing helpful comments and suggestions.

Table of Contents

Part II:

Introduction

The question how something like the mind can be purely physical is, in my opinion, perhaps the most fundamental question one can ask. It is a question that gives rise to many more questions about both the nature of the mind and the world, perhaps the most fundamental dichotomy in the human psychology. The question we are interested in answering in this thesis is a sub-question to the aforementioned, namely how can meaning be accounted for so that it is compatible with a physicalistic ontology?

Jerry Fodor has tried to answer this question by investigating the nature of intentionality and representation. And, as we will see, it is by accounting for these notions in constructing a theory of the content of mental states that he aims to account for the question of meaning. I will in this thesis try to state what Fodor's theory of content is, how Fodor's account of content relates to other theories that are similar to his, what Fodor's solution to the disjunction problem is and finally assessing Fodor's theory as a whole.

In the thesis we will focus on three pre-theoretical ideas, or intuitions we will require that Fodor's theory satisfies to successfully have accounted for how meaning can be something physical. One intuition is about the physical, and the other two are about meaning. The intuition that is about the physical is one that is implied by our ontological conviction that everything that exists is physical. This requirement is thus that the theory must account for meaning in a way that is compatible with meaning's being purely physical, and in what follows we will call this the *naturalism requirement*.

The next intuition we will focus on is one that is about the nature of meaning. The intuition, in short, is that meaning is not a feature of things in general. Not everything exhibits meaning, and a theory of meaning that implies that meaning is everywhere is one that is not satisfactory. As we will see in part II, accounting for this intuition takes the form of showing that the theory does not imply what we will be calling *pan-semanticism*, i.e. the view that everything has meaning, or is meaningful. The final, and most important intuition the theory of meaning must respect, is the intuition of the *robustness of meaning*. This is the intuition that, say, the concept CAT means what it does regardless of what causes its occurrence.

As to the structure of the thesis we can note that it is composed of two parts with three chapters in each of the parts. Part I is about what naturalistic theories are and how they usually propose to solve the disjunction problem. The disjunction problem is the main problem in this thesis because it arises from how naturalistic theories try to account for the robustness of meaning. As accounting for the robustness of meaning is, in many ways, to solve the disjunction problem, the disjunction problem is considered in both part I and part II.

Part II is about Fodor's own solution to the disjunction problem. It is in this section that we will see what Fodor's proposal amounts to, and here we will assess it in relation to some objections that have been proposed by Paul Boghossian. How Fodor accounts for the pan-semanticism worry will also be considered in part II. As the title of the thesis indicates it is about Fodor's theory of content. But for us to be able to properly asses Fodor's theory we need to establish several key notions which will serve as the foundation for articulating Fodor's theory. Part I is mainly about providing such a foundation, and Part II is thus the part where Fodor's theory will be assessed.

One can think of Fodor's overall philosophical project as the project of securing a scientific basis for our common-sense psychology. That is to defend the common intuition that beliefs and desires and thoughts are real things that figure in our minds. It is because these things are real that one can say something true when one says "I went to the store because I wanted chocolate". That one wanted chocolate is true, and it is the reason one went to the store. If one thinks that there are no such mental objects as wants then the explanation does not explain anything. Fodor's theory is a commitment to the common-sense view that our theory about ourselves and other are largely correct and worth keeping. This is a view I am inclined to endorse.

This thesis is about a part of Fodor's project of securing a basis for common-sense psychology, namely the project of accounting for mental representation and thereby providing a foundation for meaning.

Last, a word about terminology. I will try to follow existing conventions in formulating this theory. This means that when I mean to refer to a concept I will write it in caption, say, CAT; when I mean to refer to a word or sign I will mention it, say, "cat"; when I want refer to the meaning or content of symbols or terms I will try to put it in italics , say, *cat*. Some mistakes are bound to happen, and for these I apologize.

Part I:

1. Chapter I: *Requirements on a Naturalistic Theory of Meaning*

> Naturalism is the thesis that for God to create our world He needed only to have created the naturalistic entities and laws. Everything else follows from these. (Loewer, 1997, 108).

This chapter is mainly about what features a theory must exhibit for it to satisfy the naturalism requirement. We will consider what relation naturalism has to the physicalistic doctrine, and try to specify a condition that, if satisfied, suffices for the theory to be naturalistic. As we noted in the introduction the theory aims to account for meaning by accounting for representation. The last part of this chapter will be about which naturalistic relation is likely to be able to constitute representation.

1.1. Physicalism

I suppose it is not inaccurate to say that most philosophers today are physicalists in some sense. The main competing view in the philosophy of mind, substance dualism, is not widely held to be true. There are variants of this view, as of any other view, and stronger and weaker commitments one can endorse. Some physicalists are physicalists in a strong sense and only believe in the existence of some or other basic particle type (e.g. quarks) and perhaps some of the forces (e.g. gravity). Others believe not only in micro objects but also in macro objects like mountains, horses and solar-systems. The latter are usually disposed to believe in the laws that govern the things they believe in, but this is not always so.

Physicalism is a view that claims that everything that exists is physical. Fodor's theory of content is physicalistic in the sense that it assumes a basic physicalistic framework, but he sometimes seems to reject certain parts of what physicalism is normally taken to be, something that is observed by commentators such as Barry Loewer and Georges Rey in the introduction to their book *Meaning in Mind* (Loewer and Rey, 1991). We will see later in the chapter what this means. Fodor's views on the mental commit him to be a realist about mental states (Fodor, 1994, p. 3-4). He is committed to that psychological laws are real laws, which in turn imply that the properties they subsume are real properties (Fodor, 1994, p. 3). Since these properties are intentional, Fodor is committed to being a realist about the intentional. The mental states we will be interested in in this dissertation are mental states that are representational, intentional and semantically evaluable. The first two deal with how mental states can be about other things (external objects, other thoughts, etc.). The latter is about how mental states can be true of false on account of the representational properties of themselves or their constituents. One type of mental states we will not be considering is the type of mental state that is commonly referred to as qualitative states, or qualia. In addition to qualia, we will not consider questions about the nature of consciousness.

As physical beings we humans are, in principle, no different from stars and planets. The main problem for the physicalistically inclined philosopher of mind is the problem of making sense of the idea that the mental is not something radically different from the physical, or indeed, nothing but physical. Philosophers who are naturalistically inclined are often people who have great faith in the sciences. They think that science, in general, is mankind's greatest achievement, and, I think, it is hard to disagree with them. So it is not surprising that it is often these philosophers who are interested in trying to account for the mind as part of nature. Physics is often assumed to have a key importance to philosophy. Physics is often taken to be a sort of default ontology where the sciences "bottom out". So, to have a theory about something that in principle is incompatible with physics is, for a physicalist, to have a serious problem, and a theory that is incompatible with the sciences is a theory that is not naturalistic in the sense we are after. As we will see, one often assumes that the naturalization project and making a theory compatible with the sciences is one and the same project.

1.2. Naturalism

What exactly are naturalistic theories? What properties does a theory need to have to pass the tests of being naturalistic? When it comes to being the paradigm case of a naturalistic methodology, physics is it (Fodor, 1994, p. 5). Making a theory of the mind, or parts of the mind that is obviously compatible (when this means reducible) with physics is perhaps a tall order, but philosophy is not the only academic discipline that has trouble with reduction. Almost all the sciences, from chemistry through biology and to psychology and the social sciences have problems with reducing their theoretical terms to physical terms. There are several reasons for this. One reason is that even if one has established that some causal relation is a law, one has not thereby specified by what mechanism the law is implemented.

Let us take the Müller-Lyer illusion as an example. It is, we can assume, a psychological law that humans experience the lines as having different lengths when they in fact are the same length. What is the implementing mechanism in this case? It is presumably some neurological mechanism, but what it is and how it works, are not known. The failure to reduce its laws does not prevent psychology from being a science, and being one of the sciences its theories are naturalistically acceptable. Paul Boghossian (1991, p. 68) says, for instance, in a different context, that to specify something in terms of evolutionary biology is assumed to be sufficient for being naturalistic, and evolutionary biology is a special science in the required sense. Fodor says this about naturalizability in *The Elm and the Expert* (1994):

> … naturalizability, in this broad sense [i.e. not being specifically a demand upon intentionality], is a general constraint upon the ontology of all the special sciences. It's a methodological consequence of our conviction – contingent, no doubt, but inductively extremely well confirmed – that everything the sciences talk about is physical. If that is so, then the properties that appear in scientific laws must be ones that it is possible for physical things to have, and there must be an intelligible story to tell about how physical things can have them. Geologists would have no right to assume that there are mountains but that they can provide, or anyhow foresee providing, or anyhow foresee no principled reason why someone couldn't provide, naturalistic sufficient conditions for something physical to *be* a mountain. (Fodor, 1994, p. 5)

What kind of relation is the mental required to have to the physical for it to be true that the mental be physical in this sense? Metaphysically, there are two choices that are viable: reduction and supervenience. Fodor (1994, p. 4) frames the relation in terms of reduction, and reduction is often framed in terms of strict identities. There is reason to believe that he does not mean reduction in this strict sense, since this risk implying an elimination of the mental from the theoretical vocabulary, something that is far from Fodor's project. Boghossian (Boghossian, 1991, p. 65, 83) frames it in terms of supervenience. He defines a weak and a strong supervenience like so:

> A set of properties A *weakly* supervene on a set B, if no two objects in a given world could differ in their A properties without differing in their B properties. On the other hand, a set of properties A *strongly* supervene on set B, if no two objects drawn from any two worlds could differ in their A properties without differing in their B properties. (Boghossian, 1991, p. 83)

It is safe to say that for a theory to be naturalistic it is required either to reduce to or supervene on the physical.

It is easy to be confused when reading both Fodor and his commentators because they sometimes use these terms somewhat differently. For example, Fodor (1994, p. 4) says he assumes that his theory reduces to information, which as we will see, is a naturalistic notion.

Loewer and Rey (1991, p. 13), however, say that Fodor's physicalism is non-reductive. They say some clarifying things about Fodor's position in this passage from their book *Meaning in Mind*:

> Fodor's version of physicalism is, however, considerably weaker than many traditional versions. In particular, it is non-reductive: there is no requirement that there be bi-conditional bridge laws linking the phenomena of some special science to the underlying phenomena of physics. Fodor views "special" sciences in general as searching for causal explanatory laws at the level appropriate to their subject matter, developing relatively autonomously from the deeper theories whose regularities they may cross-classify. In the case at hand psychology may classify events as belonging to the same psychological type that differ in their neurophysiological properties, and neurophysiology might classify events belonging to the same neurological type that differ in their psychological properties. (Loewer and Rey, 1991, p. xiii)

The relationship between psychology and neurophysiology in this case is an example of what typically is meant by the multiple realizability of the mental, an intuition which is very important in the type/token -identity debate, a debate we will not go into here. We see that although Fodor is willing to commit to the view that every macro level property and other features of the world are fundamentally physical, he is not willing to commit to the type of physicalism that implies strict identities between mental and physical kinds. Reduction in the strong sense, i.e. that everything (every special science theory) ultimately will be expressible in some future complete physics, is one thing. The view that everything is ultimately physical is another.

Fodor's commitment to the existence of the properties and laws used in special sciences seems to be motivated by considerations about explanation, but also by considerations about causality. Fodor (Fodor, 1990, chapter 5) worries extensively that all properties other than those in the lexicon of physics are epiphenomenal. Part of his argument against the conclusion that intentional properties are epiphenomenal is that if they are, then so are all the special science properties also. Fodor argues for realism about all such properties on account of their causal responsibility in the laws they are subsumed by. It is a complex argument we will not review in detail here. It is acceptable, I think, to think that the world is ultimately composed of physical objects, and at the same time be skeptical about physics' power to explain, say, economic phenomena. It is possible to have an ontology that is richer than that of basic physics without giving up physicalism, and Fodor includes many higher level properties and laws in his. In (Fodor, 1990, p. 93) he says this:

> *Ontologically* speaking, I'm inclined to believe that it's bedrock that the world contains properties and their nomic relations; i.e., that truths about nomic relations among properties are deeper that – and hence are not to be analyzed in terms of – counterfactual truths about individuals. In any event, *epistemologically* speaking, I'm quite certain that it's possible to know that there is a nomic relation among properties but not have much idea which counterfactuals are true in virtue of the fact that the relation holds. It is therefore, *methodologically* speaking, probably a bad idea to require of philosophical analyses that are articulated in terms of nomic relations among properties that they be, as one says in the trade, "cashed" by analyses that are articulated in terms of counterfactual relations among individuals. (Fodor, 1990, p. 93-94)

I include the whole of this paragraph because it sums up Fodor's approach to several key ideas in philosophy nicely. In this thesis all of these ideas will be considered, but not all very comprehensively. However, they are important to mention because they are ideas that are constantly in the background of Fodor's thinking. So, though we will not consider these ideas much explicitly, I think it is a good idea to have seen what Fodor takes his own key ontological, epistemological and methodological assumptions to be.

So, in sum we have seen that Fodor is not a reductionist in the strict sense, and, he explicitly endorses a realist view of intentional states like beliefs and desires and he is a realist about the theoretical properties of many special sciences, like "mountain" in geology (Fodor, 1990, p. 139). The commitments Fodor has that we have reviewed are, I think, not all obviously compatible. Interesting as this is, I propose to leave this for now and turn to the question of what is required of a theory of representational content such as Fodor's to be naturalistic.

1.3. The Naturalism Condition

Fodor says this in TOC: "[W]hat we want at a minimum is something of the form '*R represents S' is true iff C* where the vocabulary in which condition *C* is couched contains neither intentional nor semantic expressions." (Fodor, 1990, p. 32). There are several things worth noting in this formulation of the minimal requirement of a naturalistically acceptable theory. As we will see, it is the representation relation which does most of the work in the actual theory. That is Fodor's account of how a mental representation, say a concept, relates to what it is about or represents. For Fodor, the assumption is that only symbols in the Language of Thought can represent, and that all other representation is derivative of this type of representation. On the informational approach, it is the relation between the individual symbols and the properties in the external world that are sufficient for causing them that constitutes the representation relation, and thus it is the representation relation that constitutes the relation between the mind and the world. The condition C is required to be a sufficient condition for the representation relation without itself being couched in semantical or intentional terms. A specific taxonomy of which terms are intentional/semantical is not available, but as we will see, causal terms and terms that are included in stating laws and properties that enter into laws are allowed. Intuitively, what one cannot do is to appeal to terms that presuppose that you have already accounted for meaning in a theory that purports to account for meaning, as Fodor's theory of content does. The naturalism condition is in many ways a demand on a theory not to be circular.

Satisfying the reduction/supervenience requirement can be done by satisfying the requirement that the theory should be stated in non-intentional/non-semantical terms. Providing a sufficient condition for intentionality that is stated without presupposing that what the theory seeks to explain is tantamount to providing a reduction/supervenience base for intentionality; which is what accounts for all the features of intentionality. Since there are no unexplained features of intentionality that the reduction/supervenience base does not account for, the conditions for supervenience is fulfilled. This condition is met if the supervenience base is framed in non-intentional/non-semantic terms, and the supervenience base is indeed sufficient for what supervenes on it.

1.4. Intuitions about Meaning

Fodor's theory of content is a theory that aims to account for meaning. So, one can ask: How does one normally go about accounting for the meanings of terms? What are our intuitions about answers to questions of the type "what is the meaning of x"? When asked to give an explanation of the meaning of, say, the word "cat", one usually tells a story about how cats are small, cute animals that have a number of legs and a tail and ears that are sort of triangular and.. etc. I think it is fairly intuitive that these types of explanations explain in virtue of exploiting the meaning relations between meaning bearing entities such as words or concepts. The mind is often assumed to be holistic in the sense that a concept means what it does in virtue of its place in a network of other concepts, or as a constituent in beliefs and the beliefs are individuated by their places in such networks. Fodor seems to think that this is the usual view. He says that "… on both sides of the English Channel, semantic holism is perhaps the characteristic philosophical doctrine of our time" (Fodor, 1994, p. 6). There are many variants of semantic holism, but all center around the idea that a given mental object gets its intentional/semantic status in virtue of its place in a network of other mental objects. Fodor opposes this tradition and thinks that theories about content should be atomistic. This is because he thinks that holism implies that one cannot generalize over intentional objects and thus not have intentional laws (Fodor, 1994, p. 7). We will see in a later chapter that there are some difficulties with intentional atomism.

Let us assume that the word “cat” expresses the concept CAT. The similarities between specifying the meaning of CAT and individuating CAT are striking and in much of the debate this seems to taken to be the same. To individuate something usually implies saying what it is that makes something unique, often by specifying something that is true of only one individual. Something is individuated if the characteristics used to identify it yields one result, i.e. the individual one wish to individuate. It is important to note that types, and not only individuals, also can be individuated by this criterion, though types and kinds often resist individuation by definition. In the common-sense example with CAT, we individuate the concept by giving a sort of description or definition that we use to single out the concept from all the other concepts. As we saw above, we can do so by specifying CAT’s relation to other concepts. There is a question whether theories that individuate contents holistically can be naturalistic in the sense we require, namely without employing intentional/semantical terms. The worry is that one needs to specify the contents of some beliefs to establish the relations that determine the contents of the other beliefs. There is also the worry that such a specification of content will, if it is to be naturalistic, imply an analytic/synthetic distinction because one arguably needs stable, necessary relations between some beliefs in order to specify the rest. These relations will then constitute relations that are necessary in virtue of the meanings of the contents, and that is tantamount to saying that some relations between contents are true in virtue of meaning, i.e. being analytically true. Philosophers who sympathize with Quine will naturally resist such a conclusion.

We have so far reviewed some criteria for what Fodor calls condition C. We have seen that most importantly it must be stated without employing intentional/semantical terms. This is because, to be naturalistic, it must supervene on something non-intentional/non-semantic that is sufficient for it. Fodor does not think he is obliged to provide a necessary condition for representation, only a sufficient one (Fodor, 1990, p. 96). The natural question to ask is what Fodor’s sufficient condition for representation is. Let us now turn to this question.

1.5. Resemblance and Causation

We remember that Fodor calls the sufficient condition for representation for C. What kind of non-semantic/non-intentional framework will satisfy C? Fodor considers two: resemblance and causation (1990, p. 33). It is important to remember here that the relation to be constituted by C is the *representation* relation. It is the representation relation that will serve as the main constituent of the theory of meaning and intentionality that Fodor proposes. This is important because this relation in and of itself does not suffice for explaining higher order mental phenomena such as thought or consciousness. Fodor argues against the view that representation is to be accounted for in terms of resemblance, very convincingly I think. We will review his arguments because the reasons why resemblance is inadequate tell us a lot about what kind of relation representation is taken to be.

Fodor first considers the proposal that representation can be accounted for in terms of resemblance. The proposal is something like this: The idea of a horse is an idea of a horse *because* it in some way resembles a horse. Generally we can say that the idea of X manages to be about Y (or mean Y) in virtue of its relation to Y. This is representation when the representation relation is framed as a resemblance relation, where resemblance presumably amounts to having features in common. Fodor considers three problems with this proposal (1990, p. 33-34), all of which seem to be fatal: (i) It is not clear what it means to say that an idea resembles what it is about. Resemblance seems to be about sharing properties, or having properties that are in the same categories. The fact that pictures resemble the objects they are of seems to suggest that not many properties need to be common at all for something to resemble something else. After all, a property like weight seems to have no impact on how pictures resemblance. On the other hand, a property like geometric shape seems to be essential, at least in visual resemblance. The point is that it is hard to imagine what property an idea should share with what it is about such that the result is that they resemble. Weight seems to be out of the question. Geometric shape might be conceivable, at least for simple geometric shapes. It is conceivable that the idea of a triangle could be realized in a triangular manner in the brain, but what about the idea "the biggest prime number"?

(ii) Representation seems to be a non-symmetrical relation. A sign represents a property without the property representing the sign. "Tiger" seems to be about tigers, but tigers don't seem to be about "Tiger"s. Contrary to representation, resemblance is a symmetrical relation, i.e. if X resembles Y, then Y resembles X. Representation does not have this feature so resemblance cannot be representation. (iii) Representation can be singular, i.e. that a sign can represent an individual object. Resemblance cannot capture this feature of representation because individual objects resemble each other and if resemblance is sufficient for representation, a sign X that represents object Y in virtue of X resembling Y, will also represent Z if Z resembles Y. Again, this is a feature not shared with representation. The sign *this tiger* will represent a unique tiger. Since tigers resemble each other *this tiger* should represent the other tigers too, but it doesn't (Fodor, 1990, p. 33-34). We can conclude that resemblance does not seem to be sufficient for representation.

These arguments point to features of representation which must be shared by whatever relation is to constitute representation. Causal relations are the obvious choice. Fodor says:

> Causal relations are natural relations if *anything* is. You might wonder whether resemblance is part of the natural order (or, whether it's only, as it were, in the eye of the beholder). But to wonder that about causation is to wonder whether there *is* a natural order. (Fodor, 1990, p. 33)

Unlike resemblance which, as we saw, has problems being what we need for representation, causal relations seem to have the features needed to constitute representation. Also, in determining what conditions a theory of intentionality must meet to be considered naturalistic we saw that being specifiable in non-semantic/non-intentional terms is paramount. Framing condition C in terms of causal relations seems to satisfy this condition.

We have so far discussed how to establish naturalistic criteria for a theory of meaning and intentionality when we have understood naturalism as being something similar to physicalism. We have seen that it is the representation relation which is to constitute the relation between the Mind and the World. It is the relation that is the foundation of intentionality, and the relation that is to constitute the basis for a naturalistic theory of meaning. We have seen that the representation relation needs to satisfy one condition, namely to be specifiable in non-intentional/non-semantic terms. This is to be done by framing it in terms of causal relations. But how are causal relations supposed to be able to reconstruct meaning and intentionality? To make sense of this idea, Fodor turns to the informational theories of meaning. Information, as we will see in the next chapter, is a notion which is naturalistic in the required sense, and a notion that can be used to construct the representation relation. The theme for the next chapter is the informational theory, and how information can be thought to constitute representation.

2. Chapter II: *Information and Causation*

In the previous chapter we tried to specify what features a theory of content needs to have in order to qualify as a naturalistic theory. That the theory should be framed in terms that are not themselves semantic or intentional we found to be the fundamental requirement. We saw that there are reasons to believe that the requirement that the theory should make content supervenient on, or reducible to, non-semantic/non-intentional features of the world will be fulfilled by framing the theory in non-semantic/non-intentional vocabulary.

This chapter is about the information relation and how this can be exploited in a theory of meaning such as Fodor's. We will start by considering what we will take information to be. This we will do by considering Dretske's theory of information, and how Fodor understands it. We will address the issue of ceteris paribus laws, an issue that is important in understanding what types of generalizations we use to express information relations. We will also see what relation information has to meaning, and some of the problems that arise when one tries to construct the latter out of the former. Let us first review some challenges to the way of doing semantics that we considered in the previous chapter.

2.1. Holism

As we have seen, and as Fodor admits, there are other approaches to semantics more popular than Fodor's approach of informational semantics (1994, p. 6). The chief alternative is the view that is called semantic holism, which is implied by conceptual-role theories of meaning (Fodor, 1994, p. 6). Fodor defines semantic/intentional holism's characterizing feature like this:

> Nothing can exhibit any intentional properties unless it exhibits many intentional properties; the metaphysically necessary conditions for a thing's being in any intentional state include its being in many other intentional states. (Fodor, 1990, p. 51).

It is important to note that, according to Fodor (1990, p. 51), there is no reason to believe that something cannot be both holistic and physicalistic, so Fodor's reservations about holism are not motivated by naturalistic concerns. What Fodor claims is rather that having a holistic theory of intentionality would preclude having a scientific, intentional psychology. He also says that the philosophers who are semantic holists often end up being semantic eliminativists (1994, p. 6), and that is obviously not an alternative for Fodor who is a realist about the intentional.

Why is holism a threat to intentional realism? What about holism makes it impossible to hold, for Fodor, without giving up intentional realism? It seems that, because of Quine's argument against the principled distinction between analytic and synthetic truths (For Fodor, at least, this is what Quine argued (Fodor, 1990, p. 52)), any intentional kind, if it is individuated by reference to other intentional kinds, cannot be subsumed by intentional laws, with the consequence that there are no intentional laws. This is what Fodor says:

> One important way that psychological laws achieve generality is *by quantifying over all the organisms that are in a specified mental state* (all the organisms that believe that P, or intend that Q, or whatever). But holism implies that very many intentional states must be shared if any of them are. So the more holistic the mind is, the more similar the mental lives of two organisms (or two time slices of the same organism) have to be in order that the same psychological laws should subsume them both. At the limit of holism, two minds share any of their intentional states only if they share all of them. And since, of course, no two minds ever do share all of their intentional states, the more [holism] is true the more the putative generalizations of intentional psychology fail, de facto, to generalize. (Fodor, 1990, p. 51-52)

It seems that Fodor thinks that holism, unless you assume an analytic/synthetic distinction, is subject to a slippery-slope type argument that shows that for two individuals to share an intentional state, they are required to share all intentional states. This is, of course, unacceptable.

2.2. The Informational Theory

Of the theories that claim to be able to account for meaning in non-intentional/non-semantic terms, and are therefore naturalistic, informational theories are what are taken to be the best option. I will start by sketching Dretske's theory of information. Once I have done this I will consider what version of the informational theory Fodor uses in formulating his theory. The best known account of informational semantics is perhaps Fred Dretske's account in *Knowledge and the Flow of Information* (1981). Dretske wants to give an account that can serve as a bridge between the cognitive sciences and computer sciences on the one hand and philosophy on the other. The way to do this is to specify a notion that is not intentional/semantical and use that in stating a condition for content. This notion is, of course, Information. In the preface to the book Dretske says this about information:

> Once this distinction is clearly understood [between meaning and information], one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes. One is therefore given a framework for understanding how meaning can evolve, how genuine cognitive systems – those with the resources for interpreting signals, holding beliefs, and acquiring knowledge – can develop out of lower-order, purely physical, information processing mechanisms. The higher-level accomplishments associated with intelligent life can then be seen as manifestations of progressively more efficient ways of handling and coding information. Meaning, and the constellations of mental attitudes that exhibit it, are manufactured products. The raw material is information. (Dretske: 1981: vii)

I think this exemplifies nicely what the project of naturalizing the mind can consist of. Though there are several theoretical alternatives one can choose from, where Dretske's is but one, this passage captures the ambition of the project as a whole, I think. In characterizing information as "an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes," he, says, in effect, that information is the notion we need to naturalize the mind. So, we will henceforth take information to be the notion that will satisfy the naturalization requirement, or, in Fodor's terms: the notion in terms of which we will state condition C.

The notion of information that Dretske employs is a quantitative notion, it is something that is measured in bits (Dretske, 1981, p. 3). Systems that can be in informational states, and signals that convey information about what caused them, are purely physical systems and signals. In this sense, information is everywhere where there is causation and the effect carries information about its cause. In what follows I will mostly rely on Fodor's exposition of Dretske's position and what Fodor himself takes information to be in *a Theory of Content* (1990). Let us look at a classic example of an information relation, namely the thermometer: A thermometer is a device we use to measure ambient temperature in a variety of circumstances, e.g. when we want to find out if the roast is cooked properly, if it is hot enough for swimming or if one should put gloves on when going out for a walk, etc. The thermometer serves this purpose in that it manages to represent the surrounding temperature in a way we have found to be reliable. The relationship we exploit in making thermometers is the causal relationship between mercury and the surrounding temperature. Mercury, we have discovered, expands in volume in a regular manner when the surrounding temperature goes up, and lessens in volume when it goes down. In making a scale on a glass column that contains mercury we can keep track of what the temperature is. We can know this because the causal relationship between mercury and temperature is of a particular kind.

Fodor considers thermometers in (1990, p. 44). First, a thermometer is a device that represents one property of the ambient medium (mean energy distribution) with another property of the mercury (volume). Both the ambient medium and the mercury have other properties that engage causally with each other and it is not, from an informational perspective, given that it is these two properties that are so related, i.e., that the volume of the mercury represents the temperature. That a thermometer represents temperature is dependent on other facts of the situation like that the vacuum in the tube that contains the mercury is intact, and so on. Secondly, the thermometer acts in this way because it is a device *for* representing the temperature. Most thermometers do this because they were designed to do this by a designer who had an intention of making the thermometer do just that. However, though thermometers need a designer to be such as to represent the temperature, the causal relation the thermometer exploits needs no such designer.

The point is this: even though a thermometer needs a designer to enable the property V represent the property T, what makes this representation possible is the underlying informational relationship between mercury and the ambient gas. Dretske calls this digitalization. He says: "The most specific piece of information the signal carries (about *s*) is the only piece of information it carries (about *s*) in digital form. All other information (about *s*) is coded in analog form." (Dretske, 1981, p. 137). Digitalization is the ability some systems have of reducing information in a signal and representing the information as a relation between singular properties. The simplest example is a system that can represent things as being on or off. The light on the dashboards of cars that represent the oil-level is typically of this kind (Dretske, 1981, p. 136). Though the physical system of the engine is a complex one and the mechanism that is the route for the information about the oil level has many different states each representing the amount of oil, the endpoint is a lamp that is either on or off. The simple property of the lamp's being off represents the complex property of the engine as having sufficient oil. The simple property of the lamp's being on represents the corresponding property of the engine's not having enough oil.

Dretske introduces several of what he calls levels of intentionality (Dretske, 1981, p. 172-173).The semantic level is level three. There are, in other words, two levels of information that are not semantical levels. It seems that Dretske introduces these levels as an attempt to say what characterizes the systems that are merely informational systems, and those that are truly cognitive, i.e., capable of entertaining beliefs. The difference between the two seems to be that while informational systems cannot distinguish properties that are nomically or analytically, as he says, connected, cognitive systems can (Dretske, 1981, p. 171-175). That is to say that a signal that carries the information that s is F when s is F logically implies that s is G, also carries the information that s is G. The fact that s is G is “nested” in s’ being F. Cognitive systems have the ability to distinguish the Gs from the Fs. An example may be that someone might represent s as being triangular without representing s as trilateral, though these properties are necessarily co-extensional. This is a feature of mentality which is hard for informational theories to account for. The ability to represent things as more fine-grained than things are in the world is one of the features for which a theory of this kind needs to account. This is something we will consider below when we consider the Frege cases.

2.3. Fodor’s Account of Information and Symbols

Fodor does not explain exactly what he takes informational theories to be. He assumes that a theory of information that is naturalistic in the required sense can be given (Fodor, 1994, p. 4), so he doesn’t worry much about the details, with the exception of how Dretske deals with the disjunction problem, which is the theme of the next chapter. But first, let us see how Fodor uses the information relation in his own theory. Fodor takes this to be the basic idea of Dretske’s:

> *“S-events (e.g., tokenings of symbols) express the property P if the generalization ‘Ps cause Ss’ is counterfactual supporting”*. (Fodor, 1990, p. 57)

This formulation is different from the way the informational theory is formulated in several important respects. First, instead of the relation “S carries information about P”, Fodor says “S expresses the property P”. To say of a symbol that it expresses a property might be read as introducing a semantic term. This is not what is happening. “Expressing a property” is here taken as a technical term replacing the “information” term used by Dretske. “Expressing a property” is here read as wholly constituted by the causal generalization linking the property and the sign.

This formulation is in essence an answer to the question of how information, something that is not in and of itself digitalized, can link particular properties with the particular representations that express them. As we saw in the thermometer case this is difficult to do without assuming a designer. Assuming a designer is, of course, not a viable option in a naturalistic theory. This definition is an attempt to give such an answer. The condition is formulated as a conditional. That means that 'the generalization "Ps cause Ss" is counterfactual supporting' is a sufficient condition for some S-event to express the property P. And, to say that a causal generalization is counterfactual supporting is in essence to say that it expresses a law. Laws are the only things that are counterfactual supporting in the sense we are after here. We can say that if there is a law that Ps cause Ss, we have a nomological relation that is sufficient for the symbol S to express the property P. Let us turn to the notion of a symbol and try to understand what is meant by that.

The term "symbol", as it is normally used, covers a wide range of applications, from how it is used in logic, to generally how words and sentences are said to be symbols. In daily life we also encounter other types of symbols. In fact, symbols are abundant in modern societies, most we hardly notice consciously since their occurrence is so natural to us. Typical examples are traffic signs, the use of the color red as a warning, various drawings that depict what situations we might get ourselves into if we are not careful, e.g., an avalanche or the like. We can also use the term when we say that Gandhi, for example, is a symbol of peace or tolerance. In order not to be question-begging the definition of the "expressing a property"-relation cannot imply that only a certain type of Ss can be symbols. This would be the case if it turns out that only mental symbols are candidates for being Ss, for example. As it stands the definition is wholly general and includes everything that can be subsumed by causal generalizations. As we will see later in chapter 4, to naturalistically specify what constitutes symbol-hood is not something Fodor does, and this poses some challenges for Fodor in avoiding that his theory implies pan-semanticism.

Something that is interesting to us is that what S expresses is a property. This might not be surprising, but it is not obvious that naturalistic accounts have properties so readily available. Causation is often something that is thought of as a relation among particulars. This was, as we saw, one of the virtues of the causal theory of representation as opposed to the resemblance theory of representation. But causation is, on this picture, a relation particulars have to each other in virtue of what properties they instantiate. This is why we can have laws that quantify over particulars, i.e., this is why particulars are subsumed by laws. This is so because particular properties are satisfiable by indefinitely many particulars, and it is this non-local or abstract feature of properties that enables generalizations to generalize in the first place. The ontological status of properties engenders exciting questions, but since we have seen that Fodor assumes that laws and properties are more fundamental than the individuals that instantiate them, we will not consider the matter in any detail here. Since the constituting relation of content is stated as a law, let us now consider what we take laws to be, in particular ceteris paribus laws.

2.4. Ceteris Paribus Laws

Special science laws typically involve macro level properties. Science taken as a whole is a pretty heterogeneous affair. The properties and laws that figure in the special sciences can cover the same cases but give different predictions and different explanations for phenomena and in this sense be theoretically incompatible. For example, the phenomenon of global urbanization might be explained very differently, and perhaps even be given conflicting explanations by economics and social science, respectively. Special science laws are not considered universal in the same sense that the laws of physics are taken to be. Usually, this is expressed by saying that special science laws are *ceteris paribus* laws. What the best way of understanding this qualification is is very much debated, and there is, as far as I know, no consensus. It seems that the ceteris paribus condition expresses that in some cases the antecedent of the law can be satisfied and the consequent not be true, though this does not mean that the law is false. Special science laws are in an important way domain specific. Unlike physical laws, which apply whenever, and where ever, it is assumed, special science laws do not. There can be circumstances that the domain of the law simply does not capture, even though the antecedent of the law is satisfied.

One example, one we mentioned above, and one we will come back to later, (Fodor, 1990, p. 155) can be the geological law that describes how the water erodes the riverbanks of a river. If suddenly, and certainly by magical means, a layer of diamond were to be placed on the whole of the banks and bed of the river, the erosion would cease though all the elements of the law's antecedent is satisfied.

Ceteris paribus laws are metaphysically unappealing since they represent a softening of the term 'sufficient'. The antecedent of a true conditional is considered to be sufficient for its consequent, though in ceteris paribus laws this is not the case. Perhaps this can be accounted for by an appeal to the internal consistency of the special science, and that when the ceteris paribus condition is violated the violation should come from without of the domain of the special science in question. This is a complex way of saying that ceteris paribus laws can be cancelled by factors outside its domain. As we will see in the quote below, they have conditions that need to be satisfied for them to apply without exception. The problem with such conditions is that it is very hard to know exactly what the conditions are, and when they are satisfied. Fodor says this about ceteris paribus laws:

> On the one hand, it's intrinsic to a law being hedged [being a ceteris paribus law] that it is nomologically possible for its ceteris paribus conditions not to be satisfied. And, on the other hand, a standard way to account for the failure of a ceteris paribus condition is to point to the breakdown of an intervening mechanism. Thus, meandering rivers erode their outside banks ceteris paribus. But not when the speed of the river is artificially controlled (no Bernoulli effect); and not when the river is chemically pure (no suspended particles); and not when somebody has built a wall on the outside bank (not enough abrasion to overcome adhesion). In such cases, the ceteris paribus fails to be satisfied *because* an intervening mechanism fails to operate. By contrast, this strategy is unavailable in the case of *non*basic laws; basic laws don't rely on mechanism of implementation, so if they have exceptions that must [be] because they're nondeterministic. (Fodor, 1990, p. 155)

Why is it important for Fodor to account for ceteris paribus laws? The reason is this: Psychological laws are special science laws and therefore ceteris paribus laws. If ceteris paribus laws cannot be accounted for properly there is no reason for supposing that they are real laws, only generalizations awaiting reduction to physical laws. For Fodor, who is a realist about properties figuring in special science laws, this is not a good result.

We have considered the notion of information and seen some examples of what kind of relation it is a notion of, and that it satisfies the condition for being naturalistic. We have also seen that Fodor frames the information relation in terms of a symbol "expressing a property". The information relation is thus taken to be a relation between a symbol and a property. This relation takes the form of a law and is thus counterfactual supporting. So, to sum up: the naturalistic relation that is to constitute the representation-relation is the relation between property and symbol in a causal law. Let us see how Fodor thinks this relation can reconstruct meaning and consider some problems with this approach: in particular what Fodor calls the Frege cases. Fodor's solution to the Frege cases is something we will not consider in detail because he solves the problem another way, one that is dependent on, the one we are considering here. But the Frege cases are generally a set of cases a theory of meaning must account for, so it is good to have reviewed them, I think.

2.5. Information and Meaning

What is the relation between information and meaning? We have seen that Fodor constructs the relation "expressing a property" from the information relation. As we shall see, the "expressing a property" -relation is actually insufficient in determining the content of a mental symbol. The reason for this is, as we mentioned above in the example of the triangular and the trilateral, the fine-grainedness of the mental. What this means is that contents and individual mental states can be individuated more finely than the objects that are in the extensions of the predicates. This is what Fodor refers to as the Frege cases (1994, p. 22). The Frege cases are a series of difficulties that have to do with how predicates can be co-extensive but nevertheless be non-synonymous. In accounting for meaning in naturalistic terms, one needs a naturalistic criterion that manages to break the connection between these types of predicates. The problem is inherent in what theoretical vocabulary one has available. Let's look at an example, the pair of predicates "triangular" and "trilateral". They are necessarily co-extensive, we can assume.

The problem for the informational approach is that we want to equate meaning with extension, by means of causation. We can say that what is responsible for a predicate's meaning is the causal relationship between tokenings of the symbol and the property the predicate expresses. So, a particular dog is responsible for tokenings of the mental symbol "dog" in virtue of being something that instantiates the property dog-hood.

2.6. Frege Cases

There are, in particular, two main problems for naturalized semantics: The Frege cases, and the Twin Earth cases. The Twin Earth cases we will get back to in a later chapter. Both cases are familiar from the philosophy of language. A standard Frege case (Fodor, 1994, p. 22) is the case where someone might believe that the Morningstar is remote but fail to believe that the Evening star is. But, "Morningstar" and "Evening Star" are names for the same object, so both names co-refer. If the meaning of a term is wholly determined by its extension then Morningstar and Evening Star should be synonymous, with the result that the beliefs are identical in meaning. The fact that this is not the case is what needs to be accounted for. The informational story is such that it aims to construct the meaning of a predicate by reference to its extension, where the extension is identified by what is sufficient for causing tokens of the predicate. The problem with both the Frege cases and the Twin Earth cases is that this strategy does not produce what we intuitively think of as the correct meanings of the terms. Thus it seems that individuating extensions is insufficient for individuating meaning. This is evident from the fact that “Morningstar” and “Evening star” mean different things even though they have the same extension. We will examine this problem more closely in the case with the frog and the fly later. But let us first briefly consider an example which is intuitively more problematic than the example with the Evening star and the Morningstar.

There are expressions which, though they differ in meaning, are necessarily co-referring, or co-extensional. “Triangular” and “Trilateral” is such a pair. This means that every possible object one can predicate the one to, one can also predicate the other to. The problem is to individuate the meanings naturalistically other than by individuating the extensions, since we have seen that this strategy fails. One approach can be to appeal to the mental analogue of the compositionality of natural languages. That is, in effect, to try to make an account about how neither “trilateral” or “triangular”, though they have the same extensions, are primitive expressions, i.e., they can be divided into their component parts, and when this is done it is seen that the component parts do not have the same extensions, and hopefully that will explain how their meaning differs. Fodor seems to try some such approach (see below). We will not go into this strategy in detail but I think some variant of this approach can intuitively seem promising, at least for these kinds of terms. The proposal conforms to the intuition that the reason the predicates differ in meaning is that though they refer to the same geometric object, they refer to different parts of that object, and it is that that accounts for the difference in meaning. When the “tri” component is removed it is immediately seen that “lateral” and “angular” are not co-extensive. This move is permitted only if one introduces something like a principled distinction between basic and non-basic predicates. There is perhaps a case to be made for that someone who has the concepts TRILATERAL and TRIANGULAR in their basic, primitive, non-composite versions, if this is even possible, they will necessarily have the same content on account of their extensions. Fodor says this about how he tries to solve the problem posed by the Frege cases:

> Propositional attitudes are relations between creatures, propositions and modes of presentation. None of the three is dispensable if a propositional attitude is to be specified uniquely. That’s because modes of presentation are sentences (of Mentalese), and sentences are individuated not just by their propositional content but also by their syntax. The identity of their content does not make wanting to marry M the same desire as wanting to marry J, any more than their synonymy makes “John is a bachelor” the same *sentence* as “John is an unmarried man.” (Fodor 1994, p. 47-48)

The example Fodor is referring to in this paragraph is the example of Oedipus who by accident married his mother (M). This happened because he did not know that Jocasta (J), whom he did want to marry, was identical to his mother, whom he didn’t want to marry. We will not consider in detail this way of solving the Frege cases, i.e. by appealing to the syntax of mentalese sentences. The reason for this is that to account for Fodor’s solution to this problem will take us to far from his proposal for accounting for the disjunction problem. The problem we will focus on is the Twin Earth problem as a case of the disjunction problem.

We have seen that for Fodor there are two main types of problems, the Twin cases (which we will discuss later) and the Frege cases. Fodor articulates them as the two ways broad contents and computational implementations might come unstuck (Fodor, 1994, p. 22). Broad content is a type of externally individuated content, a type of content we will consider in general in the discussion of the Twin Earth cases in chapter 5. The commonalities between both these types of problems are that they both arise out of the close link between content and extension. Both Frege cases and Twin Earth cases loosen the connection between content and extension, for both exemplify how content and extension come apart. The way we have been telling the informational story, it is the equation of content with extension, and extension with whatever causes the sign that is the relation that constitutes the “S expresses P” relation. As we will see in the next chapter when we consider what Fodor calls “the disjunction problem” it is the equation of the extension of a predicate (sign) with what is sufficient for causing it that is the root of the problem for the informational theory.

3. Chapter III: *The Disjunction Problem*

As we have seen there are challenges for informational semantics. Fodor's way of expressing the idea by substituting the informational relation with the "express the property" relation is an interesting suggestion since this way of formulating it makes explicit the connection between symbols and properties. We have seen that the problems arise from the restrictions placed upon the theory from its commitment to naturalism, and its attempt to construct content from the causal relation between the sign and what is sufficient for causing the sign. The terms "meaning" and "content" will sometimes in what follows be used to express the same idea. The disjunction problem is a problem about how informational semantics can account for error. Since informational theories such as the one we are considering are theories of representation the problem of error translates into being the problem of allowing for misrepresentation. This problem must be solved if Fodor is to account for one of the main intuitions we have about meaning, namely the intuition that meaning is robust. This, as we have mentioned, is the intuition that a term, or symbol, means what it does regardless of what caused its occurrence. In this chapter we will try to formulate what exactly the disjunction problem is and review some proposed solutions: Dretske's proposal and the historical/teleological theory's proposal. We will then see Fodor's arguments for why both these proposals fail to solve the disjunction problem. This will prepare us for Fodor's own proposal which we will consider in the next chapter.

3.1. Misrepresentation

The disjunction problem is a problem inherent in causal/informational theories. Fodor expresses it in the following way:

> [C]ausal theories have trouble distinguishing the conditions for *representation* from the conditions for *truth*. This trouble is intrinsic; the conditions that causal theories impose on representation are such that, when they're satisfied, *mis*representation cannot, by that very fact, occur. (Fodor, 1990, p. 34).

The problem arises from the relationship between predicate, property and the conditions for when the predicate expresses the property. Fodor expresses this idea by using the pair of terms "representation" and "truth". The representation relation, as we have seen, is constructed in terms of the predicate expressing the property. Fodor's idea of representation is that a mental representation is a tokening of a syntactic sign in a Language of Thought (LOT) (1990, p. 16). A syntactic sign in the LOT is what Fodor thinks of as what implements the S in "S expresses the property P", and the "expresses the property" relation is what constitutes the representational relation. We will often in what follows call S a predicate, and the relation as a predicate expressing a property.

Intuitively put, the disjunction problem arises from the feature of informational theories that says that the extension of a predicate is determined by what is sufficient for causing the token of the predicate (sign). The predicate expresses (means) the property that is the sufficient cause of the predicate. In sum, we can say that the predicate means whatever is sufficient for causing it. Let us look at an example. We are assuming that having the concept, DOG, say, involves having a symbol in a LOT that means *dog*, i.e. expresses the property dog-hood. The question is: What determines the content of the mental symbol? The informational theory's answer is: the causal relation the sign has with whatever is sufficient for causing it. We are supposing that in the mind we are imagining there is a tokening of the concept DOG. What determines the content of this concept? It is the concepts causal relationship with dogs. This is to say that DOG expresses the property dog-hood because dog-hood is a sufficient cause for DOG. This is the informational explanation of how something has meaning. Now, this gives rise to an obvious problem as we will see in what follows.

It can be retorted to the informational theorist, "there are surely other properties that are sufficient for causing DOG tokens. What about cases where we make mistakes?" Misrepresentation happens when something other than what is in the extension of DOG causes a DOG token. This happens when someone, for example, sees a wolf and mistakes it for a dog due to the distance to the wolf. Or, when someone sees a sheep from far away and mistakes it for a dog. Mistakes like these are common. The problem is that the representation relation is supposed to be sufficient condition for meaning. Why DOG means what it does is explained by the nomological relation between tokens of DOG and whatever causes it. It seems obvious that equating the extension of DOG with whatever is sufficient for causing DOG is not going to explain why DOG means dog, and not something else. As the example shows, the property of being a wolf can be sufficient for causing DOG tokens. That means that the informational theory implicates that wolves and dogs are in the extension of DOG. So we can express the extension of DOG as "dog or wolf" because both dogs and wolves are sufficient for causing DOG. This is why the problem is called the disjunction problem, namely because the informational theory ascribes disjunctive extensions to predicates that do not have disjunctive extensions. The concept DOG does not mean "*dog or wolf*", it just means *dog*.

Extensions can intuitively be thought of as sets of things that satisfy a property. The extension of, for example, the predicate "blue" is the set of all the things that are blue, i.e. all the particulars that instantiate the property blue. A theory of representation is required to account for the meaning/content of the symbols that represent. The concept DOG is about dogs, and only about dogs, the thought "that is a dog" is true if and only if the referent of "that" is a dog, etc. The disjunction problem threatens to make the concept DOG be about more than dogs, namely anything that is sufficient for tokening the concept, say wolves. This is not the right result for a theory of representation since it does not conform to the intuition we have been calling the intuition of the robustness of meaning. If it turns out that the informational theory is unable to attribute the correct extensions to concepts and other things that represent, then it must of course be rejected.

This is clearly an unacceptable situation. For example, there are perhaps infinitely many properties that can be mistaken for a dog. This implies that DOG is concept with an open-ended disjunctive extension, something that is clearly wrong. So, how can error and misrepresentation be accounted for? Before we turn to Dretske's proposal and the historical/teleological theory's proposal, let us see what these two proposals have in common: a distinction between what have been called type I and type II situations.

3.2. Type I and Type II Situations

The common strategy, though the actual proposals are very different, is this: It is to try to show that the situations where the predicate in question acquires a disjunctive extension are situations where the conditions for representation are not met (Fodor, 1990, p. 60). This amounts to introducing a further condition for representation than the ones we have considered so far. The new condition needs, like the ones we already have established, to be naturalistic in the same sense. The strategy is, in essence, to show that when misrepresentation occurs, something has 'gone wrong'. In these situations, i.e., situations of misrepresentation, the conditions for representation are not met, and that explains why it is a *mis*representation, or an error. As we have said, we want the theory to account for what we have been calling, following Fodor, the robustness of meaning. Robustness is the feature of meaning that we have taken to be the feature that enables a predicate to mean what it does *and nothing else.*

The basic idea in having two types of situations in which representation can occur is that one can isolate the circumstances where the symbol only applies to what is in its extension. In the type I situation the sign both represents and means the property it expresses. Everything has gone right and the sign means what it should, i.e. it applies only to things in its extension. The type II situation is the type of situation where something goes wrong and the sign does not mean the property it expresses. This way of putting it is a consequence of what we have been calling the equation of "expressing a property" and "being sufficient for causing". The disjunction problem is in a sense the problem of giving an answer to the question "when a symbol is caused by many different causes, which of these causes do the symbol express?". By appealing to the type I/type II distinction one tries answer this question. For, in type I situations the symbol always expresses the property the symbol is caused by. In type II situations the sign is free to be caused by something other than the property it expresses. Dretske proposes that the type I situation should be understood as a learning situation. Let us see what this proposal amounts to.

3.3. The Learning Situation

Dretske is aware of the disjunction problem and has proposed a solution to it (Fodor, 1990, p. 61). As we will see, his solution does not come without problems. Dretske's solution involves a principled distinction between the meaning-bestowing, type I situation, and the regular situation with the possibility of misrepresentation, type II situation. The type I situation he identifies with what he calls the learning situation. The learning situation is the situation where the meaning of a term is learned, and in so doing the representation relation is fixed in a way that specifies the meaning of the term.

According to Fodor there are reasons for thinking that Dretske's proposal is not satisfactory (1990, p. 62-63). Suppose that a student has been taught what "dog" means by, we can assume, being exposed to dogs in a way that secures that the student makes the required connection between the dog and the symbol "dog". This entails that during the whole of the training period the student has correctly applied the word "dog" to dogs, and only to dogs. Does this allow us to infer that what the student means by saying "dog" is not something disjunctive? No. He might still mean something disjunctive by it because of the fact that the training period is finite, i.e. it is a period during which the student is only exposed to a finite number of things. Fodor expresses the problem like this:

> "[…] it's the actual *and counterfactual S*-tokenings in training situations that fix the identity of the property that *S* expresses. Since it goes without saying that there must always be indefinitely many properties whose instantiations are not encountered in any finite linguistic apprenticeship, there are always indefinitely many disjunctive properties that the trainee's use of "dog" could express, *consonant with all of his actual tokenings of "dog" being dog-occasioned*". (Fodor, 1990, p. 62)

The result is a dilemma which Fodor considers (1990, p. 62). The following is how I understand the dilemma. We have the actual S-tokenings under control in the learning situation, so S-tokens are by assumption only tokened of the right properties, i.e. properties that are in the extension of S. So we can safely assume that, "dog" -tokens are only applied to dogs by the student. The problem seems to be that we cannot be sure that the student actually means *dog* and only *dog* by his tokenings of "dog". Fodor's point is that the training situation, being finite, cannot in principle guarantee that the student never will apply the predicate of something else. Why is this so? Since the relationship between the symbol and the property expressed is a nomological one, it licenses the use of counterfactuals in the individuation of the property we take the symbol to be expressing. Counterfactuals and subjunctive conditionals are intuitively conditionals that say that if x were to be the case, then y would follow. We will not discuss the nature of counterfactuals other than observing that they are different from ordinary conditionals in that they can be true in relation to other worlds than the actual one. In this case we are interested in the truth of the conditional: 'If the student encounters a cat-on-a-dark-night in the learning situation it will cause a "dog" token.' This can be true of the student even if he never encounters a cat-on-a-dark-night. Since we are assuming that all of the student's actual "dog" tokenings are dog-occasioned, we are assuming we are in the learning situation. In employing counterfactuals we will see that Dretske's proposal entails a dilemma that has the consequence that his proposal does not solve the disjunction problem. Remember that a solution to the disjunction problem requires that one allows for a symbol to be false.

Let us suppose that it is true that if a cat-on-a-dark-night had been encountered during the learning period it would have caused a "dog" tokening. That means that "dog" tokens express the property *dog or cat-on-a-dark-night* and the "dog" -tokens when caused by a cat-on-a-dark-night both in and out of the training situation are true. This is not the result we want, because it leaves no room for applications of "dog" to be false, i.e. to misrepresent dogs.

What if we suppose that the aforementioned counterfactual is false, i.e. that the property *cat-on-a-dark-night* would not cause "dog" tokens in the training situation? Then the consequence is that nothing other than dogs cause "dog" -tokens. After all, if cats-on-dark-nights could not have caused "dog" -tokens *in* the learning situation, there is no reason to suppose that they could have *outside* of the learning situation. This can be generalized such that nothing but dogs can cause "dog" –tokens, and all "dog" tokens are true. As we remember from above, this will not solve the problem. The problem is that on both alternatives "dog" tokens come out true. Symbol tokens are always true if they are applied to something in their extensions. When this extension is defined as what is sufficient for causing the symbol the result is that the symbol is always true of what causes it. This implies that they are never false, and thus cannot misrepresent. One can try to appeal to the counterfactuals concerning what the teacher would and would not have corrected, but that appeals to the intentions of the teacher, and is inadmissible in a naturalistic context. So, we can conclude that appealing to counterfactuals won't help us in the learning situation case. This, I think, is a fatal objection to Dretske's proposal. Let us turn to our other proposal: the teleological/historical theory's proposal.

3.4. The Teleological/Historical Theory's Proposal

Fodor's argument against Dretske's proposal shows that appealing to learning situations is unlikely to solve the problem, but it doesn't show that the strategy of appealing to type I and type II situations cannot be made to work. What is needed is a situation that can establish the meaning of symbol, in such a way that allows for false tokens of the symbol, i.e. tokens that are caused by properties they don't express. Fodor expresses what we are after nicely when he says:

> (i) If it's a law that *P*s cause *S*-tokens in type one situations, then *S* means *P* (and if *P* is disjunctive, then so be it); (ii) not all situations in which *S* gets tokened qualify as type one, so that tokens of *S* that happen in *other* sorts of situations are ipso facto free to be false. (Fodor, 1990, p. 64)

We see from this formulation that the meaning-defining situation is the type I situation, where it is a causal law that does the work of attaching the symbol to the property. (It is worth noting that not all disjunctive concepts are bad. Some concepts have, after all, disjunctive extensions). The informational theory does not itself contain the condition for distinguishing type I and type II situations. This condition must be stated independently of the informational theory.

What kind of condition may plausibly be one that can serve as defining type I situations? As we observed above, the intuition we have concerning misrepresentation is that when something misrepresents, then something has gone wrong. Take, for example, our belief forming mechanisms. It is often supposed that when we have false beliefs there is an explanation as to how we have gotten this belief that includes a reference to something that went wrong in the belief forming process. The guiding intuition is the intuition that if the belief forming mechanism worked properly then the belief would be true. Right and wrong are, of course, normative notions, notions that are used to express how things should or shouldn't happen. Situations of type II can be seen as the situations that allow for things that, in a sense, shouldn't happen (where things go wrong). Can this normative aspect be exploited in defining a type I situation for representation?

There are several things about this intuition of something having gone wrong in misrepresentation that I think is puzzling. Fodor quotes a passage of Stalnaker's to illustrate his point:

> Where beliefs are false … we also expect some explanation for the deviation from the norm: either an abnormality in the environment, as in optical illusions or other kinds of misleading evidence, or an abnormality in the internal belief-forming mechanisms, as in wishful thinking or misremembering. (Stalnaker, quoted in Fodor, 1990, p. 64)

Is it true that people typically excuse their false beliefs by saying something has gone wrong in their belief forming mechanisms? It is plausible, I think, that some perceptual beliefs are excusable in this manner, but people really do have false beliefs about many things that, at least not obviously, are a result of something going wrong. What, for example, has gone wrong in the belief-forming mechanisms of people who believe there are polar bears in Antarctica, or penguins at the North Pole? There are neither polar bears in Antarctica nor penguins at the North Pole, but what has gone wrong with those who believe it? It is not obvious that our brains are devices for making true beliefs with necessity when the condition that they are functioning as they should is satisfied. As with optical illusions, there is presumably nothing wrong with the belief-forming mechanisms of a person who believes the Müller-Lyer illusion is in fact a picture of two lines of differing lengths. This results in a false belief with, in my opinion, nothing having gone wrong. I think this debate is very interesting but we will not pursue these questions in detail here. The historical/teleological theory assumes that the normative aspects of our biological natures can be exploited in constructing a substantial notion of a type I situation by defining a situation that is Normal.

3.5. Normal Conditions and Functions

On the teleological/historical approach type I situations are called Normal situations (Fodor, 1990, p. 64). "Normal" is a notion that is supposed to capture several things, first and foremost condition that in a Normal situation nothing can go wrong. When the normative notion of normality is intended it is usually signaled by a capital "N" (Fodor, 1990, p. 85). Fodor says:

> *Normal* … is a normative notion, and *true* is a normative notion, so maybe it's not surprising if the former notion reconstructs the latter. … Of course, if the intentional circle is to be broken by appeal to *Normal* situations for symbol tokenings, we had better have some naturalistic story to tell about what it is for a situation to be *Normal* in the relevant respect. What might such a story look like? Roughly, the suggestion is that *Normality* should somehow be cashed by appeal to (natural) teleology; e.g., to some more-or-less Darwinian/historical notion of biological mechanisms *doing what they were selected for*. (Fodor, 1990, p. 64)

Biological functions are functions that are typically individuated "Normally". The Normal function of the heart is to pump blood around the body regardless of how many hearts in the world actually perform this function. In Normal contexts we can have a situation where only one heart fulfills its function while all the rest do not. In a normal (with no capital "n") context this would not be true. Contexts that are normal in this way are often said to be statistical in the sense that it would not be normal to, say, go to the movies on a Saturday night if only one person does it. By contrast, it doesn't matter for a Normal function how many of the individuals actually are performing the function. Normal is a normative notion, not a statistical notion (Fodor, 1990, p. 85).

The concept of Normal conditions looks like they can provide us with the means we need to provide a substantial difference between type I and type II situations, something we saw Dretske's proposal could not provide. Normal conditions look to be definable in biological terms and biology is a natural science. This, we may assume, will satisfy the naturalistic requirements. If what goes wrong can be specified with respect to some biological function that either functions inappropriately or functions in an environment that is inappropriate, it would seem we have a substantial notion between right and wrong, and thereby between type

I and type II situations. This is, in short, how the teleological/historical theory aims to account for error and misrepresentation. Let's see how Fodor sketches how all of this happens:

> … an organism's mental-state tokens get caused by, for example, events that transpire in the organism's local environment. There are, of course, mechanisms – typically neuronal ones – that mediate these causal transactions. And these mechanisms have presumable got an evolutionary history. There are presumably the products of processes of selection, and it's not implausible that what they were selected *for* is precisely their role in mediating the tokening of mental states. So there are these cognitive mechanisms, and there are these cognitive states; and the function of the former is to produce instances of the latter upon environmentally appropriate occasions. (Fodor, 1990, p. 65)

We can say that we are talking about the World – Mind relation. In this case we are talking about the Environment – Cognitive State relation as a type of the World – Mind relation. This relation, we are assuming, is typically mediated by a mechanism. Intuitively this is our sensory equipment such as eyes and ears etc. and our various information processing systems in the brain. The end of this line is a cognitive state with content, i.e. a thought that is about the environmental state that caused it. The mediating mechanisms of this causal chain are what we will call the Cognitive Mechanisms, and it is with respect to these we are talking about functions. How does this help us with the type I/type II distinction?

We said that misrepresentation happens when things go wrong. If the situation is Normal then things have gone as they should, and the representation relation between the Cognitive State and the Environment is in place. In a Normal situation things cannot go wrong. Things are Normal when the cognitive mechanisms are functioning as they should and the environment is such as the Cognitive Mechanism was "meant", or "designed" to function in, i.e., the appropriate environment. There are several places in this schema where things might go wrong: (i) The Cognitive Mechanisms might malfunction and the Environment is right; (ii) The Cognitive Mechanisms are functioning correctly, but the Environment is wrong; (iii) Both the Cognitive Mechanisms are functioning incorrectly, and the Environment is wrong. If (i), (ii), or (iii) is true of a situation it is an abNormal situation, i.e. situation of type II, and the content of the Cognitive State is free to be false.

This way of setting things up seems to constitute a substantive difference between type I and type II situations while satisfying the conditions for naturalism. If this is true it seems this account can solve the disjunction problem. Before we consider a case to which we apply this proposal let us first consider an example of what we take the notion of biological function to be.

The function of the cognitive mechanism and the environment are clearly intimately connected. Generally we can say that any biological function is closely connected to its environment. To take a common example, consider the heart. Let us suppose that the heart's function is to produce some effect in the body, more specifically to create and sustain a certain type of pressure inside of the circulatory system, i.e. to pump the blood around the body. To individuate functions by the effects they produce is a common strategy. The classic example being a doorstop, where anything and everything capable of producing the effect of keeping a door open qualifies as being a doorstop. Biological functions are perhaps individuated by further criteria, but we are assuming that it is the same in principle. So, let's suppose that the function of the heart is to create and sustain a certain pressure in the circulatory system. This is the heart's Normal effect. The heart is dependent on several factors to perform this function. The two most important for us here is that the heart itself must perform what it in and of itself must do, i.e. to contract in some sort of sequence, and, the environment in which the heart finds itself must be such as to connect the heart's movements with the effect of creating and sustaining a certain type of pressure. That environment most typically is a body. The heart can stop functioning if the heart itself is not working right, or the heart can function but in a body that does not sustain an appropriate environment. The functioning of the heart is a good example of a biological function. We are supposing in what follows that the cognitive mechanisms are biological mechanisms in the same way as the heart. If this is right we can say that for type I situations, i.e. Normal situations, both what we have been calling the cognitive mechanism and the environment must work, and be of the proper type for each other. Let us now consider the example of the fly and the frog.

3.6. The Fly and the Frog

There is a well known thought experiment often used to illustrate and argue for different views in this debate, and that is the example of the fly and the frog (Fodor, 1990, p. 70-71). It is an example used both by the advocates of teleological solution to the disjunction problem, and the ones who think that teleology will not solve the problem. As we will see, the problem with the teleological solution, according to Fodor, is that it doesn't manage to account for what we earlier called the "fine-grainedness" of meaning, a problem which can be traced to the problem of giving naturalistic conditions that will reconstruct systems that individuate content as finely as belief – and other intentional contexts. It is in many ways a similar problem to the Frege cases we considered in chapter 2. But before we go into all of this let us see how Fodor formulates the account he wants to criticize. This is his reconstruction of what a historical/teleological answer to the question of meaning is:

> [I]f you say to an informational semantical [*sic*] "Please, how does meaning work?" you are likely to get a song and dance about what happens when frogs stick their tongues out at flies. "There is," so the song goes, "a state *S* of the frog's nervous system such that: (i) *S* is reliably caused by flies in Normal circumstances; (ii) *S* is the Normal cause of an ecologically appropriate, fly directed response; (iii) Evolution bestowed *S* on frogs because (i) and (ii) are true of it." *S*, one might say, Normally resonates to flies. And it is only because it Normally does so that Mother Nature has bestowed it on the frog. And it is only because Mother Nature has bestowed it on the frog only because it Normally resonates to flies that tokens of this state *mean* fly *even in those (abNormal) circumstances in which it is not flies but something else that to which the S-tokens are resonating.* (Fodor, 1990, p. 70)

When considering this paragraph it is important to bear in mind that on the historical/teleological view, the state S of the frog's nervous system is functionally individuated. Fodor has several arguments against this view, the most important of which is this: The functional individuation of the neurological state that (i) and (ii) is true of is supposed to be sufficient for the individuation of the semantic content of the state. This, in turn, means not only that having beliefs has a Normal function, but also having particular beliefs has a Normal function. Fodor considers an example of Millikan's where the proper (Normal) function of the desire to "win the local Democratic nomination for first selectman is to bring it about that one wins the local Democratic nomination for first selectman" (Fodor, 1990, p. 67). It is by appealing to the function of the intentional state that one individuates the content of the intentional state. Fodor argues against this view, very persuasively, I think (Fodor, 1990, p. 67). For example, Fodor says:

> Stevenson wanted to win just as much as Eisenhower did, and the circumstances were equally Normal for both. But Eisenhower won and Stevenson didn't. In Normal circumstances, not more than one of them could have, what with elections being zero-sum games. So how could it be that, in virtue of a law or other reliable mechanism, in Normal circumstances everybody wins whatever elections he wants to. When the situation is Normal, the lion wants to eat and the lamb wants not to be eaten. But. … (Fodor, 1990, p. 67-68).

Millikan, as we have seen, proposes that intentional states can be individuated by reference to their proper functions (Fodor, 1990, p. 67). Fodor argues against this. However, it is important to note that Fodor's main argument against the teleological solution to the disjunction problem does not depend on assuming that the functions of the intentional states determine their content. What, on the historical/teleological theory, determines content is the Normal function of the cognitive mechanism which realizes the intentional state. Now that we have observed this let us see what the historical/teleological proposal amounts to in the example with the frog and the fly.

There is a mechanism that mediates the relation between the state S in the frogs mind and the environment the frog is in. Let us suppose the environment is Normal. The state S is about flies, i.e. means *fly*, because it is reliably occasioned by flies. We suppose this because when the flies are in the right relation to the frog, the frog will try, and often succeed, to eat the fly. This, we know, is good for the frog, because it helps the frog to survive. And the reason the mechanism is as it is, we can suppose, is because the mechanism is chosen by evolution to perform just this mediating task between flies and S because the ingestion of flies is good for the frog.

We can determine the content of the state because we know what the function of the mechanism is. This we have determined because we know what, evolutionary speaking, is good for the frog. Eating flies is Normally good, so the fly eating mechanism is functioning Normally when it mediates the relation between flies, S and the subsequent eating of the flies. We can individuate the content of the state S because that is consequent upon individuating the function of the mediating mechanism. What makes this inference from the function of the mechanism to the content of the state possible is the historical/teleological theory’s assumption that the function of the mechanism determines the content of the intentional state. So, the problem of determining content becomes, on the historical/teleological account, the problem of determining function.

But does this show us that the function of the mechanism is to make the frog catch flies? Fodor argues that it does not (1990, p. 72). The problem is that what happens in this example is that one either assumes the content of the state, and infers the function of the mechanism; or one assumes what the mechanism is designed to do, and infers the content of the state it produces. If one assumes that the function of the mechanism is to get the frog to eat flies it reasonably follows that the intentional state is about flies. And we know from biology that it is the function of the mechanism to make the frog eat flies. So, what is the problem?

For Fodor (1990, p. 72), the problem is that these types of accounts do not take into account the fact that we are in fact describing phenomena. We forget that this is only one description we could give, and if we lose sight of the fact that we are on a descriptive level we may conclude that we have determined the content of S when we have not. This is illustrated by the fact that we can tell this story in other terms. Let us assume that in the frog's Normal environment all the flies are, say, little black dots. It is, I think, plausible that if you throw something that has the appearance of a little black dot in front of a frog, the frog will snap at it. If this is right, and the function of the mechanism is that of mediating little black dot sightings to the snap guiding mechanism, then we can conclude that the content of S can equally well be taken to be *little black dot*. This is not the desired result for the historical/teleological approach. Fodor says:

> Notice that, just as there is a teleological explanation of why frogs should have fly detectors – assuming that that is the right intentional description of what they have – so too there is a teleological explanation of why frogs should have little-ambient-black-thing detectors – assuming that *that* is the right intentional description of what they have. The explanation is that *in the environment in which the mechanism Normally operates* all (or, most, or anyhow enough) of the little ambient black dots are flies. So, in this environment, what ambient-black-dot detectors Normally detect (de re, as it were) is just what fly detectors Normally detect (de dicto, as it were); wiz., flies. (Fodor, 1990, p. 72)

A condition on the teleological theory is that the function of the mechanism should be the reason why evolution has bestowed it upon the frog. The selectional advantages that come with the function of the mechanism are the reason why the organism has the mechanism. So, what selectional advantage is the mechanism in question responsible for producing? We can assume that the answer is the ingestion of flies, because that is the prime source of food for the frog. The flies are part of the frog's Normal environment and the ingestion of food produces obvious selectional advantages.

This means that the condition on the theory requires that the function should produce the right effect: ingesting flies. Fodor says: "*Darwin cares how many flies you eat, but not what description you eat them under*" (1990, p. 73). The conclusion is that the historical/teleological theory does not manage to provide a univocal description of the function in question. This is critical for its ability to ascribe content to the intentional states of the frog. When the theory assumes that the content of the mental state is determined by the function of the mechanism that produces it, the content of the mental state is obviously sensitive to the function of the mediating mechanism. When the function of the mechanism is indeterminate the consequence is that the content of the mental state is also indeterminate. Mechanisms that detect black dots are equally effective in helping the frog survive as mechanisms that detect flies in environments where all the black dots are flies. We can conclude from this that the indeterminacy of functional ascriptions results in indeterminacy of content ascriptions and thus that the historical/teleological theory does not solve the disjunction problem.

I think this line of argument is persuasive and that it shows that the teleological theory cannot distinguish between the hypothesis that the content of the mental state S is *fly* or the hypothesis that it is *ambient-black-dot.* The reason the argument works is that the historical/teleological theory does not have available a common way of individuating functions, namely by using counterfactuals. The reason this is unavailable is that the historical/teleological theory wants to cash functions out in terms of selectional advantages. For a Darwinian something can only be a selectional advantage if it is an actual advantage. On the Darwinian picture, a mechanism cannot be selected for the advantages it would have generated *if* the environment *had* been different. The mechanisms of selection are such that only actual advantages result in the survival of the actual individuals who have it. Fodor has an example of a fish that lives deep in the ocean where there is no light (1990, p. 76). This fish has a certain color. Fodor claims, rightly I think, that the reason that fish has that color, the advantages the fish has in virtue of having that color cannot be attributed to the counterfactual situation that if the fish had lived in a part of the ocean with much more light, say, at the surface, then it would have had an advantage. Organisms cannot, evolutionary speaking, have properties that are accounted for by reference to the selectional advantages the properties would have imbued in circumstances other than what the organism actually lives in.

3.7. Counterfactuals and Functions

As we have seen, the option of appealing to counterfactuals in determining function is out of the question for the historical/teleological theory. But why are counterfactuals a good option? What is it about counterfactuals that can solve the disjunction problem? The answer is that counterfactuals can split co-extensional terms (perhaps not all, but presumably enough). Let’s see what Fodor says:

> Let’s ask *how much* intentional indeterminacy one would have to put up with on the teleological story. I think that the right answer is that appeals to mechanism of selection won’t decide between *reliably equivalent* content ascriptions; i.e., they won’t decide between any pair of equivalent content ascriptions where the equivalence is counterfactual supporting. To put this in the formal mode, the context: *was selected for representing things as F* is transparent to the substitution of predicates reliably coextensive with *F*. A fortiori, it is transparent to the substitution of predicates *necessarily* (including *nomologically* necessarily) coextensive with *F*. In consequence, evolutionary theory offers us no contexts that are as

> intentional as ‘believes that. …’ If this is right, then it’s a conclusive reason to doubt that appeals to evolutionary teleology can reconstruct the intentionality of mental states. (Fodor, 1990, p. 73)

Belief contexts, which are perhaps the paradigm of intentional contexts, are what one usually calls opaque contexts. This is a trademark feature of the mental, or of what we saw Dretske calling “genuine cognitive systems”, earlier. As a type of the Frege cases this is a recurring problem for the informational theory. As we have seen, the informational theory has severe problems accounting for how predicates like, say, “triangular” and “trilateral”, can come apart. We saw in the case with the fly and the frog that the informational theory can only distinguish between predicates that have different extensions. And it seems no extensional context can split properties like triangularity and trilaterality.

This is a challenge for all types of naturalistic semantics: to show that there can be contents that are as fine grained as those needed for making propositional attitude ascriptions. The conditions for meaning/content that the teleological/historical theory postulates are in this regard not sufficient. They cannot distinguish between contents that are reliably co-extensive because the way the teleological/historical theory individuates content is by appealing to the function of the state that produces the content. The problem with this, as we have seen, is that individuating functions by way of appealing to selectional mechanisms does not yield univocal functional ascriptions. If the function cannot be determined the content cannot be determined either. The reason for this is that selectional mechanism does not carve the world finer than extensions. In environments where all little black dots are flies and all flies are little black dots all the individuals that satisfy the one predicate will satisfy the other.

If one can appeal to counterfactuals, things change. One can imagine a world in which none of the little black dots are flies and flies are instead little bright dots, say. If we imagine taking some of our frogs from this world to the world just described, we can assume, we would find that the frogs snap at the little black dots but not the little bright dots, i.e. the flies. Then we can conclude that the content of the state of the frog is *little-black-dot*, and not *fly*. This move, though attractive, is not available for the teleological/historical theorist for the reasons we considered above.

3.8. Conclusion

This concludes the first part of this thesis. We started by considering naturalistic theories in general and what is required of them. The informational theory was found to be the most plausible one. This first part has mainly been concerned with the question of naturalization, and partly with the questions concerning meaning. We have seen that the informational theory can serve as a base for a theory of representation, but that it has severe problems. We have focused on the disjunction problem and several proposals for how to solve it. We have seen that both of them are inadequate.

Of the three main intuitions we mentioned on the outset this part has only been concerned with two of them: naturalization and robustness. In the next part, when we consider Fodor's own proposal for solving the disjunction problem, we will focus more on the issue of robustness and less on the issue of naturalization. We have seen that much of the naturalization problem is solved by having the informational theory as a base. We will also address the third intuition, the intuition that meaning is not everywhere, in the next part. I will argue that Fodor does not sufficiently establish that his theory conforms to this intuition.

Part II:

4. Chapter IV: *Asymmetric Dependence*

Let us start this part by summing up what we have concluded so far. We have seen that teleological/historical solutions to the disjunction problem fail. To solve the disjunction problem the theory is required provide unique content ascriptions to a mental state without employing semantic/intentional terms. The historical/teleological theory's proposal manages to not use semantic/intentional terms, but the mechanism by which the content is to be individuated is, as we saw in the previous chapter, dependent on univocal ascriptions of function to the cognitive mechanism responsible for producing the mental state.

The problem of individuating content thus becomes the problem of individuating function. We saw that the historical/teleological theory can only appeal to actual selectional advantages in doing this. A crucial consequence of this is that the historical/teleological approach excludes appeals to counterfactuals in determining function. It follows that the teleological/historical theory cannot distinguish between reliably co-extensional descriptions in specifying function. This is because functions are individuated by the effects they produce, and teleological/historical theories cannot distinguish between effects that are equally good at producing the right selectional results. Consequently, any description of a function that makes the organism that possesses it fit with the selectional constraints, specifically that the organism survives, is a reasonable description of the function.

We remember this from the example with the frog and the fly. Teleological/historical theories could not distinguish between descriptions of the relevant mechanism that had the effect that the frog ingested flies. We have assumed that the purpose of the mechanism is to enable the frog to catch flies. But as we saw the frog eats just as many flies when we describe the function of the mechanism as making the frog ingest little black dots. That is because little black dots are reliably flies in the world the frog is in. On the historical/teleological account function determines content, so where the function of a mechanism is indeterminate it follows that the content of the state is also indeterminate. In these cases we are unable to distinguish between content ascriptions where the contents are reliably co-extensive in the frog's environment.

What, then, is the content if the frog's mental state? We know that different things are true of flies and little black dots, and we do not treat them the same. The question is if they are different for the frog. Do the frogs treat them differently? There are reasons to suppose that they don't. Frogs will snap at things that look like flies (i.e. that look like little black dots), that are not flies. In the context of informational semantics where the information relation is that of a reliable co-variation between cause and effect, we say that the effect carries information about what causes it. The effect in question is the frog's snapping. The cause is whatever is sufficient for causing the frog to snap, and it seems that flies are only a subset of everything that elicits snaps from the frog. Since there are other things that elicit snaps, flies cannot be the cause of the snaps *qua* flies. They must be the causes of snaps qua something else, i.e. little black dots. Counterfactuals can, in this way, determine the content of the frog's mental state. But is this enough to solve the disjunction problem generally?

Informational semantics' proposal is that the content of a mental state expresses whatever property is responsible for causing it. "Horse" means *horse* because horses reliably cause "horse". This, as we saw earlier, generates disjunctions problems, specifically about error. The teleological/historical solution to this problem is to distinguish between two types of situations: Normal and abNormal situations where the situation that guaranties that what causes the predicate (the content of the state) is what is in the extension of the predicate. This secures that the predicate is caused only by what it applies to, or is about. The teleologically Normal situation for the frog is when the frog gets to ingest flies. We saw that the historical/teleological theory's resources for specifying the content of a mental state are spent by specifying the function of the mechanism that produces the state. I think that Fodor's argument shows that the strategy of the historical/teleological theory does not provide the right result. The right result would be specifying a unique content to the frog's mental state, and the historical/teleological theory fails to provide such unique ascription of content because it fails to uniquely specify the function of the mechanism which is responsible for producing the mental state. This we established in the previous chapter. The main theme for this chapter is to formulate Fodor's own proposed solution to the disjunction problem. The key intuition about meaning to be accounted for by his proposal is the one we have been calling the robustness of meaning. As we will see, accounting for this intuition and solving the disjunction problem are the “same undertaking” (Fodor, 1990, p.91).

4.1. Robustness and Extensions

As an alternative to appealing to different types of situations, Fodor introduces what he calls the asymmetric dependency condition (1990, p. 90). This is his proposal for accounting for the robustness of meaning. Robustness, as we have seen, is one of the pre-theoretic intuitions about meaning that any meaning theory must account for. Meaning is something that is intuitively inherently robust. Fodor expresses what he takes robustness to be like this:

> In actual fact, “cow” tokens get caused in *all sorts* of ways, and they all mean *cow* for all of that. Solving the disjunction problem and making clear how a symbol’s meaning could be so insensitive to variability in the causes of its tokenings are really two ways of describing the same undertaking. If there’s going to be a causal theory of content, there has to be some way of picking out *semantically relevant* causal relations from all other kinds of causal relations that the tokens of a symbol can enter into. And we’d better not do this by implicitly denying robustness – e.g., by idealizing to contexts of etiological homogeneity. (Fodor, 1990, p. 91)

The most intuitive, and obvious example of the robustness of meaning is a case we have not yet considered. It is, perhaps, clearest counter-example to informational semantics in the form we have been considering. The example has to do with how thoughts relate to, on the one hand, what causes them, and on the other, what they are about. The feature of thoughts we are after here is one that, prima facie, seems to be at odds with the basic assumptions of informational semantics. The informational theory claims that the reason a predicate means what it does is because it expresses the property which is causally responsible for its occurrence. When we think of examples concerning perception this intuitively seems reasonable because it is intuitive that when we see a horse and think “there’s a horse”, it is the horse that is causally responsible for our thinking that particular thought. But what about cases where there are no horses and one is merely thinking about old western movies, and suddenly one finds oneself thinking about horses? These horse-thoughts are not occasioned by horses at all. They are perhaps occasioned by cowboy-thoughts, but this is of course not a requirement for being a horse-thought.

The crucial point is that we constantly think thoughts that aren’t caused by what they are about. They are most often caused by other thoughts. Horse-thoughts can be occasioned by almost anything, but horse-thoughts mean *horse* regardless of what causes them. Fodor says: “*… the meaning of a symbol is one of the things that all of its tokens have in common,* *however they happen to be caused*" (1990, p. 90). And, this seems to be the case the other way as well, namely that horses are not always sufficient for someone to think "horse". This is the other side of the error story we have been telling. When you mistake something for something else, say a horse for a cow, and "horse" is a misrepresentation of cow, then cow is not sufficient for the tokening of "cow". So, robustness is the intuition that the meaning of a predicate is distinct from what causes its tokening. Robustness is an absolute demand on this kind of theory, and as we have described it here, if you have a meaning theory that doesn't make meaning robust it will not qualify as a meaning theory.

As we will see, Fodor's solution to the problem of robustness is twofold. First, we can observe that the problem is not a problem about the informational approach. Rather, it is a problem about having to rely on a distinction between type I, and type II –situations. The problem, Fodor thinks, is caused by appealing to special type of situation which is such that in that situation a symbol cannot be caused by anything that is not in the symbols extension. This is equivalent to saying that there is a situation where a sign is always true about what causes it. In such a situation a symbol, say "dog", if it is caused at all, is necessarily caused by dogs and nothing else. It is this feature of the informational approach, a strategy we saw that both Dretske and the historical/teleological approach tried to use in solving the disjunction problem that Fodor dispenses with. He does not dispense with the basic framework of the informational theory.

Second, he introduces his asymmetric dependency theory to account for robustness. How is this criterion to account for robustness? The asymmetric dependence condition should provide a criterion for distinguishing the tokenings of predicates that are caused by something in the predicate's extension and those that are not. It is important to distinguish between two ways we can talk about extensions on the naturalistic view. The first way is a term for what a predicate applies to. This is the normal sense of the term. The term "dog" has dogs and only dogs in its extension because dogs are what the term applies to. In this sense "the meaning of a term" and "the extension of a term" is roughly equivalent. The second way is a way of determining the meaning of a term in an informational theory by specifying a term's sufficient causes. The goal is, by employing only non-intentional/non-semantic terms, to reconstruct a symbol's extension by specifying the symbol's sufficient causes. This is, as I see it, the core of the disjunction problem. What we called the first view of extension is roughly equivalent to the meaning of the term. The term "dog" has dogs and only dogs in its extension because it means *dog*. One can solve the disjunction problem by naturalistically specifying extensions that capture this feature of meaning. The disjunction problem is that informational theories ascribe disjunctive extensions to terms that intuitively don't have disjunctive extensions; this, of course, is a problem because the terms don't mean something disjunctive. The solution to the disjunction problem requires the theory to ascribe correct extensions to terms like "dog".

4.2. Asymmetric Dependence

As we have seen, Fodor's solution to the disjunction problem relies on finding an alternative to relying on the distinction between the type I and type II –situation. His proposal is to appeal to dependences among the causal generalizations that govern symbol tokenings, in hope of determining which causal generalization is the one that is semantically relevant. But before we get to that, let us see what asymmetric dependence is. Intuitively, asymmetric dependence is a dependence relation where the dependence does not go both ways. Let us suppose A is asymmetrically dependent on B. This means that if you have A then you have B, but not necessarily the other way around, i.e. it means that you can have B without having A. I suppose having a bike and riding a bike exemplify such a relation. You can have a bike without riding it, but you cannot ride the bike without having it. There is obviously no bike riding to be done where there are no bikes. But there can be bikes where there is no bike riding. So intuitively, riding bikes is asymmetrically dependent on having bikes. One common way of talking about asymmetric dependence is to talk about it in terms of possible worlds. On this reading we shall say that A is asymmetrically dependent on B if the worlds in which B is the case and A isn't, are closer to us than the world where A is and B isn't. This is obviously a very general condition that very many pairs of things will satisfy. This definition is intended to exemplify how asymmetric dependencies are defined in terms of possible worlds. It is possible to operate with a more narrow scope, as Fodor does when he restricts the relevant dependencies to being dependencies among laws.

To conclude the example we can say that worlds where there are bikes but no bike riding are closer to us than worlds where there are bike riding but no bikes. This latter world is arguably an impossible one, but that is only to say that bikes are necessary for bike riding. There is some controversy about how to determine distance between possible worlds, something we will get back to when we discuss Paul Boghossian's objections in chapter 6. The idea is that all the false tokens of a symbol depend on there being a true token of a symbol, and that this relation can be exploited in solving the disjunction problem. We now have a basic grasp of what asymmetric dependence is. Now, let us see how Fodor thinks that this condition can be used to solve the disjunction problem without having to face the same problems that Dretske and the historical/teleological approach did.

This is what Fodor says, when he advances the idea of asymmetrical dependence as a preferable alternative to "idealizing to contexts of etiological homogeneity" (1990, p. 91) –i.e. postulating type I situations:

> Here's a first approximation to the proposal that I favor: Cows cause "cow" tokens, and (let's suppose) cats cause "cow" tokens. But "cow" means *cow* and not *cat* or *cow or cat* because *there being cat-caused "cow" tokens depends on there being cow-caused "cow" tokens, but not the other way around.* "Cow" means *cow* because, as I shall henceforth put it, noncow-caused "cow" tokens are *asymmetrically dependent upon* cow-caused "cow" tokens. "Cow" means *cow* because *but that "cow" tokens carry information about cows, they wouldn't carry information about anything.* (Fodor, 1990, p. 91)

We remember from above that the first part of Fodor's theory is to assume the framework of the informational theory. We also remember that this takes the form of reliable causal conditionals that are counterfactual supporting and can therefore be regarded as laws (ceteris paribus). We are assuming that when a cow causes a "cow" token it does so in virtue of having the property cow, which is a causal property that has the causal power to produce "cow" tokens. As we mentioned before, there are indefinitely many properties that has the power to cause "cow" tokens. Every one of those properties can be subsumed by a causal generalization, i.e. a law with the property, X, on the antecedent side, and symbol "cow" on the consequent side. This is a consequence of the holistic character of belief fixation (Boghossian, 1991, p. 78), something we will review later. Fodor's proposal is thus that the disjunction problem is no longer the problem of determining what *situation* is the meaning bestowing one, but rather which, of the indeterminately many laws is the meaning determining one. This constitutes a radical break from the approaches we have considered so far, i.e. Dretske's proposal and the historical/teleological theory's proposal.

Fodor's proposal is, in effect, to see which laws are dependent on each other. The main idea is that if he can find one predicate governing law that all the other laws asymmetrically depend on then he has found the law that determines the meaning of the predicate. This is achieved if he can determine the law that, if broken breaks all the rest, i.e. makes the predicate not caused at all. We saw this in the example with the bike. If we take away the bike, the consequence is that there is no bike riding to be done. If one removes the bike riding, one can still have a bike and we can conclude that riding bikes is asymmetrically dependent on having bikes. But the bike example does not exemplify what asymmetric dependence amounts to in the context of semantics. First of all, bike riding doesn't *mean* bike in any meaningful sense of the word and neither is bike riding any kind of causal consequence of bikes.

In the next chapter, chapter 5, we will see how Fodor, by employing the asymmetric dependency condition, aims to account for the robustness of meaning. The rest of this chapter is devoted to considering how Fodor aims to account for the other main intuition about meaning mentioned in the introduction of the thesis, namely the intuition that meaning is not everywhere, or ubiquitous as Fodor sometimes puts it (1990, p. 93). The way of accounting for this intuition is to show that the theory does not entail what is known as pan-semanticism, which just is the implication that meaning is everywhere. We will consider Fodor's account of pan-semanticism, and I will argue that there are reasons for thinking that it is not satisfactory. I will not try to show that Fodor's view implies pan-semanticism, only that he does not show that pan-semanticism does not follow from his views.

4.3. Pan-Semanticism

There are, as we have observed, two important intuitions which a theory of meaning needs to account for, both of which, if not accounted for, implies that the theory is not successful as a theory of meaning. The robustness of meaning is one of them. A theory of meaning which doesn't make meaning robust is not satisfactory as a meaning theory. The other important intuition is that not everything has meaning. Pre-theoretically we think that there are only a few things that have meaning. Words, sentences and thoughts are common examples. A theory of meaning that has the implication that everything has meaning is not only unsatisfactory as a theory of meaning, the implication amounts to a reductio of the theory. The failure to account for one or both of the main intuitions amounts to a reductio of the theory.

In this section I want to argue that Fodor does not convincingly argue that his theory does not imply pan-semanticism. The reason for this is that he does not account for his assumption that only symbols are candidates for having meaning. Pan-semanticism is a big worry for all information based accounts of meaning since information is, just as meaning is not, everywhere. Since information is everywhere there is causality and meaning is constructed from information the theory threatens to imply that meaning is everywhere as well. This would be a catastrophic result for a meaning theory. Fodor argues that his theory does not have this consequence (1990, p. 92). I argue that his argument does not succeed in establishing this conclusion.

Fodor observes (1990, p. 93) that pan-semanticism is sensitive to the fact that the information relation is transitive. If A → B and B → C then C carries information about A. Fodor uses the example with smoke and fire. If we take “smoke” to mean smoke and smoke means fire, then presumably “smoke” means fire. But, “smoke” doesn’t mean fire, and consequently the theory has yielded the wrong result and implies pan-semanticism. How is this problem to be solved? Fodor’s solution is to appeal to asymmetric dependence to decide between the laws ‘smoke → “smoke”’, and ‘fire → “smoke”’. If we do this we see that “smoke” does not mean fire because the law ‘fire → “smoke”’ is asymmetrically dependent on the ‘smoke → “smoke”’ law. This means that the worlds where smoke causes “smoke” tokens without fire causing “smoke” tokens are closer than the worlds where fire causes “smoke” tokens and smoke doesn’t.

Fodor frames this by specifying which information relations depend on which. In his version of this argument (1990, p. 93) there are not only two such information relations, but three. The ‘smoke → “smoke”’ relation, ‘the fire → “smoke”’ relation and the ‘fire → smoke’ relation. Let’s see what he says.

> “Smoke” tokens carry information about fire (when they’re caused by smoke that’s caused by fire). But they don’t *mean* fire because their dependence on fire is asymmetrically dependent on their dependence on smoke. Break the *fire → smoke* connection, and the *smoke → “smoke”* connection remains intact; our using “smoke” in situations where there’s fire doesn’t depend on smoke’s carrying information about fire. But break the *smoke→ “smoke”* connection and the *fire → “smoke”* connection goes too; our using “smoke” in situations where there’s fire does depend on “smoke”’s carrying information about smoke. (Fodor, 1990, p. 93)

This argument, as far as I can see, establishes that “smoke” does not mean fire. However, I will argue that this does not establish that Fodor’s position does not imply pan-semanticism. As this argument stands it is assumed that only *signs* of a particular kind can carry information in the way that can be the basis for meaning. I think that there are two ways of reading this quote, two ways that the argument can be interpreted. I’m inclined to read the ‘fire → smoke’ connection, i.e. where there are no inverted commas around smoke as in the rest of the quote, as a typographical error and that the inverted commas were intended to be there. This is what I will be calling the non-literal way of reading the quote. The literal way of reading the quote is to read it as it is written. The inverted commas are, of course, what signals that the word within is a symbol that stands for something. But the assumption that there are only symbols of this kind that can enter into meaning relations is an assumption that is in need of an argument. This is something Fodor does not provide. That certain kinds of things, namely symbols, are the only things that can stand in meaning-relations with other things is what the argument is supposed to show, not what it is supposed to assume. The worry that needs to be dispelled if Fodor is to establish that pan-semanticism does not follow from his theory is not the worry that "smoke" means fire. What he needs to show is that smoke doesn't mean *fire*.

The asymmetric dependence condition is supposed to determine the content of a symbol. It does so by distinguishing between all the different laws that describe the causing of the symbol, and, by determining which of the laws all the others depend upon but which itself does not depend upon any other. The point I want to make here is that it is assumed that all the laws we are talking about are laws that are symbol causing laws, i.e. that they have a predicate X that is framed by inverted commas as the consequence. This means that all that the asymmetric dependence condition applies to are symbol causing laws, or, equivalently, laws that govern the tokenings of symbols. But to establish that the theory doesn't imply pan-semanticism Fodor simply cannot assume this. Pan-semanticism is a terrible implication of a theory for many reasons, but chiefly this: that smoke, qua standing in a reliable co-variance with fire, comes out as *meaning* fire, is nonsensical, and unbelievable. An absurd consequence of pan-semanticism and the disjunction problem may perhaps also be that smoke does, in a sense, mean not only fire, but also smoke-machine in a way that makes smoke disjunctive. This way of saying it makes clear the absurdity of the result that smoke, in and of itself, has a meaning. What would it be for smoke to, for example, have an extension, be it disjunctive or otherwise?

Let us see what happens if we try to use Fodor's theory's resources and construct an argument that shows that pan-semanticism doesn't follow from his theory. We will consider how Fodor treats the cases of inter-level relations and causal chains in *a Theory of Content* (1990). Especially the causal chain case, I argue, has similarities with the pan-semanticism case that we can exploit in constructing the argument. It is important to note that this argument assumes that one cannot independently account for a substantive notion of symbol which dissolves the pan-semanticism worry. I assume that Fodor has not presented such an account.

This problem arises from the informational theory's reconstruction of "means that" from "carries information about" (Fodor, 1990, p. 92). As Fodor puts it: "Information is ubiquitous but not robust; meaning is robust but not ubiquitous." (1990, p. 93). Fodor's solution is a solution to the disjunction problem for the *symbol* "smoke" when it is indeterminate if the symbol means fire or smoke because of the transitive nature of the information relation. It is not necessarily a solution to why smoke, in and of itself (whatever that may mean), does not mean *fire*. After all, smoke stands in an information-bearing relationship with fire much in the same way as mercury stands in an information-bearing relation to the surrounding ambient temperature, as we saw in the example with the thermometer. To solve the problem in the way that Fodor does is, in effect, to say that only a certain kind of information bearing relations qualify as symbol relations, and for those the problem of pan-semanticism never arises. I argue that what is doing all the work in this argument is a substantive notion of symbol-hood which is not accounted for, only assumed.

The reason asymmetric dependence can be thought not to be able to solve this problem is the fact that the only available dependencies it can distinguish between are the ones with the same consequent. That is, they can only distinguish between fire caused "smoke" tokens and smoke caused "smoke" tokens. In both causes the relevant laws have "smoke" as the consequent. These laws cannot, in the relevant sense, be dependent on laws that have different consequents. After all, the problem the asymmetric dependence relation is intended to solve seems to be how one and the same sign seem to have more than one meaning, not how several different signs have several different meanings. Therefore it doesn't seem that the relevant dependencies can include the fire causing smoke law. So what motivates Fodor's appeal to the asymmetric dependence relation in this case, when it is clear that the pan-semanticism implication cannot be dealt with by the asymmetric dependence relation, understood as a relation among laws with the same consequents? Let's see what Fodor says about some similar cases, namely the cases about inter level relations and causal chains.

4.4. Inter-Level Relations and Causal Chains

The result Fodor needs in these cases is to show that none of them exhibit robustness. Since they are, as we will see, cases of asymmetric dependence it is crucial that they are of the wrong type to produce robustness. Cases where there are mechanisms that implement macro laws are typical cases of this kind. This is a typical feature of special science laws and part of the reason why special science laws have to be qualified by having ceteris paribus clauses. The point here is that the macro law is asymmetrically dependent on the micro (implementing) mechanism. This example is from Fodor (1990, p. 117). Let's assume that the macro law in question is A → D, and the micro level mechanism is described by the law B → C. A → D is depends asymmetrically on B → C iff you can break the A → D connection without breaking the B → C connection, but not the other way around. This is the case if B → C is necessary but not sufficient for A → D.

We are assuming that both these conditionals express causal laws. Causal laws express informational relations, so the informational part of the story is accounted for. The relation between the laws is one of asymmetric dependence, something that should result in establishing not only information, but meaning, i.e. establishing that C *means* B, because that is the law that the other laws are dependent upon. This, of course, is not a tolerable result for Fodor. He solves this problem by saying this:

> The point of appeals to asymmetric dependence in theories of content is to show how tokens of the same type could have heterogeneous causes compatible with their all meaning the same thing; i.e., it's to show how robustness is possible. Correspondingly, if a sufficient condition for content is going to be fashioned in terms of asymmetric dependence, it must advert to the dependence of one causal law *about "X" tokens* upon another causal law *about "X" tokens*. But the sort of asymmetric dependencies that interlevel cases generate don't meet this condition. What we have in these cases is a law that governs the tokening of one thing (Ds in the example) that's dependent on a law that governs the tokening of some other thing (Cs in the example). This sort of asymmetric dependence doesn't produce robustness, so it's not semantically relevant. (Fodor, 1990, p. 117)

As we mentioned above there is a restriction on the theory that only dependencies that have the same type of symbol as consequents are potentially robust, and hence constitutive for meaning. He says: "… if a sufficient condition for content is going to be fashioned in terms of asymmetric dependence, it must advert to the dependence of one causal law *about "X" tokens* upon another causal law *about "X" tokens*." (Fodor, 1990, p. 117). This seems to have the consequence that if we consider the problem of pan-semanticism we see that the dependencies that could have solved the problem are disqualified by assumption. These considerations, I claim, seem to imply that appealing to asymmetric dependencies cannot account for the pan-semanticism case. This disqualifies Fodor's solution if we read the above quote in a way that makes the consequents the same, i.e. the non-literal way. When we do so, we see that Fodor does not solve the right problem, i.e. the pan-semanticism problem, but a problem about robustness. If we read him literally in the quote, so that he appeals to asymmetric dependencies between laws that do not have the same consequents, he breaks his own criterion for robustness, as quoted above. Let us see if Fodor can still account for pan-semanticism by using what he says about inter level relations and causal chains.

If we read the Fodor quote describing how he deals with the pan-semanticism worry literally, we see that the case is very similar to the causal chain case quoted below. Fodor's solution to the pan-semanticism case looks, on the literal reading to be an instance of a causal chain where the different parts of the chain depend asymmetrically on each other. The literal way of reading Fodor's solution to the pan-semanticism problem is reading the fire → smoke connection as being asymmetrically dependent on the smoke → "smoke" connection. Let us for the moment ignore that this disqualifies it from potentially being robust. This, due to the transitivity of the information relation, can be reconstructed as a causal chain such as A → B → C where A is fire, B is smoke and C is "smoke". We remember from the pan-semanticism case that the answer Fodor needs for his theory not to imply pan-semanticism is that C means B, and not A because the A → C connection is asymmetrically dependent on the B → C connection. So, it is, I think, surprising to see what he says about the causal chain case:

> Suppose that *A*s (qua *A*s) cause *B*s (qua *B*s), and *B*s (qua *B*s) cause *C*s (qua *C*s), and assume that *A*s are sufficient but not necessary for the *B*s. Then the law *A* → *C* is asymmetrically dependent on the law *B* → *C*. Why doesn't it follow that *C*s mean *B*? *Answer*: Because, although the causal chain makes the *A* → *C* connection asymmetrically depend the *B* → *C* connection, the dependence of *C*s on *B*s that it engenders is not ipso facto robust, and content requires not just causal dependence but robustness too. The dependence of *C*s on *B*s *is robust only if there are non-B*-caused *C*s. But the causal chain *A* → *B* → *C*, engenders an asymmetric dependence in which *all the A-caused Cs are also B-caused*. So the asymmetric dependence of *A* → *C* on *B* → *C* doesn't satisfy the conditions on robustness; so it is not semantically relevant. (Fodor, 1990, p. 118)

I claim that on the literal reading of Fodor's treatment of the pan-semanticism case what he says constitutes a case similar to the causal chain case. Fodor's argument in the pan-semanticism case purports to show that "smoke" means smoke, and that smoke does not mean fire. This is the right result for dispelling pan-semanticism. However, in the causal chain case, Fodor's account, which I argue is analogous to the literal reading of the pan-semanticism case, does not produce the right result, namely that C ("smoke") means B (smoke). He denies that C means B. If what I claim is right and the arguments are analogous then Fodor seems to provide contradicting results.

Are the cases only superficially alike, or is there an actual structural likeness? Several things are alike. Both cases exhibit, at least on the literal reading, a causal chain where the whole is dependent on the part in a way that makes the dependence asymmetrical. As we have seen this apparently is not sufficient for meaning. There is another requirement, namely that the dependence should produce robustness. This provides us with another clue as to how to understand asymmetric dependence. Consider the sentence: "The dependence of Cs on Bs *is robust only if there are non-B*-caused Cs." (Fodor, 1990, p. 118). This seems to say that for C to be robust, i.e. meaningful, none of the causes that are sufficient for causing C can be necessary. In the causal chain case we see that B is necessary for C, by assumption.

If we insert the pan-semanticism case into the causal chain case above we get that fire (qua fire (A)) causes smoke (qua smoke (B)), and smoke (qua smoke (B)) causes "smoke" (qua "smoke" (C)). And, we can assume that fire (A) is sufficient but not necessary for smoke (B). Why, then, do the two cases come out differently? It seems that to make the latter case come out right one needs to assume that *B*s are necessary for *C*s. That is the only way to guarantee that *C*s are not robust. Fodor does not assume this in the pan-semanticism case where "smoke" is free to be caused by other things than smoke, and one gets the asymmetric dependencies between the right kinds of laws, i.e. laws that aren't apart of chains, and "smoke" means smoke. When we set up the case this way we see that by assuming that *B*s are necessary for *C*s we can get the right result: that C does not mean B. Though we will not pursue the matter here I think there is a possibility inherent in talking about ceteris paribus laws that one can challenge Fodor's assumption that in inter level cases the *B*s will always be necessary and thereby try to show that inter level cases can produce robustness.

4.5. Conclusion

We have seen that the condition for robustness is that when appealing to asymmetric dependences among laws those laws must have the same consequents. We saw that there are two ways of reading Fodor's treatment of the pan-semanticism case, one literal and one non-literal. On the non-literal reading where we read the paragraph as satisfying the condition for robustness we see that Fodor does not account for the pan-semanticism case in a satisfactory manner. Instead he assumes a substantive notion of symbol-hood which does all the work of dispelling the pan-semanticism worry. This substantive notion of symbol-hood is not accounted for naturalistically, and is therefore, I argue, not available for Fodor in dispelling the pan-semanticism worry.

On the literal reading of Fodor's treatment we saw that it is structurally similar to the way Fodor accounts for inter level relations and causal chains. By exploiting the similarity between the pan-semanticism case and the causal chain case I argued that Fodor risks contradicting himself. Fodor can avoid the contradiction by assuming that "smoke" has no necessary sufficient cause, something which is plausible. I conclude that the literal line of argument can account for the pan-semanticism worry, but at a cost I do not think Fodor could accept. His options are, as I see it, to provide an independent account of a substantive notion of symbol-hood that explains how it is possible to read the pan-semanticism case in what we have been calling the non-literal way. The prospects for this, I think, are challenging. The other option (the literal argument) is to loosen the condition for robustness, i.e. to allow semantically relevant asymmetric dependencies between laws that do not have the same consequents. This is to allow what we have been calling the literal reading of the paragraph. If this is purchased at the cost of accounting for robustness, which it appears to be, it is surely not an option for Fodor.. This concludes this chapter. In the next chapter we will continue to investigate if Fodor's theory can account for robustness by seeing whether it accounts for the Twin Earth cases.

5. Chapter V: *Twin Earth*

In the last chapter we saw Fodor's proposal for dealing with the pan-semanticism worry and I argued that it is not satisfactory. Now we will look at some of the other challenges to Fodor's proposed solution to the disjunction problem, particularly the ones that have to do with the modal aspects of the theory. Since Fodor does not think highly of speaking in terms of possible worlds we will not take him as fully committed to such views (Fodor, 1990, p. 95). But, since it is the way he formulates the theory, and it is not obvious what other way it can be formulated, there is some commitment to analyzing elements of the theory in terms of possible worlds. We will also see that Fodor is committed to a form of verificationism. This is important in that it introduces an epistemological aspect and thus a break with the purely metaphysical considerations up until now. We will also see that the verificationism is closely connected with the way Fodor accounts for the contents of kind terms. This has important implications for another view Fodor holds, namely intentional atomism. Intentional atomism is the view that it is possible for systems to have a single intentional state. I will argue that though this might be possible, Fodor's account excludes the possibility of this state having a content that is the same as the content of a kind term. What this amounts to we will see below.

The main concern in this chapter is that the informational approaches to semantics, being a species of what is called content externalism, need to be able to account for what normally are referred to as Twin Earth cases, or problems. There are numerous variants of the Twin Earth cases, but all involve reference to worlds that are similar to ours but different in some crucial aspect. The standard (Fodor, 1994, p. 22-26) example owes to Putnam and is usually taken to be an argument for content externalism, i.e. the view that content does not supervene only on aspects internal to the organism that is in the states that have content. Content externalism claims that content does not supervene merely on the internal states of the organism, rather, content supervenes on the internal states of the organism plus states external to the organism. The view opposing content externalism is content internalism. This view is that content does in fact supervene only on internal facts about the individual having the state with content. Some forms of the latter view imply holism about content, but this is a consequence only of one holds that there is no analytic/synthetic distinction, something Fodor does. We will not discuss this debate in what follows. Let us see what the standard Twin Earth case is.

5.1. Standard Case

Putnam’s classic argument for content externalism can be formulated like this (Fodor, 1990, p. 114-115): Assume a world that is physically exactly like our world, with one exception. Instead of the chemical substance H_2O, which we have in abundance in our world, there is the equally abundant substance XYZ. XYZ, though having a radically different chemical makeup than H_2O, shares all of the macro properties of H_2O, i.e. it tastes the same, you can use it to cook, fish live in it, etc. In other words, the world we are assuming is a perfect copy of our own except that all the H_2O in our world is replaced with XYZ. Now, assume we make a perfect clone of some individual from our world and put him in the Twin world. The clone is a perfect copy and is internally identical to the original. We are, of course, ignoring the problems raised by the fact that the human body consists largely of H_2O.

The two individuals are, by assumption, identical, including their internal mental states. The question is whether this fact implies that the content of their mental states are also identical. Is the fact that their internal states are identical enough to secure that the contents of their mental states are identical? When they use the word “water” to talk about the substance before them that is wet, that they use to cook, that fish live in etc., they are not talking about the same stuff. What they are referring to are by assumption different substances. So, do “water” -tokens, either in their brains or when they talk, mean the same thing or have the same content? Fodor, and all other externalists, think no. “Water” means two different things in the different worlds because they refer to different substances even though the people entertaining them are internally identical. This seems to imply that content cannot supervene only on the internal facts about the individual, but must also take into account external facts about the environment. This is a result Fodor’s theory must account for.

We can think of the example above as the basic formula for the Twin cases. All the variants include some sort of indistinguishable external kind who partly determine content, and an individual who is by assumption ruled out as the source of the differing content. This point can also be expressed by appealing to how we individuate content. A thing is individuated when conditions are given that uniquely specifies that thing. We can say that when something is individuated it is done by specifying features of the X such that the X is a uniquely determined individual. The Twin cases seem to show that to individuate mental states is not sufficient for the individuation of their content.

The Twin Earth cases, I think, strongly suggest that content supervenes not only on the internal states of the individual, but also on external states. If this is the case, it seems that the internalist cannot account for content. But the Twin Earth argument shows that to individuate the content you need more theoretical resources than to individuate the mental state, namely you need to appeal to the reference of the mental state, an external fact, to individuate the content.

It is the individuation story that is our main concern here. To give a naturalistic story about content is to give naturalistic specifiable conditions for the individuation of the content of a mental state. I take the Twin Earth case to imply that internalist theories, such as we have defined them, cannot account for content. But it is not obvious that Fodor can either, and, as we shall see, his answer has some interesting epistemological implications. But what is the problem the Twin cases pose that Fodor must account for?

The worry that needs to be dispelled by a theory of meaning is that the Twin cases might imply that "water" means XYZ when the intuition is that XYZ is not in the extension of "water". That "water" has XYZ in its extension is intuitively the wrong result, because the intuition is that "water" only has H_2O in its extension. These intuitions are plausibly connected with the fact that we are likely to think about the Twin cases in terms of dispositions and counterfactuals. In light this, the Twin cases can seem to imply about ordinary English speakers that if they were magically (or otherwise) to be transported to the Twin world they would be disposed to call XYZ for water, i.e. apply "water" to XYZ. After all, XYZ does not distinguish itself in any way from H_2O, so there would be no reason to not apply "water" to XYZ. Cases where the world contains both substances will be addressed below. It is this disposition to apply "water" to XYZ that creates the problem for Fodor in that this would seem to imply that the best way to describe the extension of "water" is by the disjunction "H_2O or XYZ", which we have seen in earlier cases, imply that "water" means something disjunctive, which it does not. So, Twin Earth cases seem to present another type of disjunction problem that needs to be accounted for. As we saw earlier, a theory of meaning like Fodor's must reconstruct the right extensions of symbols to be satisfactory.

In his solution to this problem Fodor (1990, p. 115) appeals to the fact that "water" is a kind term and that part of what it is to be using terms as kind terms is to treat them in accordance with certain intentions. He thinks, reasonably, that treating things as natural kinds involves having intentions to treat all the objects in the environment that one takes to be relevantly similar, as the same kind. Things that are taken as being not of a kind, e.g., that they are relevantly dissimilar, are not covered by the same kind term. He also notes (Fodor, 1990, p. 115) that not all expressions are controlled by such intentions, but they are not natural kind terms. How does appealing intentions help to secure the result Fodor needs, namely that "water" only has H_2O in its extension?

5.2. Verificationism

> My point is that the intention to use "water" only of stuff of the same kind as the local samples has the effect of making its applications to XYZ asymmetrically dependent on its applications to H_2O ceteris paribus. Given that people are disposed to treat "water" as a kind term (and, of course, given that the local samples are all in fact H_2O) it follows that – all else equal – they would apply it to XYZ only when they would apply it to H_2O; specifically, they would apply it to XYZ only when they *mistake* XYZ for H_2O; only when (and only because) they can't tell XYZ and H_2O apart. Whereas, given a world in which they *can* tell XYZ and H_2O apart (and in which their intentions with respect to "water" are the same as they are in *this* world), they will continue to apply "water" to H_2O and refrain from applying it to XYZ. (Fodor, 1990, p. 115)

We see in this paragraph that Fodor introduces two aspects of his theory which we have not so far considered in detail. The first is that he appeals to intentions in accounting for kind terms. This is something we will consider below and I will argue that this has implications for Fodor's views about intentional atomism. The other is that he introduces the term "mistake" in its epistemological sense. Thus, the focus shifts from the conditions of truth of a symbol to whether or not the subject would recognize something as true. We can assume that to apply a word to a thing according to some intention one has is naturally thought of as performing an action. Actions, at least intentional actions, have conditions for when they are successful and not. Presumably, the relevant condition in this case is that the action should correspond to the intention, i.e. the application of the word should only be applied to the objects that the intention dictates. The mistake in this case is the misapplication of the word to objects that are not approved by the intention, so to speak.

What makes appealing to intentions epistemological is that the objects that are candidates for being attributed with a certain kind-hood must be recognized as such, and such recognition is dependent upon the possibility of deciding the truth of certain relevant conditionals, especially the conditionals concerning the attributes of the object. If what one takes to be of a kind does not behave like the rest of its kind in some situation of other, then this is evidence for that one has to do with more than one kind. This implies that mistakes are dependent on the possibility of being right. Where there is no possibility of deciding the truth of certain conditionals, e.g. finding a world where some conditional is true of H_2O but false of XYZ, there is no possibility of being right and hence there is no sense to the idea that one is wrong. In this kind of world "water" probably has a disjunctive extension. This is a point Fodor accepts (Fodor, 1990, p. 119, 91). He says:

> … the theory I'm selling says that false tokens can happen whenever they like; only if *they* happen, so too must tokenings of other kinds: No noncow-caused "cow"s without cow-caused "cow"s; false tokens are metaphysically dependent on true ones. (Fodor, 1990, p. 91)

This will also be evident from Fodor's considerations about verificationism below. As we will see he thinks a certain amount of verificationism is unavoidable on any causal account of intentionality.

5.3. Fodor's Proposal

Let us see how Fodor proposes to account for the problem posed by the Twin Earth case, and how he responds to some challenges to his view. Lynne Rudder Baker has presented a challenge to Fodor which he responds to which we will consider (Fodor, 1990, p. 103). I depend on Fodor's exposition also on this case. This is a case where two kinds that are not distinguished share a world. This case will have some important implications as to how we are to understand certain features of Fodor's theory. But first let us consider how Fodor accounts for the intuition that content depends on external facts.

The case is this: On the Twin world the substance XYZ can cause "water" tokens. The informational theory's treatment of the case has the result that "water" means XYZ because XYZ is in "water"'s extension. This is because, as we remember, the extension of "water", in the informational theory, is specified by reference to whatever is sufficient for causing "water". This result, as we saw earlier, is a bad result for a meaning theory because it does not conform to our intuitions that "water" means water and has H_2O, and only H_2O its extension. What we want is some way of distinguishing between the H_2O caused "water" tokens and the XYZ caused "water" tokens. Fodor claims that there is a difference (1990, p. 115) and the difference is that in worlds where H_2O and XYZ are indistinguishable, if you break the one connection then you break both, *and* vice versa; but in the worlds where H_2O and XYZ are distinguishable the following asymmetric dependence hold: if you break the H_2O – "water" connection, you also break the XYZ – "water" connection, but *not* vice versa. That is, the H_2O – "water" connection holds where all other X – "water" connections are broken, and this is what accounts for the fact that "water" tokens only have H_2O in their extensions. This is a promising result. If this is true, then Fodor seems to have found a difference between XYZ caused and H_2O caused "water" tokens. This is what is required for solving the problem. But this way of determining a difference does not make perspicuous what the difference consists in. In particular, the role of the intentions and kind terms are not explained on this account. Let us try to specify in what way these notions contribute to Fodor's proposed solution.

The standard Twin Earth story imagines what would happen in a world where all of the H_2O is replaced by XYZ. And, for the people who belong in that world, "water" means XYZ and not H_2O. But for us, were we to go to this world, "water" would mean H_2O because it is true of us that if we performed some tests on XYZ and discovered that what we had taken to be H_2O in fact was XYZ, we would stop using "water" in thinking and speaking of it. This is explained by the fact that we have intentions of using the term "water" only of things that has the chemical make-up H_2O.

As we mentioned above there are interesting cases of the Twin Earth argument where one takes the same world to contain both of the relevant substances. Fodor looks at several of these cases. We will consider what he says about Baker's robot-cat case (Fodor, 1990, p. 103). This is a world where there are both regular, ordinary cats, and artificial cats, i.e. robot-cats which look and act just as ordinary cats. We also have a person S whose mental symbol "cat" has only ever been caused by robot-cats. Then, one day, S experiences a "cat" tokening that is caused by a real cat. The question is what the meaning of that particular "cat" token is? According to Fodor (1990, p. 103), Baker thinks that there are three alternatives, none of which she thinks is tolerable.

The alternatives are: (i) the "cat" token means *cat* and is true of the cat, (ii) the "cat" token means *robot-cat* and is false of the cat, and (iii) the "cat" token means something disjunctive, namely *cat or robot-cat*, and is true of both. Baker thinks that the first one can't be right because the dependence seems to be going the other way than we want, i.e. the disposition to token "cat" tokens when presented with a real cat seems to be asymmetrically dependent on the disposition to token it when presented with robot-cats. After all, that is what has caused "cat" tokens so far, and presumably can cause "cat" tokens even if S never encounters any real cats. That the "cat" token means *robot-cat* is supported by the conditional: S is disposed to apply "cat" to robot-cats even if she never encounters any cats. But, Fodor and Baker claim, this too ignores relevant counterfactuals, particularly that cats would have caused "cat" tokens had S encountered any. Plausibly, the counterfactual governing S' "cat' tokenings is that both cats and robot-cats (that is the disjunction 'cats or robot-cats') can cause them, and that S is disposed to token "cat" of both. That S' "cat" tokens have all been caused by robot-cats is purely accidental. This leads us to the final alternative where "cat" means *cat or robot-cat*, and is true of both. And this seems to be the disjunction problem all over again, i.e. ascribing a disjunctive extension to something that intuitively does not have a disjunctive extension.

5.4. Kind Terms

By exhausting the options and showing how none of them can be accepted this argument purports to be a reductio of Fodor's proposal. This is true if Fodor cannot accept any of the alternatives. But Fodor does in fact accept one of the alternatives, namely the third option: that "cat" means *cat or robot*. But how can he accept this? What is it about this case that makes it special, as it must be, since this is the only case we have seen so far that Fodor admits that a normal kind concept has a disjunctive extension? Hopefully, the explanation will not only explain how Fodor can hold this seemingly wrong conclusion without giving up his account, but also why Fodor can perform the surprising move in the original Twin Earth case, namely to specify the direction of the asymmetric dependence by appealing to the intentions of the subjects. The puzzle is how this can be a valid move given the naturalistic conditions that restrain the theory. Presumably, if an asymmetric dependence holds between two or more laws, it does not hold *in virtue* of the intentions of the organisms that instantiate the consequent of the law. If it does, it seems that Fodor's account is not naturalistic after all.

Fodor (1990, p. 104) explains that both he and Dretske share the intuition that in cases where the alternatives (cats and robot-cats, or H_2O and XYZ) inhabit the same world, but when some speaker has learned "cat" from only one of the alternatives, the extension is disjunctive. Fodor takes the fact that only one of the alternatives actually has caused the "cat" tokens as accidental. The more important feature of the situation is the fact that the speaker *would* token "cat" even if they were to be caused by cats, though they so far only have been caused by robot-cats. So, both the 'cat → "cat"' law and the 'robot-cat → "cat"' law are in place. Fodor explains his intuition why this is not a case of the disjunction problem like this:

> It is OK for *some* predicates to be disjunctive as long as not all of them are. One can perfectly consistently hold, on the one hand, that "cat" means *robot or cat* when it's *accidental* that you learned it just from robot-cats; while denying, on the other hand, that it would mean *cat or robot* if you had learned it in a world where all you *could* have learned it from were robot-cats (e.g. because there aren't any cats around.) Similarly, Dretske can consistently hold that "water" is true of H_2O or XYZ in the case he describes while agreeing that it is true of H_2O and false of XYZ in the case Putnam describes. (Fodor, 1990, p. 104)

It seems that we have several different alternatives as to the types of worlds we are imagining. In worlds where there are cats and H_2O, and no XYZ and robot-cats "cat" and "water" mean what they do in the actual world. In worlds where there are robot-cats and XYZ and no H_2O and cats, "cat" and "water" would have robot-cats and XYZ in their extensions and therefore mean something else than they do in the actual world. In worlds where there are both cats and robot-cats, H_2O and XYZ, "water" and "cat" have disjunctive extensions, and are true of both alternatives and thereby disjunctive. In a world where "cat" expresses something disjunctive I take it that the laws that govern "cat" tokens do not exhibit an asymmetric dependence, but rather a symmetric dependence. This implies that it is true of S that she would stop using "cat" about cats if she stopped using it about robot-cats, if she is disposed to take them to be the same kind.

We assume that S believes that robot-cats and cats are of a kind. The features that distinguish the two are all internal and unavailable to S at present. But, it is presumably true that if S were to discover that cats and robot-cats indeed are different, she would stop treating them as being of a kind. Does this mean that she would decide that it was *cat* she really meant all along because she had always assumed that she was referring to, say, a biological kind and therefore not to robot-cats? Or does it mean the opposite? Is there a fact of the matter about what other intentions govern the tokens of "cat" such that it answers the question? The example does not say. But let us ask what happens when S does find out that what she took to be of a kind actually was not, and that she had been mistaken all along. Fodor asks: "If [S'] "cat" tokens meant *cat or robot*, then they were true of *both* the cats and the robots that she applied them to. Is she, then, mistaken to suppose that she was mistaken?" (Fodor, 1990, p. 105). Fodor distinguishes between what he calls an easy and an interesting answer to this question. The easy answer is that S' mistake was to not distinguish between cats and robot-cats, and that is what explains her misapplication of “cat” to robot-cats. It seems that her application of “cat” to both cats and robot-cats is true and not a mistake given her disposition to treat them as of the same kind. It is treating them as being of the same kind that is the mistake. In S’ application of “cat”, in her mouth, so to speak, “cat” meant something disjunctive.

The interesting answer clarifies this point, I think. Fodor introduces a difference between the notions “being in the extension of” and “meaning that”. We have been taken this to be roughly equivalent up until now, because it is by specifying the extension of a term that you on the informational theory determine the meaning. But meaning and extension are separate notions in the following manner. Something can be in the extension of a term without the term therefore expressing what is in its extension. Fodor is using ‘to mean x’ as meaning ‘to express the concept x’. He says:

> …if *S* used to use “cat” in the way that Baker imagines, [then] cats and robots were both in its extension. But this doesn’t, of course, imply that *S* used “cat” to express the disjunctive concept CAT OR ROBOT (i.e. to mean *cat or robot*). Quite the contrary, S *couldn’t* have used “cat” to express that concept because, by assumption, she didn’t *have* that concept. Nobody can have the concept CAT OR ROBOT unless he has the constituent concepts CAT and ROBOT; which by assumption, *S* didn’t. (Fodor, 1990, p. 105)

So, one cannot conclude from the fact that an extension is disjunctive and that what is meant by the symbol expressing the extension is disjunctive that the symbol itself is disjunctive. Primitive symbols can have disjunctive extensions (from our point of view), and so it is in this case. The mistake S made was not a misapplication of “cat” to robot-cats. Her applications were in fact true. The reason why she applied “cat” to robot cats was, according to Fodor, ‘because she took it that the robots that she called “cats” had a certain nondisjunctive property which they shared with everything else in the set {cats U robots}.’ (Fodor, 1990, p. 105). Fodor continues by saying that what she learns is that there is no such property, and that the only property shared by cats and robot-cats is the disjunctive one of being a cat or robot-cat. So, S’ mistake was not to falsely apply “cat” to robot-cats, but rather to assume that everything she applied ‘cat’ to shares a common property.

This seems to be the same point Fodor makes in discussing the case with H_2O and XYZ, i.e. that what the extension of a term contains is sensitive to the intentions of the speakers. By having the intention to use a term as a kind term the speaker is committed to applying the term only of stuff that are indistinguishable in local samples, or at a minimum that they are not obviously distinguishable. But as we observed before, intentions are not available as a theoretical notion in a naturalistic theory of content. So, how can we, by using non-intentional vocabulary formulate the same point as we have been using intentions to do? One way is to think of this use of intentions as meaning dispositions for applying terms to things. This can perhaps be construed as a psychological law that states that humans are disposed to apply kind terms only of things that are not obviously dissimilar. These dispositions can be specified in terms of counterfactuals, and, as we have seen, Fodor reserves the right to postulate what counterfactuals are to be counted as true on account of only being committed to providing a sufficient and not necessary condition for intentionality (1990, p. 94). And, as we will get back to, it is clear that Fodor assumes some facts about the psychology of both humans and other organisms in formulating the theory.

It seems clear that what counterfactuals are true of a given speaker S at some time depend crucially on the epistemic situation S is in. This means that some counterfactuals have antecedents that can be instantiated by a change in S' epistemic status. This is obvious from the fact that S' "cat"-dispositions changed when she discovered that what she had taken cats to be really were two distinct kinds of things, namely cats and robot-cats. Whether one wants to describe this as the existing disposition changing or S' acquiring a new disposition is, as far as I can see, of no consequence. But, we can ask, what happens to S when she learns that what she used to think of as cats turned out to be robot-cats? Her disposition to use the term "cat" of robot-cats is surely gone. This, presumably, translates into a new counterfactual that is true of S, e.g. were she to be exposed to a robot-cat she would not use the word "cat" of it. This is a counterfactual that, by assumption, is not true before she learns the difference, evident by the fact that she did use "cat" of robot-cats.

This section has been about various intuitions and cases that add up to being about how Fodor aims to solve the problem posed by the Twin Earth cases. The rest of this chapter will be about the role intentions play in Fodor's account. I will argue that Fodor's commitment to intentions has implications for his commitment to intentional atomism. We will also consider the proposal made some paragraphs back, namely to analyze intentions in terms of dispositions and counterfactuals and thereby satisfy the requirement that the theory should be naturalistic. "Intention", as we know, is an intentional term.

5.5. Intentions

We have seen that Fodor's solution to the Twin cases involve appealing to the intentions of the subjects. We have also seen that it is necessary for a term to be a kind term that it is governed by an intention to use the term as a kind term. In this section we will consider some implications of this view, in particular whether this view is compatible with Fodor's commitments to intentional atomism. Let us first see whether Fodor's notion of intention here can reasonably be replaced by a notion of disposition. As we established before, the notion of intention cannot have a substantive role in a naturalistic theory. Even though it can be retorted that appealing to intentions is admissible in an independent account of kind terms, and it this Fodor has and he merely uses his independently motivated account of kind terms in the context of naturalization, I argue that this nevertheless has the consequence that kind terms are incompatible with intentional atomism. If this is true it means that in a mind that has only one intentional state, this state cannot have a content that is equivalent to the content of a kind term.

Fodor sometimes appeals to the psychology of both humans and other organisms in deciding which counterfactuals are true in a given disjunction problem. The frog, he says, necessarily mistakes little black dots for flies, and that is the reason why the content of the frogs' state when it sees a fly is *little-black-dot* (Fodor, 1990, p. 108). Since it is impossible for the frog to differentiate between the two alternatives, there are worlds where there are no flies where frogs still snap at little black dots, but not the other way around. What role do these assumptions about psychology play in solving these problems? Can one suppose that frog-hood implies (ceteris paribus) a set of laws that describe the frogs' dispositions in such a way that these cases end up having the right result?

Fodor sums up his point about psychology in this paragraph. He compares the frog to Macbeth, who, famously, made mistakes concerning daggers. Macbeth's problem is analogous to the frog's in that the frog also makes mistakes, namely mistaking black dots for flies. The Macbeth case can be seen, as Fodor does, to be a problem of misrepresentation, i.e. a symbol tokening ('dagger') that is caused by something not in its extension (dagger-appearances). This is what he says:

> There is no world compatible with the perception mechanisms of frogs in which they can avoid mistaking black dots for flies. Whereas even if, freakishly, I mistake all the dagger appearances I

> actually come across for daggers; and even if, still more freakishly, I never do recover from any of these mistakes, still, that would be an *accident* since it is nomologically consonant with the way that I'm constructed that I should distinguish daggers from dagger appearances some of the time. But it is *not* nomologically consonant with the way that frogs are constructed that they should ever distinguish black dots from flies. So Macbeth and I have dagger detectors and not dagger-or-dagger-appearance detectors but frogs have black-dot detectors and not fly detectors. (Fodor, 1990, p. 108)

Let us review the case with the cats in light of what Fodor says here. When S sees what she takes to be cats, let us suppose the mental symbol "cat" is tokened in her brain. Unbeknownst to S, her "cat" symbol has a disjunctive extension and applies to objects of several kinds, more precisely both cats and robot-cats, which are indistinguishable from real cats. Not indistinguishable in principle, only their appearance and behavior are indistinguishable. What distinguishes S from the frogs is that S' psychology is such that presented with evidence that what she until now has taken to be the same kind is in fact two distinct kinds, she will stop applying "cat" in the way that she used to. Specifically, it is true of S that when she encounters a robot-cat she will no longer apply the term "cat" to it. What are responsible for S' ability to revise her own applications of "cat" are features of her psychology, specifically the features that govern kind terms. Earlier we saw that Fodor appealed to S' intentions for explaining this, and that this, arguably, is not allowed on a naturalistic account.

We are trying to determine what features of psychology that can be assumed on this picture? There is reason to be restrictive, because the theory Fodor proposes claims to be not only sufficient and naturalistic, but also atomistic (Fodor, 1994, p. 6). Atomism is the view that systems can have only one intentional state. If the content of a kind term is dependent on the intention of treating the term as a kind term, it seems that a system that has an intentional state that has the content of a kind term necessarily needs to also have another intentional state: the intention that secures that the first intentional state is a kind term. The consequence is that intentional states that have kind term contents imply at least one other intentional state. This amounts to the system not being atomistic in the relevant sense. I think this is a serious challenge for Fodor's commitment to intentional atomism. Kind terms are abundant in both natural languages and, it seems, in our conceptual vocabularies. We have seen that kind terms are terms we require to be about things that are relevantly similar to each other. It is safe to assume, I think, that we can conclude that if Fodor uses intentions in accounting for kind terms and contents that amount to the contents of kind terms, then these contents will not be admissible to figure in an atomistic system. And just as important, they cannot figure in examples of atomistic systems such as this:

> Intentional theories are, on the face of them, *atomistic* about content. If all that matters to whether your thought is about dogs is how it is causally connected to dogs, then, prima facie, it would be possible for you to have *dog* thoughts even if you didn't have thoughts about anything else. (Fodor, 1994, p. 6)

But, as we have seen, thoughts about dogs are not just determined by their connection to dogs, they are partly determined by the intention to treat dogs as a kind. The content of "dog" and other kind term contents cannot figure in atomistic system, or so I have argued.

5.6. Conclusion

In (Fodor, 1990, p. 115 - 116) Fodor seems to think of the intentions governing kind terms as dispositions for applying the kind terms, and uses the terms interchangeably. Let us see if we can determine if intentions can be reconstructed in terms of dispositions. We have seen that in the original Twin Earth example it is assumed that the speaker intends the term used as a natural kind term, with the implication that the term will not be used of obviously dissimilar things. This has the result "of making the semantically relevant asymmetric dependencies true of his use of the word." (Fodor, 1990, p. 116) He continues by saying that Baker doesn't say if S uses the word "cat" as a kind term or not, and that he assumes that S doesn't in her example.

> So, Baker's "cat" means *cat or robot* because, on the one hand, *S* would (indeed does) use "cat" for either; and, on the other, there's nothing in Baker's description of the case that suggests a mechanism (such as an intention to use "cat" as a kind-term) that would make the use for the robots asymmetrically dependent upon the use for the cats (or vice versa). (Fodor, 1990, p. 116)

Fodor assumes that S does not have an intention to use "cat" as a kind term (1990, p. 116). How reasonable is this claim? Even though Baker does not explicitly state that S does intend to use "cat" as a kind term it seems from the description of the case that this is what S does. Her response to finding out that cats and robot-cats are actually very dissimilar is evidence of this. To assume that if S had no intention of using "cat" as a kind term then she would not stop using "cat" of both cats and robot-cats when she found out that they are dissimilar is plausible, I think. It is plausible because the counterfactual 'if S finds a dissimilarity between cats and robot-cats she will stop using "cat" indiscriminately of them', one that is true by assumption, plausibly implies that the likeness of cats and robot-cats is relevant to S' use of the term "cat". So, if we're assuming that the fact that S stops using "cat" of robot-cats is not an arbitrary fact about S, but something that is consonant with S' psychology, then we can assume that S has a standing disposition to use “cat” as a kind term, since this disposition just is to stop using a kind-term of things that one discovers are unlike.

It is not clear that Fodor does, in fact, assume that intentions can be analyzed in terms of dispositions and counterfactuals. I think it is more reasonable to think that appealing to intentions is Fodor’s way of accounting for kind terms in general. The appeal to intentions is then imported in to the theory of content in accounting for the content of kind terms. I have argued that this implies a challenge to Fodor’s atomism theses. To assume that intentions can be reduced to dispositions is tantamount to solving the naturalization problem by assumption. I do not think it is likely that Fodor does this. Now that we have seen how Fodor deals with the Twin Earth cases we have stated the theory with enough precision to consider some arguments that are critical of his position. In the next chapter we will mainly consider Paul Boghossian’s argument that Fodor’s theory is really a type I theory in disguise, and that being a type I theory it is subject to problems. By reviewing this objection, and related objections, I hope to be able to assess the features of Fodor’s theory we have focused on in this thesis. These features, naturally, are those that concern the theory’s ability to account for the three intuitions we mentioned at the outset, except the implication of pan-semanticism which we concluded chapter 4 by arguing that Fodor does not convincingly account for. So, the next chapter will mainly be about trying to asses Fodor’s theory in light of the intuitions that meaning should be robust and the intuition that a theory of meaning should be accounted for naturalistically.

6. Chapter VI: *Objections and Replies*

Fodor’s theory has been widely criticized by many different people, for many different reasons. In this chapter we will review an objection made by Paul A. Boghossian in the book *Meaning in Mind* (Loewer and Rey, 1991). The book is a collection of articles criticizing many aspects of Fodor’s theory, and Boghossian’s is of particular interest to us. It is interesting in particular because the objections are aimed at the parts of Fodor’s theory that we have been considering, in particular issues surrounding the verificationist aspects of the theory. We have also discussed some of the modal features of Fodor’s theory, and in this section we will consider Boghossian’s objection to how Fodor determines distance between the possible worlds. Boghossian’s argues that Fodor’s theory does not account for the robustness of meaning. It does so by constructing a version of the Twin Earth case that is supposed to imply that Fodor’s theory cannot in that case ascribe correct extensions to kind terms. As we will see, Fodor tries to block Boghossian’s argument by attacking the main premise, but I will argue that he is unsuccessful. We will also find reason to believe that Boghossian’s objection poses a real challenge to Fodor’s theory.

Boghossian’s argument is a complex one, and we will go through it in detail. As we said, Fodor has responded to the argument so we will review his response as well. Boghossian argues that the asymmetric dependency theory is a type I theory in disguise. This is the premise for all the other arguments in his article and thus the most important aspect of his argument against Fodor. Fodor, as we saw earlier, argues extensively against type I theories. Boghossian agrees with Fodor’s arguments against the teleological type I theories and considers them to be fatal (Boghossian, 1991, p. 68). Fodor argues, as we saw, that it is not only the teleological variant of the type I theory that is wrong, rather it is the whole project of defining a meaning-determining situation where what causes a symbol necessarily is in the symbols extension, i.e. relying on a substantive distinction between type I and type II situations in the first place. Boghossian claims that this is just what Fodor does, and thus is subject to his own arguments. Boghossian’s main argument is, as we will see, that any specification of a type I situation will not yield unique content ascriptions because it cannot rule out the possibility that the symbol is caused by something not in its extension. We remember from before that type I theories are theories that define a meaning determining situation where the symbol necessarily is caused by what is in its extension. For example, in a type I situation, if the symbol “dog” is caused it was necessarily caused by dogs. Boghossian’s argument aims to show that such a situation is impossible, with the consequence that no type I theory produces correct content ascriptions. As we will see, the reason for this is the type I theory’s inherent verificationist aspects. Boghossian aims to offer a counter-example to Fodor’s theory by showing (i) that Fodor’s theory is a type I theory, and (ii) that type I theories do not produce univocal content ascriptions. If this argument is true, the conclusion is that Fodor’s theory has not accounted for the robustness of meaning and is thus not a satisfactory account of meaning. Let us see the first stage of Boghossian’s argument.

6.1. Fodor and Type I Theories

What reason does Boghossian have for claiming that Fodor’s theory is of the type I kind? We have seen that type I theories are theories that define a situation that is such that only the referent of a symbol can cause the symbol to be tokened. This is to say that only what is in the extension of the symbol can cause it (here we mean extension in the normal sense, i.e. dogs, and nothing else, is in the extension of “dog”). As we have seen, the teleological version of this approach suffers from several difficulties. We have focused on that the teleological theories do not provide a solution to the disjunction problem because they make the content ascriptions of the mental states depend on univocal functional ascriptions to the mechanism that realizes the mental state. We concluded part I of the theses with the result that Fodor’s arguments against this approach are convincing.

Boghossian makes a point that Fodor does not consider very much, namely the assumption that the cognitive mechanisms are never selected for hiding rather than tracking the truth (Boghossian, 1991, p. 68). I think that this is a valid concern, both in considering the functional individuation of biological kinds, but also considered as a general intuition. This is for the same reasons I outlined earlier in considering the intuition that when a belief is false, something in the belief forming mechanisms have malfunctioned. We will not pursue the matter further here. Boghossian also mentions perhaps the most important counterexample to type I theories, namely tokenings in thought that are caused by other thoughts. We saw this point earlier, but it is worth repeating. This point challenges the assumption that, ‘”when things go right” S will be tokened only in application to its referent’ (Boghossian, 1991, p. 68). Thoughts can cause other thoughts without thereby being tokened only in application to their referents. We saw that horse thoughts can be caused by cowboy thoughts. This is the most important aspect of what we have been calling the robustness of meaning. Thoughts mean what they do regardless of what happens to cause them in a particular instance. The teleological theory cannot account for this feature of meaning. This is not because it is teleological, but because it is a type I theory. Type I theories are dependent upon being able to non-intentionally and non-semantically specify a situation where a symbol is guaranteed to be caused by its referent, a specification we have seen convincing reasons not to believe is forthcoming.

The asymmetric dependency relation is what, for Fodor, is supposed to determine what causal relation is the meaning-determining one. S means P if “S is caused by X” is asymmetrically dependent upon “S is caused by P”. This means that if P is not able to cause S then nothing is. But not the other way around, i.e. any non-P can stop being able to cause S without this effecting P’s ability to cause S. Boghossian observes a possibility implicit in this definition it is important to get out of the way, so to speak, before we move on (Boghossian, 1991, p. 70). Fodor also considers this possibility (Fodor, 1990, p. 108-110). The definition seems to open for the possibility of symbols being dependent on their proximal causes and not their distal ones. The reason is this: if one makes the proximal cause of a symbol, say, a horse image on the retina, unable to cause “horse” tokens, then horses are also made unable to cause “horse” tokens, and presumably nothing will cause “horse” tokens. And conversely, if you make the horse unable to cause “horse” tokens, then, presumably, horse images on the retina will still cause “horse” tokens. This is plausible because it is plausible that horses cause “horse” tokens in virtue of their proximal projections, i.e. producing horse-images of some kind on the retina. Since there are indefinitely many ways horses can produce proximal projections, this leads us to conclude that horse caused “horse” tokens are asymmetrically dependent upon a disjunction of proximal projections. So the meaning of “horse” is not *horse*, but the disjunction of proximal projections. It is important to note that nothing depends upon that the proximal projection is some kind of image on the retina. The important point is that the proximal cause of the symbol tokening is the asymmetric dependency base for the other causes, and thus constitutes the meaning of the symbol. This is, of course, the wrong result for the theory to imply.

Fodor answers this objection (Fodor, 1990, p. 108-110) by pointing out that just about any proximal cause can cause a symbol tokening like “horse”. This is because belief fixation is open-ended and holistic (Boghossian, 1991, p. 70). This means that “horse” tokens are theory mediated, so that what you take to be evidence for there being horses about, depends upon your theory about horses. This has the result that any evidence can be taken as evidence that there are horses about and then cause a “horse” tokening. Fodor claims that the disjunction of proximal projections is not eligible as an asymmetric dependency base because the disjunction is open, and open disjunctions cannot be the referents of primitive symbols. At least, that is what Boghossian takes Fodor to mean (Boghossian, 1991, p. 70). A more important reason is that proximal causes are not eligible for entering into the causal generalizations which are potentially meaning-determining, i.e. causal laws. What we are left with as candidates for the asymmetric dependence bases are the distal causes.

6.2. Possible Worlds and the Distance Metric

Boghossian makes a remark that is clarifying, I think. The theory states that “cow” has cows in its extension but not cats because the cat → “cow” law is asymmetrically dependent upon the cow → “cow” law. He says:

> Put in the language of possibilia, this seems best interpreted as suggesting that “cow” has cows in its extension and not cats, provided that although there are worlds in which *cow* can cause a “cow” token but *cat* can’t, there are no worlds in which *cat* can cause a “cow” token, but *cow* can’t. This, however, can’t be precisely what’s meant. The point is, of course, that even by the theory’s own lights, there have to be *some* worlds in which the property of being a cat can cause “cow” tokens even if the property of being a cow can’t, for there presumably are some worlds in which “cow” means *cat*. (Boghossian, 1991, p. 71)

He goes on to say, by quoting Fodor, that the theory must be stated more accurately to reflect this. The more accurate version is this: there needs to be a world W where (i) cows cause “cow”s and non-cows don’t; and (ii) W is closer to the actual world than a world where some non-cows cause “cows” and no cows do (Boghossian, 1991, p. 71). This has the effect of introducing a nearness clause for the possible worlds we are considering. What this means is that the relative distances between the possible worlds and the actual world is significant in determining content. There is very little agreement on what criterion for determining distance between possible worlds is the best one. The term “distance” in this context is used to capture differences between the possible worlds. Worlds that are significantly different from the actual world are thought to be farther away than worlds that are similar. Possible worlds are individuated by what counterfactuals are true in them. For example, there is a possible world where it is true that Al Gore is president. This world can be quite similar to the actual one, whereas a world where it is true that humans have wings is quite different. So the world where humans have wings is farther away than the world where Al Gore is president because the fact that humans having wings is a bigger change in our planets history than Al Gore being elected president instead of George W. Bush is. These distances are supposed to reflect our intuitions about how much change particular changes imply. These intuitions may include such intuitions that changing a biological fact implies a bigger change than changing a historical fact; changing a physical fact (for example making sugar not soluble in water) is a bigger change than a biological change etc.

The result is that the world where you have to change more to make the counterfactual true is the one that is the farthest away. We will see, in considering Boghossian's argument, that there are semantically significant cases where this intuition of distance is problematic. What is important to note is not only, as in (ii), that the relevant world need to have some distance relation to some other worlds, but, as in (i), that there must be a world where cows cause "cow"s and nothing else does.

6.3. Boghossian's Argument

Boghossian claims that Fodor's theory, though it does not look like it, is a type I theory (Boghossian, 1991, p. 71). Fodor states his theory in terms of dependencies among causal laws which in turn are stated in terms of counterfactuals. So, the theory does not contain any references to situations where the referents of a symbol are the only thing that can cause it, which is what Boghossian takes type I theories to be (Boghossian, 1991, p. 71). This is also what Fodor takes type I theories to be (Fodor, 1991, p. 272). The question is whether saying that there is one unique world where only the meaning-determining law is in place, and none of the other laws that depend on the meaning determining law is tantamount to specifying a type I situation. Boghossian wants to show that if there is such a world, a world it seems Fodor is committed to the existence of, then to specify such a world is tantamount to specifying the type I situation required. And it is hard not to agree, I think, that to specify a world in which the only law that governs "horse" tokens is the one that has horses as the only possible cause, is tantamount to specifying a situation where a symbol can only be caused by something in its extension, which, as we saw, is the condition on being a type I theory.

Boghossian's argument depends on establishing that asymmetric dependence necessarily presupposes a world where the only causal relation that possibly can cause a symbol is what is in the symbols extension. Boghossian aims to show this by showing that the only plausible alternative to presupposing such a world will not result in an asymmetric base for the symbol. If Boghossian can show that a commitment to the asymmetric dependence relation necessarily implies specifying a unique world where the only cause of S is P, then asymmetric dependence is a type I theory. Boghossian thinks he can show this by showing that the only other plausible alternative for formulating the asymmetric dependence relation between laws, i.e. without appealing to worlds that can be characterized as type I situations, is unable to produce dependencies that are meaning-determining and establish unique laws as the dependence base for symbol tokenings (Boghossian, 1991, p. 72). What is this other option?

Boghossian thinks that the only other plausible way of establishing an asymmetric dependency base is "… a distribution of nearby worlds which contains worlds in which both P and Q can cause S tokens, and worlds in which both P and R can, but no world in which only P can and no world in which only non-Ps can …" (Boghossian, 1991, p. 72). The problem with this proposal is that we have no way of determining that it is the P → S law that is the dependency base and not the (P & R) v (P & Q) → S law. Since the theory does not provide a condition for deciding these cases we are forced to conclude that the dependence base in this case is disjunctive. Boghossian (1991, p. 72) considers the worry about how we can be sure that whenever we have P as a candidate as the dependency base for S we will also have (P & R) v (P & Q) as a candidate. One might worry that Boghossian's case is not general. However, Boghossian thinks that this worry is irrelevant because the case is not required to be general. Since Fodor is offering a sufficient condition Boghossian thinks that "… any case in which S has both P and (P & R) v (P & Q) as candidate asymmetric dependence bases will be a case in which [Fodor's] theory yields either the wrong result or an indeterminate one." (Boghossian, 1991, p. 72). He continues by saying that the only requirement is one case where (P & R) v (P & Q) → S is a law whenever P → S and R → S is laws, and that he finds it obvious that there will be plenty such cases (Boghossian, 1991, p. 72). The problem for Fodor is, of course, that Boghossian's alternative entails ascribing a disjunctive extension to P. The argument is intended to force Fodor to admit that his theory is a type I theory by eliminating his options. I do think that this is a worry Fodor needs to take seriously.

Let us, for the sake of argument, assume that this argument establishes that Fodor's theory is a type I theory and consider the rest of Boghossian's argument. On the basis of this premise Boghossian wants to argue two further claims: (i) the enterprise of naturalistically specifying a circumstance in which a symbol necessarily is caused by its referent (a type I theory) is impossible; and (ii) even if it were possible to specify such a circumstance we could never certify that we had succeeded in specifying them. The first claim is to establish that type I theories are impossible, and the second is that even if they were possible, we would never be able to know if we had succeeded in specifying one. Claim (ii), though interesting, will not be considered here.

6.4. The Argument for the Impossibility of Type I Situations

To establish that type I theories are impossible one needs to argue that the situation that type I theories depend upon, i.e. the situation where a symbol can only be caused by its referent, never can arise. This is achieved, we can assume, if it is shown with generality, that whenever a symbol is tokened in such a situation as specified, there are always more candidates than one for being the asymmetric dependence base for the symbol. If this is shown it will establish that no unique law serves as the asymmetric dependence base for any symbol, and all symbols will have disjunctive contents.

As we have seen, Boghossian claims that it is impossible to specify a situation such that a symbol necessarily is caused by its referent, or, as is the same, specifying a type I situation. What is the argument for accepting this conclusion? Boghossian says:

> … any situation in which X is a possible cause of my S-tokens is also a situation in which any other property Y, indistinguishable from X in all physically possible circumstances accessible to me, is *also* a possible cause of my S-tokens. Since there are no physically possible circumstances accessible to me in which X and Y can be told apart, any circumstance in which X can cause my S tokens is also a circumstance in which Y can. It follows, therefore, that on a type 1 theory, if S has X in its extension, then it also has all these other "X-equivalent" properties equally in its extension. But is this plausible? Is it really true that my having a symbol that means X but not Y depends on my being able to tell Xs and Ys apart? After all, it surely doesn't follow from the fact that Xs and Ys can only be told apart in worlds that are too far for *me* to get to, that being X and being Y are the same property. But, then if the difference between being X and being Y is real, then so too, presumably, is the difference between being X and being (X or Y). And if this difference is real, then why shouldn't we be able to think in ways that respect that difference? (Boghossian, 1991, p. 73)

This paragraph introduces several key ideas in the problems we are considering. We recognize from the discussion of the Twin cases that there are reasons for thinking that what a symbol means may depend on what one thinks fall under the same kind, and, that at least some symbols, i.e. symbols that stand for kinds or have the contents of kind terms, get their meanings partly from the fact that we can distinguish their referents from other kinds of things. We also remember Fodor distinguishing between cases where two kinds are mistakenly taken to be the same as cases where the mistake is accidental and cases where the mistake is necessary.

A possibility in this way of setting things up we have not considered yet is the possibility of worlds that are too far for us to get to. But what is meant by the phrase "too far for us to get to"? Boghossian distinguishes between physically possible worlds and physically possible worlds that we can get to (1991, p. 75). As I understand it, the physically possible worlds that we can get to are worlds that are physically possible and where the circumstances are such that we can survive there. The worlds that are inaccessible to us seem to be worlds, though physically possible, where we will not survive in principle, i.e. where our biological and medical limitations don't allow us to go (Boghossian, 1991, p. 75-76).

Boghossian, like Fodor, take natural kind terms to be paradigm examples of terms to be considered in the context of naturalization. Boghossian lists some intuitions he thinks are plausible when it comes to the semantics of kind terms that Fodor's theory should account for (Boghossian, 1991, p. 74-75). The kind term should be such that (i) "some sort of basic naturally explanatory property unite all the things that are correctly said to fall in its extension" (Boghossian, 1991, p. 74); (ii) it should be "[t]he kind exemplified by all, or better *most*, of certain local samples" (Boghossian, 1991, p. 74); and (iii) "If investigation uncovers that there is no single hidden structure uniting the local samples, but that there are two (or so) well-defined such structures, then there appears to be a temptation to say that the expressions has both of those structures in its extension" (Boghossian, 1991, p. 75). He also notes that if there is no common property at all, only a "messy motley of basic particles" (Boghossian, 1991, p. 75) in the local samples, then the term fails to refer. He uses 'jade' as his example of (iii). 'Jade' is a kind term that has two chemically different substances in its extension, namely jadeite and nephrite (Boghossian, 1991, p. 75), and thus seems to be a disjunctive natural kind concept.

These are intuitions that Boghossian argues are plausible, but that information based semantics cannot account for. The reason for this is a special case of the Twin Earth case. In the normal Twin case we have two substances, H_2O and XYZ. The kind term "water" functions so as to pick out all the things that are of the same natural kind as the local samples, i.e. H_2O in the actual world. The normal Twin case we are considering here is no counterexample to Fodor's theory because you can, under the right circumstances, tell H_2O

and XYZ apart. This is so even though in encountering XYZ one is normally disposed to wall it “water”. The circumstances in which you can tell H_2O and XYZ apart are circumstances humans can survive in, and therefore physically possible in the required sense. Presumably, a chemist with the relevant know-how could distinguish them easily. The point is that “water” does not have both H_2O and XYZ in its extension because the world where we can tell them apart is physically accessible to us, and were we to go there and discover that they are different substances we would stop using “water” of XYZ.

6.5. ABC

Now to the main part of Boghossian’s argument. Boghossian wants us to consider a case where to distinguish the substances in this manner is not an alternative. In this case, he says (Boghossian, 1991, p. 75), we have a substance that is just like H_2O and XYZ in having all the same macro properties, e.g. you can drink it, it boils at 100°C, fish live in it etc. Boghossian calls this substance ABC. The only difference is that the worlds you need to get to to distinguish ABC from H_2O and XYZ are worlds that you can’t get to, on account of the extreme conditions there. The example Boghossian uses is that you can only distinguish ABC from H_2O in gravitational fields that are characteristic of the insides of black holes, which is a place that is physically possible, but also where humans in principle cannot survive. The question is if ABC is in the extension of “water”. If it is, then “water” has two distinct kinds in its extension and is disjunctive in the same sense as S’ original “cat” concept, i.e. as a primitive concept with a disjunctive extension, but with the difference that the mistake is necessary.

The role the notion of physics plays in this example is somewhat puzzling, I think. First of all we have the difference between physically possible worlds and a subset of these worlds that are the physically possible worlds we can get to. This difference, I think, is initially fairly intuitive. There are, after all, conceivable circumstances that are physically possible but where humans would not survive. And, where humans cannot survive, humans cannot perform experiments to show that two things are kind-distinct. And if two things are kind-distinct but cannot be verified as such then our language dispositions cannot respect that difference. As we have seen, this comes down to which counterfactuals are true of an individual in a world, and in these cases it seems that all the counterfactuals that govern the tokening of “water” of H_2O and not of ABC are false in worlds that can be inhabited by us.

But this is a strange view of physics. First of all, it is a view that takes for granted that all kind-distinctions must be made on the basis of experiment. This is, as it stands, not unreasonable. But in this context this implies that, since there are circumstances where humans cannot perform experiments or observations, there are physically possible worlds that are impossible to describe in physical language. This follows from Boghossian's view since he claims (i) that ABC is a physical kind, and (ii) that it can in principle not be discovered experimentally. Given the premise that kind distinctions must be made on the basis of experiment, it follows that on Boghossian's view, ABC is, though a physical substance, it is physically indescribable. ABC is, after all, a substance that we necessarily cannot discover. It is also a substance that cannot be discovered indirectly, as it were. To discover something indirectly means in this context that you can discover it on the basis of examining your existing theory for inconsistencies or weaknesses and, as a solution, we can imagine, postulate some theoretical kind that solves the problem. I take it that scientific history is full of this kind of indirect discovery. The discovery of Neptune comes to mind as such a case. ABC can in principle not be discovered like this. Objects that are discovered indirectly have features that are such that they can be discovered, given the right experimental environment. There are no experimental environments where ABC can be discovered, by assumption. This, I think, is the puzzling part about Boghossian's thought experiment. What is a physically possible world like that contain kinds that cannot, in principle, be described physically? What makes it the case that this world can be said to be *physically* possible? I think it is not unreasonable to say that physically possible worlds are worlds where the laws and the individuals subsumed by those laws are not inconsistent with some ideal physics humans ideally can provide. I am puzzled by that the worlds Boghossian postulates as physically possible does not satisfy this condition. If the ABC worlds are worlds that are in principle not describable in even an ideal physics, by what criteria are they *physically* possible? I take Fodor to express a similar worry in (Fodor, 1991, p. 275).

This worry can be addressed by relaxing the condition that kind-distinctions must be established experimentally. But this is just to relax the condition that the world in which the kind-distinction is detectable must be available to us, and, as we have seen, that is a crucial premise of Boghossian's argument. I make this point as an observation and will not pursue it here. Fodor also expresses doubts as to whether this feature of the though experiment is plausible (Fodor, 1991, p. 275). Let us return to the main argument.

6.6. Verificationism and the Actual History Condition

That informational semantics imply verificationism is something Fodor is explicitly aware of (Fodor, 1990, p. 120). In fact he even spells out the argument for it:

> Such theories [informational ones] distinguish between concepts only if their tokenings are controlled by different laws. Hence only if different counterfactuals are true of their tokenings. Hence only if there are (possible) circumstances in which one concept would be caused to be tokened and the other concept would not. […] …, i.e. a world where H_2O and XYZ are distinguished (a forteriori, a world where H_2O and XYZ are distinguishable. That is how you get from informational semantics to verificationism. Correspondingly, the way you avoid the verificationism is: You relax the demand that semantic relations be construed solely by reference to subjunctive conditionals; you let the actual histories of tokenings count too. (Fodor, 1990, p. 120)

The verificationist aspects of the theory commits the informational semanticist to the principle that 'If *"X"* expresses at least *X*, and if there is a *Y* which it is not nomologically possible for you to distinguish from *X*, then *"X"* expresses *Y* as well as *X* (e.g., it expresses the disjunctive property *X or Y.*)' (Fodor, 1990, p. 119). And this is the feature of Fodor's theory Boghossian's argument seeks to exploit. The problem is that if one cannot distinguish ABC from H_2O, then both H_2O and ABC are in the extension of "water", and this is intuitively not the right result, Boghossian thinks. One of the intuitions we postulated that the theory should respect is that kind-terms like' "water" either picks out something that is exemplified by most of the local samples, or it picks out nothing at all. And, it is hard not to agree with Boghossian that "water", intuitively, does not have a disjunctive extension.

Fodor's response to the verificationism worry is to introduce a new condition he calls "the actual history condition". This condition postulates that if S is to express P then some Ss are actually caused by Ps (Fodor, 1990, p. 121). This seems to exclude ABC from the extension of "water": since ABC is only a possible and not actual substance, no "water" token has ever been caused by ABC. Given the actual history condition, ABC is thus not in the extension of "water". But the actual history condition has some disadvantages over the pure informational approach. One of the great advantages of the informational approach is that it can treat uninstantiated properties the same way as it treats instantiated properties. Since all that is relevant for the tokening of a symbol is the law that governs the tokening, and that law is accounted for in terms of subjunctive conditionals, there is nothing that precludes us from accounting for the symbol "unicorn" by saying there is a nomic connection between unicorns and "unicorn"s. We can say that "unicorn" means what it does because it would be caused by unicorns if there were any. On the actual history condition this is not an option. Unicorns don't exist and therefore cannot have caused "unicorn" tokens.

As Boghossian points out (Boghossian, 1991, p. 77), this seems to have the consequence that there is a connection between being an instantiated property and being the cause of a primitive symbol. He says: "… there seems no reason to believe that every concept that has an empty extension in the actual world will turn out to be *complex*." (Boghossian, 1991, p. 77). This has the potential implausible consequence that there is a connection between being an uninstantiated property and being a complex symbol. Boghossian thinks Fodor owes an argument for the plausibility of this (Boghossian, 1991, p. 77). I do agree with Boghossian that Fodor owes an argument for this. But I do not think that it is prima facie not plausible that there is a connection between being an instantiated property and being a simple symbol. Perhaps one can make a case for the claim that all simple symbols must have a corresponding instantiated property. From there one can by deducing that no simple symbol has a corresponding uninstantiated property conclude that all uninstantiated properties, if they correspond at all, correspond to complex symbols. There is nothing, on the face of it at least, why this approach is implausible.

A second difficulty is that it is not obvious that the actual history condition actually solves the problem. Boghossian considers the case where ABC actually exists in the actual world, in trace amounts (Boghossian, 1991, p. 77-78). The idea is that ABC stands to H_2O as iron pyrites stands to gold, i.e. as an impurity that is not in the extension of the kind term, but that still occasionally is causally responsible for causing the tokening of the kind symbol. So ABC exists all around us, we are supposing, but only in trace amounts. We are assuming that this is sufficient for that some "water" tokens have actually been caused by ABC, and that this satisfies the actual history condition. This seems to imply that "water" has both H_2O and ABC in its extension, and this is, Boghossian says (1991, p. 78), contrary to intuition, and thus constitutes a counterexample to Fodor's theory. The consequence of this seems to be that the actual history condition doesn't produce the effect Fodor hopes, namely to secure univocal content ascriptions even in cases where one cannot verify that the asymmetrical dependence base consists of a single law. This constitutes Boghossian's full argument against Fodor's view. I do not think that Fodor can reply effectively to this challenge, and in considering Fodor's response I will try to formulate why.

6.7. Fodor's Response

How does Fodor respond to this argument? Fodor responds by attacking the argument that shows that the asymmetric theory is a type I theory, i.e. the crucial premise of Boghossian's argument. We remember that the argument tries to show this by showing that the only plausible alternative to the world where P → S is the only S causing law is not available as an asymmetric dependency base. The thought is that if the asymmetric dependency base of all the laws that causes S is defined as a world where the only cause of S is P, then, that is in effect to say that this world is a situation where the only thing that can cause a symbol is its referent. This is, as we saw, equivalent to the definition of a type I situation. If Fodor can undermine this argument he has shown that his theory is not a type I theory.

For Fodor to undermine the argument he must show that Boghossian's argument does not force him to accept the conclusion, i.e. that the asymmetric dependence base for S is a world where the only thing that can cause S is P, which is the referent of S. He must do this because the alternative is to admit that the theory is a type I theory. Boghossian's argument consists in assuming that Fodor denies this conclusion and then considering Fodor's options. Boghossian thinks Fodor has one option, namely to insist upon "a distribution of nearby worlds which contains worlds in which P and Q can cause S tokens, and worlds in which both P and R can, but no world in which only P can and no world in which only non-Ps can, …" (Boghossian, 1991, p. 72). But this is tantamount to saying that the asymmetric dependence base for S tokens is (P & R) v (P & Q), which does not yield P as the unique cause of S tokens, and hence not the meaning of S. Boghossian concludes, reasonably, that Fodor's theory is a type I theory because the only alternative specification of asymmetric dependence bases is not sufficient for producing robustness. This, we remember is the chief condition a theory of meaning must satisfy.

As we have set up the argument here it seems reasonable that the way for Fodor to disarm the argument is either to show that there is indeed a way of specifying the asymmetric dependence base for S other than by specifying a world where the only cause of S is P, which amounts to finding a better alternative than Boghossian's alternative, or by somehow denying that the way we have set up type I theories here is in fact what is meant by type I theory. These, as far as I can tell, are the only options for countering Boghossian's objection.

Fodor chooses a different strategy, one I find puzzling. Call the law (P & R) v (P & Q) → S for C. We remember that C is Boghossian's alternative asymmetric dependence base to admitting an asymmetric dependence base that is equivalent to a type I situation. As we have been setting things up, we have found reason to accept that C is the most plausible alternative Fodor has other than admitting to a type I theory. To accept that C is the asymmetric dependence base for S is, of course, to admit that S means something disjunctive, but this is presumably a problem that could be solved by the actual history condition. So, admitting that C is an alternative to P as the asymmetric dependence base for S tokens is presumably not fatal to Fodor's theory.

The puzzling part about Fodor' reply is that this is not the approach he takes. Fodor argues against that C should be regarded as a candidate for the asymmetric dependency base for S (Fodor, 1991, p. 273), and he does so very convincingly. He says this:

> Notice, to begin with, that for S to mean P in this world, P → S has to be (not just true but) lawful in (relevant) worlds where other X → S laws fail. The argument for this is straightforward. The intuition that underlies the asymmetric dependence story is that S's meaning P depends on P → S but not on any other X → S connections. But this implies that any other X → S can fail consonant with whatever is required for S to mean P. But for S to mean P, Ss must carry information about P; and Ss to carry information about P, [P → S] has to be a law. So X isn't a candidate for what S means in W_i unless X → S is a law in W_i. So, then, the question arises whether we're guaranteed that C is a law in W1 and W2; and the answer is that we aren't. (Fodor, 1991, p. 273)

There are two main reasons why C is likely not to be a law (Fodor, 1991, p. 273). The first reason is one we looked at when we considered whether proximal causes could be the antecedents of laws. Fodor considers the disjunctive antecedents of such laws to be subject to the same considerations that makes proximal causes not eligible as antecedents in laws, namely that the disjunctions are open. He does not say why he thinks this is so, and I don't quite see how the cases are connected. But, we will not pursue the matter more here. The second reason is that it seems that Boghossian makes some assumptions about what laws imply, assumptions Fodor can deny. Fodor (1991, p. 273) takes Boghossian to assume that the inference from if X → S and Y → S are laws to the conclusion that it follows that X & Y → S is a law holds. Fodor makes the observation that "*ceteris paribus* laws are notorious for not satisfying this pattern of inference" (1991, p. 273) as an example to show that the inference is likely not to hold. This is, I think, a valid concern on Fodor's part, but I do not think he conclusively shows that there cannot be a law like C that has a disjunctive antecedent that satisfies both the condition that it should be a proper law, and that its antecedent is not open.

What is confusing in this discussion, I think, is the dialectic of the exchange between Fodor and Boghossian. Fodor says:

> I conclude that Boghossian has given us no reason to believe that C will be a candidate for determining the meaning of S whenever P → S is a candidate for determining the meaning of S. From which I further conclude that Boghossian has given us no reason for supposing that the asymmetric dependence story is committed to type I situations. (Fodor, 1991, p. 273 – 274)

The puzzlement stems from the fact that this conclusion does not seem to follow from the way we have set up Boghossian's argument. We have considered C as the only plausible alternative Fodor has as an asymmetric dependence base if Fodor wants to deny that his theory is not a type I theory. It seems that C, being a law that secures that Ps can cause Ss without relying upon specifying a situation or world where the only thing that can cause Ss is Ps, is a candidate Fodor can employ to demonstrate that his theory in fact is not a type I theory. Instead, Fodor seems to treat Boghossian's argument as a disjunction problem where he needs to dispel the disjunction, something he arguably succeeds in doing. But, in doing so he does not actually answer Boghossian's objection. Boghossian's objection, as we have seen, assumes that Fodor wants do deny that his theory is a type I theory. This implies, Boghossian believes, that one needs another way of specifying the asymmetric dependence base of a symbol than appealing to the nearest possible world in which all S tokens are caused by Ps, since this proposal is equivalent to the type I way of specifying the asymmetric dependence base. Boghossian then considers what he thinks is the most plausible option and argues that it also is insufficient for being the asymmetric dependence base for S tokens. In effect, Boghossian wants to show that specifying a situation (world) where all S tokens are caused by Ps, i.e. being a type I theory, follows *necessarily* from Fodor's theory of asymmetric dependence, because no other option is available for specifying the required asymmetric dependence base.

The reason why Fodor's reply is puzzling, I think, is that Fodor does nothing to counter Boghossian's argument. He, in fact, strengthens it by adding arguments to why Boghossian's alternative to being a type I theory is flawed. If 'We want it to come out that ["cow"] means [cow] because *cowhood* is the only property whose instances cause "cow"s in every world where anything does' (Fodor, 1991, p. 273) which is Fodor's revised condition after he has replied to Boghossian's objection, unpacks as something other than "… (i) in W, cows cause "cow"s and non-cows don't; and (ii) W is nearer to our world than any in which some non-cows cause “cow”s and no cows do” (Fodor, quoted in Boghossian, 1991, p. 71), which is how Boghossian’s argument presupposes that it does not, then Fodor owes an argument for the difference, I think. In the next section we will consider some difficulties relating to relying on distances between possible worlds in specifying the correct asymmetric dependence bases. As we will see, Boghossian argues that such distances cannot be specified naturalistically.

6.8. The Distance Metric

We mentioned above that there are problems connected to the theory’s modal aspects, and especially connected to the conception of distance between various possible worlds. Boghossian is skeptical about whether the relevant “nearness”-relation can be specified non-question-beggingly, i.e. in non-semantic and non-intentional terms (Boghossian, 1991, p. 81). He formulates the problem like this:

> Clearly, everything depends on whether the relevant similarity relation can be specified non-question-beggingly – without the benefit of sidelong looks at the meanings of the expressions in question. What the success of Fodor’s theory depends on, in other words, is that when nearness of worlds is judged from a purely non-semantic and non-intentional – for our purposes, therefore, from a purely physical – point of view, the H_2O-only world always turns out to be closer than the XYZ-only world. Will this be true? (Boghossian, 1991, p. 81)

It is important to note here that Boghossian takes himself to have shown that Fodor’s theory is a type I theory where “to say that P is an asymmetric dependence base for S is simply to say that P is the sole cause of S tokens *in the closest world where S has a single cause*” (Boghossian, 1991, p. 81, my italics). Above, we saw reason to believe that Fodor must accept this conclusion. So there are worlds where the only law that govern S tokens is P → S, and these worlds are the ones that are the candidates for being the asymmetric dependence base for S. The one of these that is the closest to the actual world is the one that is the asymmetric dependence base for S. But by what principle should we decide the distances between these worlds? We said in the beginning of this chapter that it is the changes one must make to get to a world that are important in this context.

Fodor has an account of this procedure (quoted in Boghossian, 1991, p. 83) which we will review. The question is: why does “water” track H_2O and not XYZ? What do we need to change to make it the case that “water” tracks H_2O and not XYZ? Kind terms are, as we have seen, governed by our dispositions and what we can distinguish between. Let us consider a world where we can infallibly tell H_2O and XYZ apart. What do we need to change to make this the case? We could perhaps make it so that XYZ has a particular color or smell that makes it instantly recognizable. This world will then be a world in which our “water” tokens track H_2O and not XYZ. Next we will consider what needs to be the case for “water” tokens to track XYZ and not H_2O. Plausibly we need to change the world enough that the different substances are recognizably different, e.g. by introducing some distinguishing feature, like in the previous world. And, in addition to this, we need to change our dispositions, i.e. we need to change such that we are disposed to use “water” for XYZ and not H_2O. The important point is that we need to change more to get “water” to track XYZ (i.e. make XYZ and H_2O distinguishable, and change our disposition to token “water” of XYZ) than to make it track H_2O (to make H_2O and XYZ distinguishable). Since we need to change more to get to an XYZ world than to an H_2O world, the H_2O world is closer and “water” has H_2O in its extension. This way of specifying distances between possible worlds is Fodor’s. Boghossian seems to say that this way is question-begging and not naturalistic. Let us try to specify why he thinks so.

We have seen that “water” tokens H_2O and not XYZ because we need to make more changes to get to the world where “water” is tokened only by XYZ (a world where the substances are obviously distinguished and that our dispositions are different) than we need to make to get to the world where “water” is tokened only by H_2O (only that the substances are distinguished). So it seems that the H_2O world is closer on account of the fewer changes needed to get there, and that explains why “water” means H_2O and not XYZ.

Boghossian objects that this way of specifying the distance between worlds is question begging (Boghossian, 1991, p. 82-83). His point is that it is not obvious that particular physical changes and intentional changes are equally big changes i.e. that they count for as much as the other changes. Fodor seems to think of the changes individually as having the same value as any other change. This has the obvious advantage of allowing distances between worlds to be accounted for quantitatively. And this, it seems, is what Fodor does. As we have seen, the XYZ world is farther away from the actual world than the H_2O world because to get to the XYZ world you need to make two changes (make XYZ uniquely detectable, and change our dispositions), whereas to get to the H_2O world you only need to make one (make H_2O uniquely detectable).

Boghossian’s point is that this way of looking at changes is not very plausible. He makes the observation that to make H_2O infallibly detectable from XYZ one only needs to change H_2O in such a way so that it has some property that is instantly recognizable (Boghossian, 1991, p. 82). One does not also need to make XYZ infallibly detectable from anything else. He continues by saying that making these changes depends, from a physical perspective, not only on our detecting capabilities, but also on the chemistry of the substances. From this it is reasonable to infer that some substances will be easier than others to change in such a way as to imprint them with a property that is instantly recognizable by us. This seems to show that physical changes cannot be regarded as equal in such a way as to license a quantitative account of distance between worlds in the way we have seen that Fodor does. Boghossian continues by presenting the following case. He assumes that XYZ has been made uniquely identifiable by an alteration of our sensory apparatus that is such that XYZ smells horrible. H_2O is not identifiable in the same manner, and can be confused with other substances that have similar traits. It is also assumed that the change that made XYZ instantly recognizable is a very small physical change. The change needed to make H_2O instantly recognizable is assumed to be a complicated affair, implying bigger changes needed than the XYZ case. Boghossian poses the following question:

> To get, then, from our world – in which “water” means H_2O – to a world in which “water” gets applied only to H_2O, you have to make a big physical change; to get to a world in which it gets applied only to XYZ, you have to make a physical change and a small intentional change. Now: which world is closer to ours? (Boghossian, 1991, p. 82-83)

He continues by concluding that the only way of getting the desired result – that the H_2O world comes out as closer than the XYZ world – one needs to assume that “all physical changes are on par, and every intentional change counts for as much as every physical change.” (Boghossian, 1991, p. 83). This is non-naturalistic in the following sense: Fodor’s argument depends on the physical changes being on par because then the intentional change can settle the question of which world is closer. But, as Boghossian’s argument show, there is reason to believe that physical changes are not on par, so that situations where Fodor’s theory yields the wrong results are possible.

Fodor responds to this argument by stressing that the laws in these cases are ceteris paribus laws and that this changes the case. He says:

> What's needed is that each world in which H_2O and XYZ is distinguishable and we apply 'water' to H_2O but not to XYZ is closer to us than the world *that is closest to it* in which H_2O and XYZ are distinguishable and we apply 'water' to XYZ but not to H_2O. (This is a way of saying that *ceteris paribus* worlds in which water rules are closer to us than corresponding worlds in which XYZ does.) Any adequate distance measure should have this property because, as we've seen, you have to change more things to get to XYZ-but-not-H_2O) [*sic*] worlds than you do to get to H_2O-but-not-XYZ worlds. And though, as Boghossian very properly reminds us, it is not required that the more things you have to change to get to a world, the further away that world has to be, it *is* required that the more things you have to change to get to a world, the further away that world has to be *ceteris paribus*. (Fodor, 1991, p. 276-277)

Fodor seems to be saying that the reason Boghossian's case yields the wrong result is that he does not take in to account that the laws in question are ceteris paribus laws. The feature of ceteris paribus laws that make it the case that Boghossian's example does not apply is not easy to understand, and I am a little puzzled by this. I do not quite understand what good it does to make the distance metric to be a relation between A – B and B – C instead of A – B and A – C. Actually, I think there is an ambiguity in how the distance metric is articulated in the quote above. This is something I think can have serious implications for Fodor's theory, but we will only consider it superficially here. Let's suppose the actual world is A, the H_2O world is B and the XYZ world is C. Now, as we remember, the question we want an answer to is this: why does "water" mean H_2O and not XYZ?

I claim that there are two ways of understanding Fodor's account in the quote above: The A – B distance is less than the B – C distance. We are imagining that C is the closest world where XYZ causes "water" tokens in relation to B. The ambiguity arises from the fact that Fodor doesn't say if the distance that matters is in relation to A or not. And there is no reason to assume, since we are dealing with ceteris paribus laws that can cancel each other out, that the worlds we are considering are arranged linearly. It is in the cases where they are not that the ambiguity becomes relevant. If the A – B distance is less than the B – C distance and it is these two distances that matter, then the case comes out right. This, however, implies the assumption that possible worlds are arranged linearly, an assumption that is in need of an argument. If the distances that matter are in relation to A and the worlds are arranged linearly then C is further away from A then B as a trivial fact, and there seems to be no reason to express it like Fodor does. But if the worlds are not arranged linearly, as is possible since ceteris paribus laws can cancel each other out, and we resolve the ambiguity by postulating that it is in fact the relation to A that matters, then the possibility arises that even though B – C is greater that A – B, A – C can still be less of a distance than A – B. This has the plausible consequence that the XYZ world can still be closer than the H_2O world. This is, in my opinion, a result Fodor cannot tolerate. If it is true that this result only can be avoided by assuming that possible worlds are arranged linearly I think it is a result that comes at a price for Fodor. Intuitively I do not think it is plausible that possible worlds are arranges in a strict, linear way. This assumption will also imply that Fodor must provide a naturalistic and non-question begging account of the arrangement of possible worlds that has the result that they are arranged linearly. I do not think such an account is forthcoming.

6.9. Conclusion

In this chapter we have primarily focused on Boghossian's argument against Fodor. The argument aims to show that Fodor's theory of asymmetric dependence is a type I theory in disguise. We have seen reasons in support of this claim. We also reviewed Fodor's reply to this objection and found it lacking in that it seems to address something else than what we have taken Boghossian's argument to be. From this I think it is reasonable to conclude that Fodor's reply to Boghossian is not successful in dispelling the worry the argument poses, and that the argument poses a considerable challenge to Fodor's account. Boghossian's argument poses several worries about Fodor's theory's ability to account for what we have said are the three main intuitions we have about a naturalistic theory of meaning: the intuition that the theory should be naturalistic, that it should construe meaning as robust, and that it should not construe meaning as being ubiquitous, i.e. everywhere. Mainly, Boghossian's objection challenges Fodor's theory's account of the robustness of meaning in that it claims to show that the theory will always ascribe disjunctive extensions to symbols. It also challenges the theory's virtue of being naturalistic in that it claims that Fodor cannot get the right content ascriptions unless he makes assumptions about how to arrange the possible worlds, assumptions that Boghossian claims are question-begging and non-naturalistic. We have also seen some evidence for this claim. I think that what we have considered in this chapter suggests that Fodor's theory has severe problems to tackle before it can claim to have succeeded in naturalizing intentionality and meaning.

Summary and Conclusions

At the outset we identified three things we pre-theoretically and intuitively think a theory of meaning should account for. Two of these intuitions are explicit features of meaning: that meaning is not everywhere (the disproval of pan-semanticism) and that meaning is robust (that the meanings of things are insensitive to, for example, what causes them). The third intuition is a more general intuition about how the world is. It is the intuition that everything that exists, in some form or other, is physical. This is the intuition which is the basis of what we have been calling physicalism. I called this a more general intuition because it applies to not only meaning and intentionality but to all things. Physicalism implies that also the rest of the mental, including consciousness, in some form or other, is physical. And everything that is physical in this sense should, in principle, be accounted for naturalistically. Fodor's theory, as we have seen, does not attempt to account for the rest of the mental or consciousness, only for meaning and intentionality.

On this background we can see that Fodor's project of trying to naturalize meaning and intentionality by providing a naturalistic theory of content is an extremely ambitious project, but a project that must be attempted if one is convinced that everything that exists is physical, in some form or other.

It is perhaps the requirement that the theory should be naturalistic that Fodor's theory satisfies best of the three requirements, or intuitions. In part I we saw that, by using a notion such as information, the Fodor's theory is specified in naturalistic terms. As we saw in the discussion of the historical/teleological theory Fodor's theory has the advantage of not having to assume something like the teleological notion of biological function. We saw that Fodor's reasons for finding an alternative solution to the disjunction problem that does not involve a principled distinction between type I and type II situations are sound. His rejection of this distinction is, as we saw, motivated by considerations about robustness.

In part II we have seen Fodor's own proposal for solving the disjunction problem. His asymmetric dependence theory does seem to account for many cases of the disjunction problem, but as we saw when we considered Boghossian's objection there are reasons to believe that it cannot account for all. I do think that some of these considerations pose severe difficulties for Fodor.

In conclusion I would like to observe that though we have concluded here that Fodor's theory does not account for the three main intuitions in a satisfactory manner, we have not shown that Fodor's theory, suitably revised, cannot meet the requirements we have posed. In fact I do believe that Fodor's theory is the best theory available that aims to naturalize meaning. It succeeds in accounting for many things in this context, but not everything it sets out to do. As we observed above, we must assess Fodor's theory in relation to the ambition of the project. In this case the project is to place the mind, and everything else, where it belongs, namely in nature.

Bibliography

Boghossian, P. A. 1991, “Naturalizing Content”, in B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and his Critics,* Blackwell, Oxford UK & Cambridge MA.

Dretske, F. I. 1981, *Knowledge and the Flow of Information,* Blackwell, Oxford.

Fodor, J. A. 1990, *A Theory of Content and Other Essays,* MIT Press, Cambridge MA.

------. 1991, “Reply to Boghossian”, in B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and his Critics,* Blackwell, Oxford UK & Cambridge MA.

------. 1994, *The Elm and the Expert: Mentalese and Its Semantics,* MIT Press, Cambridge MA.

Loewer, B. and Rey, G. 1991, “Editors’ Introduction”, in B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and his Critics,* Blackwell, Oxford UK & Cambridge MA.

Loewer, B. 1997, “A Guide to Naturalizing Semantics”, in B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*, Blackwell, Oxford UK & Malden MA.