THE UNIVERSITY OF CALGARY

How to Make Mistakes: The Problem of Error in Information-Based Semantics

by

Hugh A. R. Leighton

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF

MASTER OF ARTS

DEPARTMENT OF PHILOSOPHY
CALGARY, ALBERTA

AUGUST, 1989

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

ISBN 0-315-54266-7

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "How to Make Mistakes: The Problem of Error in Information-Based Semantics" submitted by Hugh A. R. Leighton in partial fulfillment of the requirements for the degree of Master of Arts.

A. A. Kazmi

Supervisor, A. A. Kazmi
Department of Philosophy

C. B. Martin

C. B. Martin
Department of Philosophy

A. W. Colijn

A. W. Colijn
Department of Computer Science

30 AUGUST, 1989

Abstract

Interest in semantics or theories of content based on information theory has increased with the development of the interdisciplinary field of cognitive science. Accounting for error, mistakes and misrepresentation presents a major problem for information-based semantics. This thesis critically examines two leading information-based theories of content which have been proposed by Fred I. Dretske and Jerry A. Fodor, focussing on how they approach the Problem of Error. It is concluded that neither has been successful in providing a model which gives an adequate explanation of content and solves the Problem of Error.

Acknowledgements

I would like to thank my supervisor Dr. Ali Akhtar Kazmi for his advice and philosophical guidance and for his patience while I tried to do too many other things at the same time. I would also like to thank Dr. Charlie Martin for helpful comments regarding a draft of this thesis and his encouragement and many discussions over the course of my program. I think I finally see some of the edges in the bigger picture.

Dedication

For my partner, Cheryl Peters, my parents, Art and Evelyn Leighton, and the memory of my friend Doug Reierson.

Table of Contents

Chapter 1. Introduction

1.1 Information-Based Semantics

The study of semantics is the study of meaning. For instance, while the syntax of a language describes how components such as phrase units and words combine to form grammatical sentences, the semantics of the language focusses on what those linguistic units mean, refer to, or express; it deals with the content of the linguistic units. Semantics does not, however, simply provide a list of words and their associated meanings, as might be found in a dictionary, since this would be of limited value to understanding how meaning comes to be connected with words or the nature of the connection between a specific word and its particular meaning. Rather, semantics aims to provide a general model of the relationship between symbols and their content, in a way that is analogous to how syntactic theory models the form of a language using rules to describe the way in which symbols combine to produce grammatical sentences rather than just listing grammatical sentences. Where syntax is concerned with form, semantics addresses content.

There is a great variety in the theories of content which have been proposed at different points of the philosophical history of semantics. One goal which is currently pursued in the context of modern philosophy of mind and cognitive science is to provide a naturalized theory of content. Cognitive science aims to bring together research, theories, practical and conceptual constraints, and models from such disciplines as psychology, neurophysiology, artificial intelligence, philosophy of mind, and linguistics. The goal of this enterprise is to synthesize a general model and understanding of cognition and other mental phenomena building on the contributions, successes and failures of these various fields of study, figuratively taking the blindfolds off people attempting to model an elephant by feeling just a leg or ear or trunk. By pooling and using what has been learned in other fields, it is hoped that the models generated from this wider perspective will become more accurate. Cognitive science aims to unify somewhat disparate fields, constructing models of the mental built on a foundation of natural entities and processes. Such a naturalistic account of cognition would provide an explanation of how mental phenomena such as thinking, perceiving, remembering, intending or meaning are related to the material world which physics, chemistry and biology describe.

A naturalized theory of content would provide cognitive science with an account of how one piece of the material world can can be about or represent another. Such a notion of content would be general enough to apply not just to words and linguistic units, but also to a wider range of symbols or content bearing entities. This generalized idea of a symbol could include such things as natural signs, like smoke representing fire or tree rings expressing the age of the tree. More importantly for cognitive science, such a theory of content would apply to neurophysiological entities, such as neuronal or other brain state groupings allegedly corresponding to thoughts, beliefs and other mental phenomena which are about the outside world. A naturalized theory of content would provide a model for reconciling, on one hand, a description of some state of affairs such as a belief that it is raining given in terms of such semantic or intentional concepts as belief and meaning, with a description of the same state of affairs in terms of relationships between portions of the physical world which does not make use of any semantic or intentional notions.

Central to much of cognitive science is what has been called the computational view of the mind.[1] This view claims that cognition (and, in general, the activity of any intentional or representing system) can be described and understood in terms of computations (formal rule-governed operations or transformations) on representations (states of the system) which are directed, in part, by the content of the representations (what those states represent). On this view, the behaviour of such a system is determined, not only by syntactic rules involving the formal nature of the system states, but also by semantic considerations which relate the contents of the states. These models require a naturalised theory of content which will provide an account of how and what these states represent; without such a theory, the computational view of the mind cannot get off the ground. Cognitive science has looked to information theory as a potential basis for a naturalized theory of content which can ground the computational view of the mind.

---

1 See [Pylyshyn, '84].

Around the mid-point of this century Claude Shannon developed a mathematical theory of information[2] which quantified in an objective way the information associated with a state of affairs. Shannon proposed a technical notion of information which holds that the occurrence of any event or state of affairs generates some information. The amount of information associated with the occurrence of an event or state is a function of the number of alternative events or states of affairs which could have obtained. On this theory, communication is understood as the receipt of information from some source, where the state of the source is seen as generating an information-bearing signal which is transmitted across a communication channel to the receiver. Shannon's quantitative analysis of such aspects of a communication system as the capacity of the channel to handle varying amounts of information, the average information generated and received, the effects of noise, equivocation and coding schemes on the transmission of a signal provide the basis for much of modern telecommunications, and portions of electrical engineering and computer science.

Information theory as described by Shannon is not by itself sufficient to provide a theory of content. Shannon's work addressed only quantitative issues surrounding information such as measuring the maximum amount of information a channel is capable of

---

2 See [Shannon, '49].

carrying or the amount of information received from a source, so classical information theory needs to be extended to include consideration of not only how much information but also what information is carried by a signal. An information-based theory of content will have to provide a qualitative account of the technical notion of information.

Fred Dretske[3] and Jerry Fodor[4] have produced two of the leading models in information-based semantics. Dretske builds quite substantially on Shannon's foundation developing a somewhat direct extension that work. Beginning with a Shannon's account of the amount of information carried by a signal, Dretske goes on to formulate a notion of the informational content of a signal (a symbol, in the sense noted above) and builds a general theory of content from this information-theoretic base. The key to informational content of a signal, on Dretske's model, is the dependence of a signal on the state of affairs at the source which generated it. If a signal can only occur when some particular state of affairs obtains at the source, the presence of a signal of that sort tells the receiver which state the source is in; that signal carries information about the source, namely, that it is in that particular state. The idea that a signal carries information about what must be

---

3 See [Dretske '81].

4 See [Fodor '87].

the case when it is present forms the core of Dretske's model of informational content.

Fodor, writing several years later than Dretske, with the benefit of discussion and debate regarding the strengths and weaknesses of Dretske's approach takes a slightly different orientation to providing a theory of content. Fodor bases his theory on a causal analysis of the relationship between a symbol and its content. At the heart of his model is the idea that symbols are about or represent what causes them. The content of a symbol such as "horse", for example, can be traced to horses which have caused that symbol to be tokened. This model can be viewed as a variety of information-based semantics since a symbol can be said to "carry information" about what causes it.

1.2 The Problem of Error

There are many problems which arise for a theory of content, including how to deal with intentionality, knowledge, and shared content. The so-called Problem of Error involves accounting for error, mistakes and misrepresentation. It is not unusual for symbols to be used or applied erroneously. Consider, for example, a token of the symbol "horse" used to refer to a four-legged animal standing in a field far away, which, as it turns out is not a horse but a cow that has been mis-identified as a horse. Another example of erroneous content would be a mistaken belief that today is Saturday. A theory of content must be able to account for such errors and misrepresentations as well as correct applications if it is going to provide an adequate account of symbolic content. If in cases of error or mistake, a theory of content fails to provide a correct assignment of meaning to a symbol or a belief state, it cannot be deemed a successful model of content adequate to account for the range of applications required.

The Problem of Error is particularly difficult for information-based theories of content. For instance, on Dretske's model, the informational content of a symbol or a signal is the state of affairs which must be the case at the signal's source. It is possible to talk of more or less content of this sort, but informational content so defined can never be false; if the symbol (brain state) corresponding to the belief that today is Saturday can be generated by mistake, or in error, then there is no single state of affairs which must obtain when that symbol is present, thus no informational content corresponding to the (false) belief that it is Saturday. On the other hand, if, as Fodor suggests, a symbol receives its content from what causes it, another version of the Problem of Error arises. It is not clear how a symbol such as "horse", which can be caused both by horses and cows seen at a great distance under poor lighting conditions can avoid being assigned a content which includes not only reference to horses, but also cows and anything else which might be mistaken for a horse.

Attempts by information-based theories of content to deal with the Problem of Error are further constrained to avoid any direct or indirect appeal to semantic or intentional notions such as "what the symbol really means". A theory of content could only use such notions at the risk of circularity and, in any event, would not meet the criteria of naturalized theory. A naturalized semantics needs a theory which can provide an assignment of content to symbols based on only naturalistic considerations, an assignment which coincides with independent intuitions about the content of symbols in cases which include mistaken or erroneous symbol tokens.

1.3 The Objective of this Thesis

This thesis critically examines the theories of content constructed by Dretske and Fodor, focussing on how they approach the Problem of Error. It concludes that neither of them have been successful in providing a model which gives a naturalistic explanation or reduction of symbolic content and solves the Problem of Error. Since the particular details of the Problem of Error and the strategy for handling it depend greatly on the context of the theory of content to which it applies, a fairly detailed presentation of the general framework of these theories is required.

In Chapter 2 Dretske's theory is discussed, outlining the development of his conception of semantic content from a foundation of information-theoretic notions. How the Problem of Error arises for Dretske's model and the solution he proposes is examined and it is concluded that his approach to dealing with false content is not consistent with other critical aspects of his model. In the final section of Chapter 2 the relationship between some aspects of Dretske's model and Shannon's information-theoretic foundations from which the model was derived is considered.

Fodor's theory of content is presented in Chapter 3. Fodor's view of content must be considered within the context of his Representational Theory of Mind, which is summarized prior to discussing his causal theory of content. How that theory of content is vulnerable to the Problem of Error, and a related, more general problem, is explored, as is Fodor's suggested solution. Chapter 3 concludes with a criticism of the adequacy of Fodor's theory of content and the success of his strategy for handling the Problem of Error.

Chapter 2. Dretske

This Chapter outlines Fred Dretske's account of content as presented in [Dretske '81]. The Problem of Error, how it applies to Dretske's model, and his proposed solution are examined. The final section of this chapter examines some aspects of the relationship between Dretske's model and the information theoretic foundations from which it was derived.

2.1. Dretske's Information-Based Account of Content

Dretske's model is built upon a framework which he traces back to Claude Shannon, whose work has had a major influence on such fields as electrical engineering, computer science and telecommunication theory, to name just a few. In the middle of this century, Shannon developed a mathematical theory of communication[1] which has come to be known as communication or information theory. Information theory

> provides a measure for how much information is to be associated with a given state of affairs and, in turn, a measure for how much of this information is transmitted to, and thus available at, other points. The theory is purely quantitative. It deals with *amounts of information* - not, except indirectly and by implication, with the information that comes in those amounts.[Dretske '81, p.3]

---

1 See [Shannon '49].

Classical information theory, as described by Shannon, is not a theory about the nature or content of information. Instead, it provides a set of tools for quantitatively describing the generation and transmission of information. A typical model of a communication system includes, in Shannon's terms, an information source which generates or produces signals that are encoded and transmitted across a communication channel, which may introduce an element of noise corrupting the signal, and a receiver which decodes the signal when it arrives at its destination. A typical sort of problem to which classical information theory might be applied might involve determining the maximum amount of information which can be sent through a communication system with specified characteristics, such as channel capacity, noise levels, and signal encoding schemes.

Dretske's goal is to build on this quantitative base and extend it to a semantic theory of information which accounts for the semantic and intentional features of language and propositional attitudes. The first step in this project is to supplement the basic communication theory with a theory of the qualitative aspects of information; a theory of informational content. Having established a notion of informational content, Dretske proposes a model to show how an information bearing structure (something which has informational content) can possess a content with intentional properties appropriate to a semantic content and can be related to belief (as a paradigm example of a propositional attitude). It is useful to have a rough sketch of how the parts of the model are supposed to eventually come together. Informational content which meets certain criteria is defined to be semantic content. A notion of semantic type which generalizes from tokens of semantic content is used to establish the content or meaning of a belief. A belief, in turn, is defined as a cognitive structure which exhibits both a functional aspect, given in terms of its causal role within a larger system in which it is imbedded, and a representational aspect, exhibited by its content. Dretske's model is best described by first presenting his theory of informational content, then showing how semantic content is related to informational content and, finally, linking these ideas to beliefs.

2.1.1. Informational Content

While it is not necessary to go into detail explaining the mathematical basis of Shannon's theory, there are a few ideas which Dretske uses that should be explicitly identified. If, in a given situation, there are a number of events or states of affairs possible, the actual occurrence of one of those possibilities is said to generate information. Classical quantitative communication theory says that the amount of information generated by or associated with any particular event or state of affairs is a function of the number of alternatives that were available or could have happened. To put this another way, the amount of information associated with an event is a function of the amount of uncertainty that was eliminated by the occurrence that event. Consider, for example, the rolling of a single six-sided die. When the die stops rolling, six possible states of affairs (one for each face which may have ended up showing on top) have been "reduced" to one actual state, for instance, the one which has the face with two dots showing. The die may be seen as a signal source which generates information when it settles on one side. The amount of information generated by a roll of a six-sided die is less than that associated with the roll of a ten-sided die, since in the latter case, more alternative states of affairs were available, more uncertainty was eliminated.

A theory of informational content must provide an account of what it is for a signal generated by a state of affairs to have a particular content, to carry a particular piece of information. To continue the example, the theory must say what it is for a signal to have the informational content that a source *s*, the die, is in some particular state *F*, showing a face with two dots. For this example, the light which transmits the dot pattern on the showing face of the die to the receiver can be considered the signal.

Dretske gives the following definition of the informational content of a signal:

> A signal *r* carries the information that *s* is *F* = The conditional probability of *s* 's being *F*, given *r* (and *k*), is 1 (but, given *k* alone, less than 1) [Dretske '81, p.65]

This definition may be formalized as follows:

(1) $I(r, \text{that } s \text{ is } F)$ iff
$$[P(\text{that } s \text{ is } F / (r \text{ and } k)) = 1 \text{ and } P(\text{that } s \text{ is } F / k) < 1]$$

where the relation of carrying information is symbolized as I (x, y) where x is the signal which is carrying the information and y is the informational content being carried and the conditional probability of A given B is symbolized as P(A/B).

The idea on which Dretske's definition is built is that a signal carries the information that the die is showing a face with two dots when the conditional probability that the die is showing a face with two dots, given the presence of the signal, is 1. In other words, if the presence of the signal is a perfect indicator of the presence of a particular state of affairs, that signal carries the information, has the informational content, that that state of affairs obtains. If the signal only occurs when the state of affairs obtains, then the conditional probability of that state of affairs obtaining given that that signal is present is 1. If, however, a given signal could be generated by more than one state of affairs, the probability of any particular state obtaining given the presence of the signal is less than one, since the signal may have been generated by one of those alternative states of affairs which could generate the signal. In such a case, the signal does not carry the information that any particular generating state obtains, although it may carry a disjunctive piece of information that some generating state or other obtains.[2]

---

2 It should be noted that Dretske's model tacitly assumes that the sort of probabilistic analysis which works for a designed or engineered communication system with a finite number of well defined possible states can be straightforwardly extended to the real world with a multitude of possibilities, which, it may be argued, are uncountable and not amenable to such an arithmetic analysis. This assumption will not be challenged here.

Dretske's definition includes references to '*k* ', which "is meant to stand for what the receiver already knows (if anything) about the possibilities at the source."[Dretske '81, p.65] The parenthetical portions of the informational content definition are intended to deal with cases where a receiver's prior knowledge about the possible states of affairs which may obtain at the source affects the information transaction. Consider a shell game, involving three shells A, B, and C, under only one of which is a pea. Person X lifts shell A, finds no pea and replaces the shell. Along comes person Y, who did not see what X saw, lifts shell B and finds no pea. Call the signal that shows no pea under shell B, signal r. Since X knows that the pea had to be under one of the shells and also knows that it is not under shell A, X receives from signal r the information that the pea must be under shell C. The probability of the pea being under shell C, given signal r and what X already knows about the possibilities, is 1. The probability of the pea being under shell C, given just what X knew (a pea is under one of the shells) prior to the receipt of signal r, was less than one, since for all X knew, the pea could have been under either shell B or C. Y, however, does not receive the same information from signal r. The probability of the pea being under shell C, given signal r and what Y knew about the possibilities (a pea is under one of the three shells), is less than one, since for all Y knows after seeing no pea under shell B, the pea could be under either shell A or C.

There are some formal problems with Dretske's definition which are immediately apparent from the formalization given in formula

(1). In particular there is a problem of scope with respect to the variable k since it appears on the right-hand side of the biconditional but does not appear on the left-hand side. Unbound variables are conventionally interpreted as universally quantified, but there are at least two ways of interpreting Dretske's definition and formula (1) with respect to the scope of k 's quantification:

(2) $(\forall k)$ {I (r, that s is F) iff
[P(that s is F /(r and k))=1 and P(that s is F /k)<1)]}

(3) I (r, that s is F) iff
($\forall k$) [P(that s is F /(r and k))=1 and
P(that s is F /k)<1)]

Consider, first, the difficulties with formula (2) as illustrated for some signal r and content, say, that snow is white. This yields formula (2'):

(2') ($\forall k$) {I (r, that snow is white) iff
[P(that snow is white/(r and k))=1 and
P(that snow is white/k)<1)]}

Now consider an instance of (2') where k is instantiated such that the probability conditions specified in the right-hand side of the biconditional are satisfied, by k_1. That is, the probability that snow is white, given r and k_1 equals 1 and the probability that snow is white given k_1 alone is less than 1. See (2'a):

(2'a) I (r, that snow is white) iff
[P(that snow is white/(r and k_1))=1 and
P(that snow is white/k_1)<1)]}

Since the right-hand side of the biconditional in (2'a) is true it follows from the truth of (2') and hence (2'a) that the left-hand side of the bi-conditional is true; I (r, that snow is white) is true. Signal r carries the information that snow is white.

What about an instance of (2') where *k* is instantiated by some k_2, such that the probability conditions specified in the right-hand side of the bi-conditional are not satisfied? For instance, if k_2 ("what the receiver already knows ... about the possibilities at the source") already includes the information that snow is white, the probability that snow is white given k_2 will equal 1, making the second conjunct false and, in turn, the right-hand side of the biconditional false. See (2'b):

(2'b) I (r, that snow is white) iff
[P(that snow is white/(r and k_2))=1 and
P(that snow is white/k_2)<1)]}

Since the right-hand side of (2'b)s biconditional is false, it follows from the truth of (2') and hence (2'b) that the left-hand side must also be false; I (r, that snow is white) is false. Signal r does not carry the information that snow is white. But this contradicts the earlier result arrived at from (2'a). Signal r cannot both carry and not carry the information that snow is white. Since instances of formula (2) lead to a contradiction, it cannot provide an appropriate characterization of informational content. *K* 's scope cannot extend over the entire definition as formalized in formula (1).

The interpretation of Dretske's definition of informational content characterized in formula (3) is not so seriously flawed as (2). According to (3), I (*r*, that *s* is *F*) is true just in case [P(that *s* is *F* / (*r* and *k*))=1 and P(that *s* is *F* /*k*)<1)] is true for <u>all</u> instantiations of *k* ; that is, whatever the receiver might know about the possibilities at the source. If there is any instantiation of *k* which could make [P(that *s* is *F* /(*r* and *k*))=1 and P(that *s* is *F* /*k*)<1)] false, then I (r, that *s* is *F*) is false. The signal *r* only carries the information that *s* is *F* if all knowledge states satisfy the probability conditions specified by the open sentence in right-hand side of the biconditional. The existence of a single knowledge state *k* which makes the right-hand side of the biconditional false, for instance someone who already knows that *s* is *F*, means that the signal *r* does not carry the information that *s* is *F*.

This may not be an unreasonable way to define the informational content of a signal. It does not, however, appear to be what Dretske had in mind. In the paragraph immediately following the definition, he writes,

> for example, if one already knows that *s* is either red or blue, a signal that eliminates *s* 's being blue as a possibility (reducing this probability to 0) carries the information that *s* is red (since it increases this probability to 1). For someone who does not know that *s* is either red or blue (given what they know, it could be green), the *same signal* might *not* carry the information that *s* was red. [Dretske '81, p. 65]

The scope of the quantification of k in formula (3) is not compatible with the analysis Dretske provides in this example. Formula (3') describes the situation for the signal r that eliminates s 's being blue as a possibility, and for the content that s is red and (3'a) shows the right-hand side of (3')s biconditional with k instantiated by the knowledge state k of the individual who does not know that s is either red or blue:

(3') I (r, that s is red) iff
($\forall k$) [P(that s is red /(r and k))=1 and
P(that s is red/k)<1)]

(3'a) [P(that s is red /(r and k))=1 and P(that s is red/k)<1)]

The individual concerned does not learn from r that s must be red ("given what they know it could be green") so the probability that s is red given r and k is less than 1. (3'a) is, therefore, false, and hence the universal generalization of (3'a) which is the right-hand side of (3')s biconditional is false. Since the right-hand side of the biconditional in formula (3') is false, it must also be false that signal r carries the information that s is red regardless of what any receiver might know about the possibilities at s . This is contrary, however, to Dretske's initial statement in the passage just quoted. Formula (3) cannot be what he has in mind.

Can any alternative interpretation or formalization of Dretske's definition of informational content be molded or patched to coincide with his interpretation of this example? Only if the left-hand side of the bi-conditional is modified to include in the information-carrying relation some reference to k.[3] One way that this modification could be achieved, is to re-define the information-carrying relation to be information-carrying for k. If signal r carries the information for k that s is F, this would be symbolized as $I_k(r, \text{that } s \text{ is } F)$. Dretske's definition, so amended, would then be formalized as,

$$(4) \quad (\forall k)\ \{I_k(r, \text{that } s \text{ is } F) \text{ iff } [P(\text{that } s \text{ is } F\ /(r \text{ and } k))=1 \text{ and } P(\text{that } s \text{ is } F\ /k)<1)]\}$$

Formula (4) avoids the problems identified above with formula (2), since if, for some k_1 which satisfies the right-hand side of the biconditional, the relation which holds between the signal and the content is $I_{k_1}(r, \text{that } s \text{ is } F)$, whereas for some k_2 which does not satisfy the right-hand side of the biconditional for the same signal and content, the relation which does not hold is $I_{k_2}(r, \text{that } s \text{ is } F)$. There is no contradiction if $I_{k_1}(r, \text{that } s \text{ is } F)$ is true and $I_{k_2}(r, \text{that } s \text{ is } F)$ is false. The problems of formula (3) are avoided since the information carried by a signal is explicitly relativized to the receiver's knowledge. A signal r may carry the information that s is red, for someone with the appropriate background knowledge k_1, while, for someone with a different background knowledge k_2, the same signal r may not carry the information that s is red.

---

[3] This would eliminate any version of formula (3) as a potential formalization of the informational content definition, since the entire definition would have to be within k's scope.

Although such a relativization of informational content is not indicated in his definition of informational content, quoted above, it does indeed seem to be what Dretske had in mind. He suggests[4] that the informational content of a signal is relativized to what the receiver knows about the possibilities at the source, to k. The typical situation, however, according to Dretske, is one in which "the assessment of the information contained in a signal ... is carried out against a background of communally shared knowledge in which individual differences are submerged."[Dretske '81, p.80] He claims that usually everyone knows the same thing about the possibilities at the source, has the same k, so the details of relativization drop out of the analysis. While this claim is certainly not self-evident and may be disputed, this is not the place for a detailed critique about the validity of his idealization. It is enough to note that the amended formulation of the definition of informational content as set out in formula (4) is in line with what Dretske says regarding the relativization of the informational content of a signal to k.

2.1.2. Nested Information

In building an account of semantic content Dretske supplements his definition of the informational content of a signal with a notion of nested information. The idea behind nested information is that,

---

4 See [Dretske '81, pp.79-81].

> if there is a natural law to the effect that whenever *s* is *F*, *t* is *G* (thus making the conditional probability of *t* 's being *G* , given *s* 's being *F*, 1), then no signal can bear the message that *s* is *F* without also conveying the information that *t* is *G*. [Dretske '81, p.71]

Dretske includes as natural laws such nomological or causal laws as might relate temperature and the expansion of mercury, as well as analytic or definitional relations such as those which require that if *s* is a square then *s* is also a rectangle. Thus if a signal carries the information that *s* is a square then it also carries the information that *s* is a rectangle. Similarly for a message which contains the information that the mercury has expanded, it also carries the information that the temperature of the mercury has risen. The two pieces of information are related by the nesting relation:

> The information that *t* is *G* is nested in *s* 's being *F* = *s* 's being *F* carries the information that *t* is *G*. [Dretske '81, p.71]

Dretske comments that a signal which carries the information that *s* is a square "will also carry the information that *s* is a quadrilateral, a parallelogram, *not* a circle, *not* a pentagon, a square *or* a circle, and so on." [Dretske '81, p.70]

A problem arises when this nesting relation, as it is presented, is combined with the definition of informational content of a signal. Take Dretske's example of some signal r which carries the information (for someone with background knowledge k) that s is a square; that is, I_k(r, that s is a square) is true. According to the definition of informational content, it must be the case that P(that s is a square/(r and k))=1 and P(that s is a square/k)<1). Since it is an analytical law that anything that is a square is also a rectangle, information that s is a rectangle is nested in the information that s is a square. Given what Dretske says about nested information, r must also carry that nested information (for k); I_k(r, that s is a rectangle) is true. So, according to the definition of informational content, it must be true that P(that s is a rectangle/(r and k))=1 and P(that s is a rectangle/k)<1). But, suppose, further, that prior to receiving the signal r, the individual already knows that s is a rectangle, without knowing that s is a square; k includes the information that s is a rectangle. In this case, P(that s is a rectangle/k) equals 1. But if this is true, according to the definition of informational content I_k(r, that s is a rectangle) is false, r does not carry the information that s is a rectangle, contradicting the result arrived at by applying the definition of the nesting relation.

This situation may be clarified if the nesting relation is broken into two claims. First is the claim that if there is an analytic or nomic law that says that if *s* is *F* then *t* is *G*, then the information that *t* is *G* is nested in the information that *s* is *F*. The second claim is that a signal which carries the information that *s* is *F* carries all information nested in the information that *s* is *F*. The first claim is, perhaps, not too controversial, being primarily a matter of definitional labelling. The problem presented above is a result of the second claim in combination with the second conjunct of the right-hand side of the biconditional of the definition of informational content (formula (4)). It is always possible (and often likely) that someone's background knowledge about the possibilities at the source includes some information that t is G, which is analytically or nomically nested in the information that s is F, without also including the information that s is F. But then the information that t is G cannot be contained in the signal r for that k, since P(that t is G /k) will equal 1 in every such instance, contrary to Dretske's definition of informational content. To eliminate the final conjunct of the informational content definition would also eliminate the relativization to k that Dretske desires, as was pointed out above. The only alternative solution to this problem if the informational content definition is to be preserved intact, is to reject the second claim of the nesting relation, that a signal which carries the information that s is F carries all information nested in the information that s is F.

One result of rejecting the second claim would be that not all information nested in information carried by a signal will necessarily be information that is carried by that signal. This will cause problems for some of the later details of Dretske's theory relating to the semantic content of a signal which makes use of nested information to distinguish between methods of encoding information. A process is described which is supposed to be able to extract nested information from a signal. This process will run aground if the nested information actually carried by a signal were relativized to the knowledge of the receiver, omitting some information which is already known by the receiver. Some information nested in information that is carried by the signal would not be available for extraction from the signal, since it is not actually carried by the signal. Something which is not present in a signal cannot be extracted. Even if, as Dretske suggests in an earlier quote, people generally share the same background knowledge, this problem persists. It only means that the gaps in the nested information carried by a signal are also shared generally. This point will be pursued further in section 2.4. It should also be noted that this problem is not a result of the amendment to the definition of informational content to formula (4); the same problems arise even with the original formulation of the definition.

2.1.3. Semantic Content

Recall that the task at hand is to relate informational content to semantic content and, in turn, to meaning and belief, and to show how structures bearing the latter types of content can be developed out of information bearing structures. Dretske looks to the intentionality exhibited by semantic contents and beliefs for a distinguishing feature which will provide a handle for relating them to informational contents. He notes that,

> *all* information-processing systems occupy intentional states of a certain low order. To describe a physical state as carrying information about a source is to describe it as occupying a certain intentional state relative to that source. If structure *S* carries the information that *t* is *F*, it does not necessarily carry the information that *t* is *G* even though nothing is *F* that is not also *G*. The information embodied in a structure defines a

> propositional content with intentional characteristics.[Dretske '81, p.172]

It is not exactly clear what difference Dretske sees, if any, between informational content as defined earlier and propositional content. However, what he has in mind when speaking of the intentional characteristics of an information-bearing structure is clarified by three orders of intentionality which he defines:

> *First Order of Intentionality*
>
> (a) All *F* 's are *G*
> (b) *S* has the content that *t* is *F*
> (c) *S* does not have the content that *t* is *G*
>
> When this triad of statements is consistent, I shall say that *S* (some signal, event, or state) has a content exhibiting the first order of intentionality.[Dretske '81, p.172]

For example, while it may be the case that all people in the room are beardless, if *S* carries the information, has the content, that *t* is in the room, it may still be the case that *S* does not carry the information, does not have the content, that *t* is beardless. Even though some *S'* s might carry both pieces of information (a photograph of *t* in the room showing *t* 's face, for instance), there is no guarantee that information about such contingently coincident properties will be carried together. If the words "*t* is in the room" carry the information that *t* is in the room, presumably, those words alone do not also carry any information about *t* 's facial appearance. Dretske notes that all information-processing systems exhibit this low level of intentionality.

The second order of intentionality is defined similarly to the first with clause (a) changed to read, "It is a natural law that *F* s are *G*". It is possible, to use Dretske's example, to believe (know) that the water is freezing without believing (knowing) that the water is expanding even though there is a natural law that says water expands when it freezes. The third order of intentionality likewise amends clause (a) to read, "It is analytically necessary that *F* s be *G* ." It is possible to believe (know) that $t = 81$ without also believing (knowing) that t = the square root of 6561, even though it is impossible for 81 to not be the square root of 6561.

Dretske notes that belief and knowledge exhibit higher order intentionality, as seen from the examples. While it is not clear if Dretske thinks that this is all that is involved in the intentionality of beliefs and knowledge, he identifies any propositional content exhibiting the third order of intentionality as a semantic content. To qualify as a semantic content in Dretske's model, the criteria of exhibiting the third order of intentionality must be met by an information-bearing structure.

Semantic content cannot be identified with informational content. Dretske says,

> A signal (structure, state, event) does not possess this higher-order intentionality [exhibited by a semantic content] with respect to its informational content. If properties *F* and *G* are nomically related (there is a natural law to the effect that whenever anything has the property *F*, it also has the property *G*), then any

> structure that carries the information that *t* is *F* will also, necessarily, carry the information that *t* is *G*. Indeed, the information that *t* is *G* will be *nested* in the situation described by "*t* is *F* " in such a way that no signal can carry one piece of information without carrying the other.[Dretske '81, pp.174-5]

Likewise, if *F* and *G* are analytically paired. A signal cannot carry the information that *t* is a bachelor without also carrying the nested information that *t* is an unmarried adult male.[5] In contrast with a belief, which exhibits higher order intentional characteristics because it has a specific content, a structure cannot carry the information that *t* is *F* without also carrying all of the information that is nomically and analytically nested in the information that *t* is *F*. No single piece of information counts as *the* informational content of the signal.

> For this reason a physical structure has no *determinate* or *exclusive* informational content. The pieces of information embodied in a physical structure, although they qualify as propositional contents exhibiting (first-order) intentional characteristics, do not qualify as the kind of *semantic content* characteristic of belief. To occupy a belief state a system must somehow discriminate among the various pieces of information embodied in a physical structure and select *one* of these pieces for special treatment – as *the content* of that higher-order intentional state that is to be identified as the belief.[Dretske '81, p.174]

---

[5] The difficulties with the nesting relation discussed in section 2.1.2 will be ignored until section 2.4.

Since Dretske wants to ground his notion of semantic content in terms of informational content, but the informational content (or, more to the point, the multitude of informational contents) of a structure do not endow the structure with the proper intentional characteristics, he needs pick out a single piece of information contained in the structure which can qualify as its semantic content. If a way is found to single out a piece of information as *the* exclusive content of a structure, that specific content will satisfy Dretske's criteria of semantic content, that is, it will exhibit higher order intentionality; if *S* has as its sole content[6] the information that *t* is *F* then, regardless of any nomic, analytic or any other sort of relation (save, perhaps, identity) between *F* s and (any)*G* s, *S* will <u>not</u> also have the information that *t* is *G* as its content.[7]

To identify the content of a structure, from among the many informational contents that may be carried by that structure, which is a candidate for being its semantic content, Dretske utilizes a distinction in the way information is encoded in a signal or a structure.

> [A] signal (structure, event, state) carries the information that *s* is *F* in *digital* form if and only if the signal carries no additional information about *s*, no information that is not already nested in *s* 's being *F*. If the signal *does* carry

---

6 The sole content of this sort. Of course, *S* will still have its whole ensemble of informational contents.

7 Of the relevant sort.

> additional information about s, information that is *not* nested in s 's being F, then I shall say that the signal carries this information in analog form. When a signal carries the information that s is F in analog form, the signal always carries more specific, more determinate, information about s than that it is F. Every signal carries information in both analog and digital form. 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 '81, p.137]

Consider a signal which carries the information and no other more specific information about s than that s is a square. It does not, for instance, carry any information about s 's colour, location or size; only that s is a square. The signal will carry other less specific information that is nested in the information that s is a square, such as that s is a rectangle, that s is a quadrilateral, and that s is a quadrangle. The information that s is a square is said to be encoded or carried by the signal in digital form while the other pieces of information are in analogue form.

With this coding distinction in hand, Dretske identifies the semantic content of a structure as follows:

> Structure S has the fact that t is F as its *semantic content* = S carries the information that t is F in digital form [Dretske '81, p. 177][8]

---

[8] The definition of semantic content needs to be amended slightly to avoid cases where although the information that t is F is encoded in digital

and points out that this definition results in semantic content having the higher-order intentional properties desired. If a structure *S* has as its semantic content that *t* is *F*, then any information which might be (analytically) nested in the information that *t* is *F*, and thus part of the informational content of *S*, is not the semantic content of *S*, because it is not in digital form; only one piece of information, the information that *t* is *F*, is in digital form. "In this respect, *S*'s semantic content is *unique* in a way that its informational content is not. And this uniqueness is generated by the fact that we have identified semantic content with information that is *coded* in a particular way."[Dretske '81, p.178] In this way the conditions of

---

form, it itself is nested in other information that is <u>not</u> about *t*. The result is:

> Structure *S* has the fact that *t* is *F* as its semantic content =
> (a) *S* carries the information that *t* is *F*
> (b) *S* carries no other piece of information, *r* is *G*, which is such that the information that *t* is *F* is nested (nomically or analytically) in *r*'s being *G*. [Dretske '81, p.185]

Information which qualifies as the semantic content of a structure according to the revised definition is said to be in completely digitalized form. "This definition implies that if *S* has the fact that *t* is *F* as its semantic content, the *S* carries the information that *t* is *F* in digital form. But the reverse implication does not hold. A structure can carry the information that *t* is *F* in digital form but not have this fact as its semantic content."[Dretske '81, p.185] Only the single piece of non-nested information carried in *S*, the piece of information in which all other information carried in *S* is nested, is in completely digitalized form and is the semantic content of the structure. This refinement does not impact on the matters discussed below, so the original definition of semantic content will be used.

higher order intentionality are satisfied by his formulation of semantic content.

Information that is carried in analogue form is eligible for becoming a semantic content by changing the way that it is encoded. A process can be described which would convert a piece of information from analogue to digital form, a so-called digitalization process. Such a process would be selectively sensitive to that piece of information which is to be digitalized. For example, consider a situation where shapes are being classified according to the number of sides they have. A signal which carries the information that *s* is a square, also carries the nested information that *s* is a quadrilateral. The latter is encoded in this signal in analogue form and the former is, assuming that it is the most specific information encoded in the signal about *s*, encoded in digital form. A process which takes this signal and converts it into one which carries the information that *s* is a quadrilateral in digital form, ie., it carries no information about *s* which is more specific would digitalize that piece of information. Such a process can be viewed as classifying *s* as a quadrilateral, "filtering" out other more specific information about *s* ; in the course of digitalization, the information that *s* is a square is lost (in the digitalized representation). Alternatively, the digitalization process can be seen as extracting a piece of analogue information and converting it into digital form, making it the semantic content of the resulting signal or structure.

Dretske thus moves from the informational content of a structure to the semantic content of a structure via a particular coding scheme.

The semantic content of a structure must be encoded in digitalized form. To convert any particular piece of information into a semantic content involves a process whereby that piece of information is encoded in a structure in digitalized form. Typically this will involve the loss of some, more specific, information.

2.2. Belief and the Problem of Error

Semantic content as described above cannot alone be sufficient to give an account of belief. The semantic content of a structure is derived from the information carried by the structure. Informational content, as Dretske has defined it, cannot be false. If a structure S is present only when some state of affairs A obtains, then S is a perfect indicator of and carries information about A obtaining.[9] If, however, S is not a perfect indicator of A then it is not the case that S carries false information about A obtaining, S does not carry information about A obtaining at all. There is no room for false or partial information in Dretske's model. But if informational content cannot be false, the semantic content of a structure which is derived from the informational contents of the structure likewise cannot be false. Beliefs, on the other hand, are frequently false. Therefore the content of a belief cannot be identified with the semantic content of a structure which embodies the belief. Dretske needs some mechanism by which the content of a belief can be based on a semantic content (which is defined in terms of informational content), but which has

---

[9] This leaves aside considerations of background knowledge, *k*.

the added feature of the possibility of being false, the possibility that it could misrepresent a state of affairs. He needs to explain how falsity can arise from a background of truth. This is the Problem of Error as it arises for Dretske's information-theoretic account of content and belief.

Dretske wants to capitalize on intuitions which arise in cases of artificial symbols representing and misrepresenting some state of affairs. A blue patch on a map represents a lake because of accepted conventions for what such a symbol means. It misrepresents the state of affairs if the actual state of affairs does not correspond in the relevant ways to that conventional meaning.

> [A] map can misrepresent the geography of an area only insofar as its elements (the variously colored marks) are understood to have *a meaning* independent of their success in carrying information on any *given* occasion.[10] A particular configuration of marks can *say* (mean) that there is a lake in the park without there actually being a lake in the park (without actually carrying this piece of information) because this particular configuration of marks is an instance (token) of a general type of configuration which *does* have this information-carrying function.[Dretske '81, p.192]

The point is that not only do the symbols on a map have conventional meanings but that, "the symbol token inherits its

---

10 Presumably, an appropriate story could be told about the veracity of the cartographer responsible for the map which could justify the claim that the symbols on the map carry information about the actual geography.

meaning from the symbol type of which it is a token; and the symbol type has an information-carrying role independent of the success (if any) of its tokens in carrying this information."[Dretske '81, p.192] It is in virtue of being a token of the type of symbol which means "lake here" (or, more precisely, "lake in the corresponding location relative to other geographical features which are symbolized") that a particular patch of blue on a map has that meaning, whether or not that representation is accurate.

In the case of conventional symbols such as those on a map, it is easy to determine the meaning of a symbol type by simply consulting the symbol key provided which describes the conventions. The presence of the key is what, in the map example, establishes the meaning or content of the symbol type independent of the actual success of any instances of the symbol to actually represent "what they are supposed to." Dretske wants to use a similar notion which has tokens of cognitive structures such as beliefs inheriting their content from the structure type to which they belong; structure types which have their content spelled out in informational terms, but have that content independent of whether its particular tokens carry such information. A false belief could then arise as token of a cognitive structure which, although it has inherited its content from its structure type (its belief-type, as it were) it does not actually carry the corresponding information. To make this work, Dretske needs to specify the cognitive equivalent of the symbol key on the map. How do structure types get matched to information-based meanings independent of the success of their particular tokens in carrying information?

Dretske's solution is to divide the life-span of a cognitive structure into two parts: a learning phase and an application phase. During the learning phase *L*,

> a system is exposed to a variety of signals, some of which contain the information that certain things are *F*, others of which contain the information that other things are not *F*. The system is capable of picking up and coding this information in analog form ... but, at the onset of *L*, is incapable of digitalizing this information. ... [D]uring *L* the system develops a way of digitalizing the information that something is *F* : a certain type of internal state evolves which is selectively sensitive to the information that *s* is *F*. This semantic structure develops during *L* in response to the array of information-bearing signals (assisted, presumably, by some form of training or feedback).[Dretske '81, p.193]

During the learning phase, a process is developed which digitalizes the information that something is *F* so that when the system is presented with a signal which contains the information that *s* is *F* in analogue form it generates or instantiates a token of a structure which has the information that *s* is *F* as its semantic content, a so-called semantic structure. The development of this structure type in response to signals which carry the information that *s* is *F* and the fact that each of the structure tokens during the learning phase have the information that *s* is *F* as their semantic content fixes a meaning for the general structure type; the learning phase links structure types to contents providing the "key" that maps structure types to meanings.

During the application phase, which amounts to all tokenings of the structure subsequent to the learning phase, the structure type has a content or meaning which its tokens inherit whether or not they carry the information that *s* is *F*, in a manner analogous to tokens of lake symbols on maps.

> [O]nce we have meaning, once the subject has articulated a structure that is selectively sensitive to information about the *F* -ness of things, instances of this structure, tokens of this type, can be triggered by signals that *lack* the appropriate piece of information. When this occurs, the subject *believes* that *s* is *F* ... And if, in fact, *s* is not *F*, the subject falsely believes that *s* is *F*. We have a case of misrepresentation – a token of a structure with a false content. We have, in a word, meaning without truth.[Dretske '81, p.195]

Thus, once the learning period is over and a process has been developed which digitalizes the information that *s* is *F* and yields a structure which has as its semantic content that *s* is *F*, any subsequent instances of that structure type will have the semantic content that *s* is *F* in virtue of being an instance of that structure type, regardless of whether that particular instance has the informational content that *s* is *F*.

In this way Dretske describes how, from informational origins, structures with intentional characteristics appropriate to beliefs can be developed through a process of digitalization which isolates a particular semantic content of a structure and how during the development of the digitalization process, a general content can be associated with the resulting structure type, associated in such a way that a subsequent token of that type can inherit the general content, can have that content as its meaning without actually carrying the content as information. Dretske summarizes his information-theoretic account of beliefs as follows:

> ... a semantic structure has a more or less unique content, a content with a degree of intentionality comparable to that of a belief. If, then we identify beliefs with the *particular instances* (tokens) of these abstract semantic structures, we solve the problem ... of how (in terms of informational structures) to account for the possible falsity of beliefs, the problem of *mis-* representation. The way this problem is resolved is by realizing that a *type* of structure (a concept) may have informational origins (in the sense that that type of structure *developed* as a system's way of coding certain sorts of information) without (subsequent) instances of that structure having similar informational origins. [Dretske '81, p.196-7]

Before moving on to evaluate Dretske's account of belief as a solution to the Problem of Error, in order to complete the summary of Dretske's account of belief, a few words are in order to very quickly sketch how he accommodates the functional aspect of belief in his theory. Dretske's makes a distinction between semantic and cognitive structures, following up on an observation that in addition to having a representational content, a typical feature of beliefs is their effect on behaviour; having a particular belief may cause a person to act in a certain way. "A semantic structure qualifies as a *cognitive* structure ... insofar as its *semantic content* is a causal determinant of output in the system in which it occurs." [Dretske '81, p.199] This functional aspect of a structure in virtue of its semantic content is an essential quality of beliefs. These two aspects, functional and representational, are used to individuate cognitive structures or concepts. What concept a cognitive structure embodies is determined, in part, by its representational or semantic content, on its origins from information-bearing structures. But, two concepts or cognitive structures may be semantically equivalent but remain distinct concepts, if at least one of them is a complex concept and distinguishable from the other in its functional characteristics.

> Having identical semantic contents is not enough to make two structures the same *cognitive* structure. They must, in addition, be *functionally* indistinguishable for the system of which they are a part. And if the one structure is a composite structure, having parts that are themselves cognitive structures and the other is not (or is built up out of simpler cognitive structures in a different way), then the two structures constitute different concepts despite their *semantic* equivalence. It is this fact that makes it possible to distinguish, on purely information-theoretic grounds, between the *belief* that *s* is water and the *belief* that *s* is H_2O despite the fact that the information that *s* is water cannot be distinguished from the *information* that *s* is H_2O.[Dretske '81, p.218]

2.3. Critique of Dretske's Solution

2.3.1. Learning Period

There are problems with Dretske's account of false belief and, in particular, the learning period that are immediately apparent. How, it is reasonable to ask, is the learning phase to be distinguished from the application phase? At what point is the meaning of a particular representation finally "learned" so that it is now possible to have false beliefs involving it? Are there any principled criteria which could be brought to bear to help make this distinction? Intuitions might suggest that a meaning is learned once it starts to be used correctly, that the concept of *F* -ness is acquired once it is regularly correctly applied to things which are *F*. But this is of little help to Dretske's model, since it presupposes a mapping of structures to "correct" meaning or content which the learning phase is intended to establish. Dretske provides no non-arbitrary, non-circular way of delimiting the learning phase, where the meaning of the cognitive structure is in the process of being established through exposure to information bearing instances, from the point where the meaning of the structure is set, the application phase begins, and it is possible to misrepresent the state of affairs. Without detailed answers to these crucial questions, it is not clear how this approach can get off the ground.

2.3.2. Internal Problems

Even if, as Dretske supposes, the learning phase could be distinguished from the application phase for cognitive structures, the more serious problems for his model involve internal theoretic consistency. Consider the process of digitalization which converts a piece of information encoded in analogue form into a digital representation, that is, into a structure which has that piece of information encoded in digital form as its semantic content. Recall that a piece of information is said to be in analogue form if it is nested in other information carried in the structure and is in digital form if it is non-nested information. This transformation would involve taking an analogue representation of the information generated by the source and mapping it onto a digital representation of the same information. It is important to note that information is not generated by the digitalization process, rather, more specific information is stripped away, or filtered out as the digital representation is formed; the process goes from some piece of information in one form to that same piece of information encoded in a different form. While details about just how such a transformation might take place are sketchy in Dretske's account, this is probably just as it should be; it is more a matter for empirical investigation than for philosophic conjecture. But it is important to have a clear idea just what role this process is going to play within the context of Dretske's model.

This digitalization process takes structures or states in which a particular piece of information, say, that *s* is *F*, is carried in analogue form, to structures or states in which the information that *s* is *F* is carried in digital form, that is, a structure or state which carries no information in which *s* is *F* is nested; no information about *s* which is more specific. The digitalization process yields a structure that has the information that *s* is *F* as its semantic content. Dretske's account of false belief requires that instances of such a cognitive (belief) structure which has the semantic content that *s* is *F* can be triggered by signals which *lack* the information that *s* is *F*. On one hand, some tokens of this cognitive structure type result from a digitalization the information that *s* is *F*, corresponding to a true belief that *s* is *F*, and on the other hand, some tokens are triggered by signals or structures which lack the information that *s* is *F*, corresponding to a false belief that *s* is *F*.

Consider an information theoretic analysis of the situation described above. Some structures carry the information that *s* is *F* in analogue form. Call these A-structures. Since A-structures carry the information that *s* is *F*, the conditional probability of *s* being *F* given the occurrence of an A-structure is 1. If an A-structure is present, then *s* <u>is</u> *F*, without exception. There may be many different general structure types whose tokens count as A-structures, but each token of an A-structure carries the information that *s* is *F*. When A-structures undergo a process which digitalizes the information that *s* is *F*, that process strips away any more specific information about *s*, leaving the information that *s* is *F* as the only non-nested information in the resulting structure, as the semantic content of that structure. Call structures which during the learning period get type associated with the semantic content that *s* is *F*, tokens of type B. Some B-structures are the result of the digitalization of the information that *s* is *F*.

Take, for instance, a simplified example of this digitalization process. Suppose that a colour photograph shows a red square on the wall. Let s refer to the pictured item and assume that the photograph carries the conjunctive piece of information that (s is red & s is a square & s is on the wall). Clearly, the photograph carries the information that s is square in analogue form. Suppose further that a digitalization process has been developed which extracts the information that s is a square and presents it in digital form. Perhaps the process generates a black and white copy of the original photograph which obscures or leaves out any background indications of location. No more specific information is available in the resulting photograph than that s is a square. The original colour photograph is an A-structure and the black and white result of the digitalization process is a token of a B-structure.

According to Dretske's analysis of false belief, some non-A-structures, structures that do not carry the information that *s* is *F*, can also "trigger" B-structures. Some B-structures can occur when *s* is not *F*. This is the import of Dretske's comment that,

> Subsequent tokens of this [B] structure type inherit their meaning from the type of which they are tokens. What this means, of course, is that subsequent tokens of this [B] structure type can *mean* that *s* is *F*, can have this

> propositional content, despite the fact that they fail to carry this information, despite the fact that the *s* (which triggers their occurrence) is not *F* . [Dretske '81, p.193, underscoring mine]

To continue with the example, suppose that the black and white photograph of a square can be generated when there is no colour photograph of a square provided; perhaps a photograph of a blue rectangle taken at an angle will generate the same result, or it is possible to etch the outline of a square directly onto the black and white film.[11] Some tokens of the B-structure, the black and white photograph of a square, result from the digitalization process on A-structures, as in the case of the colour photograph of a red square and some tokens of the B-structure are "triggered" by structures or states which may arise when s is not a square and do not carry the information that s is a square.

There is a serious problem here. When Dretske introduced his notion of a signal or structure *r* carrying the information that *s* is *F*, he explained what he intended by conditional probability as follows:

> In saying that the conditional probability (given r) of *s* 's being *F* is 1, I mean to be saying that there is a nomic (lawful) regularity between these event types, a regularity which *nomically precludes r* 's occurrence when *s* is not *F*. ... A conditional probability of 1 between *r* and *s* is a way of describing a lawful (exceptionless)

---

11 Since, by hypothesis, the original colour picture did carry the information that s is a square, it must be assumed that it, and other A-structures, are veridical and immune to such mistakes or tampering.

> dependence between events of this sort, and it is for this reason that I say (in the text) that if the conditional probability of *s* 's being *F* (given *r*) is 1, then *s* is *F*. [Dretske '81, p.245, note.1, underscoring mine]

Now, if a structure of type B can be triggered by a non-A-structure, a structure which does not carry the information that *s* is *F*, it is possible for a B-structure to occur when *s* is not *F*. If this is possible, there is not an exceptionless nomic dependence of the sort required between the occurrence of B-structures and *s* 's being *F*. But, if that nomic dependence does not hold, on Dretske's conception of conditional probability, the conditional probability that *s* is *F* given the presence of any B-structure is less than 1. It is possible for a structure of type B to occur and for *s* not to be *F*. B-structures are not perfect indicators of *s* 's being *F*. Thus according to Dretske's definition of informational content, if a structure of type B can be triggered by a structure which does not carry the information that *s* is *F*, then no structures of type B carry the information that *s* is *F*. The black and white photograph of a square does not carry the information that s is a square. If an instance of the cognitive structure type corresponding to the belief that *s* is *F* can occur when *s* is not *F*, no instances of that structure type can carry the information that *s* is *F*.

According to Dretske's formulation of conditional probability, a B-structure cannot carry the information that *s* is *F* *even when it is a result of an A-structure, even when s is F*. Only if B-structures are "perfect indicators" of *s* being *F*, if there is a regularity which nomically precludes the possibility of B-structures occurring when *s* is not *F*, can structures of that type carry the information that *s* is *F*. Since on Dretske's account of false belief B-structures result both from A-structures when *s* is *F* and non-A-structures when *s* is not *F*, their occurrence is consistent with the possibility that *s* is not *F*, they are not "perfect indicators" of *s* being *F*. B-structures do not carry the information that *s* is *F* and, *a fortiori*, B-structures do not carry the information that *s* is *F* when they result from structures which themselves do carry the information that *s* is *F*, *vis.*, A-structures.

It can be assumed that the required nomic dependence exists between the state types corresponding to s being a square and the presence of the colour photograph since the colour photograph, by hypothesis, carries the information that s is a square in analogue form. But, since the presence of a black and white photograph of a square is consistent with the possibility that s is not a square (the black and white photograph might have been "faked") no black and white photographs of squares can carry the information that s is a square, not even the ones that were actually photographs of the square s.

This points out the conflict between Dretske's account of false belief and his description of the digitalization process which claims that B-structures resulting from the digitalization of A-structures have the information that *s* is *F* encoded in digital form. It cannot be true both that B-structures do not carry the information that *s* is *F* and that B-structures which result from a digitalization of A-structures carry the information that *s* is *F* in digital form.

The result that B-structures cannot carry the information that *s* is *F* holds even if a B-structure has never actually been instantiated as a result of a non-A-structure, that is, if all of the B-structures to date have been as a result of a digitalization process operating on structures carrying the information that *s* is *F*. As long as the possibility exists that a B-structure could occur when *s* is not *F*, that something other than an A-structure under the appropriate conditions *could* generate a B-structure, the required nomic dependency is not available. B-structures as a type are equivocal from an information theoretic point of view and, according to Dretske's formulation of informational content, cannot carry the information that *s* is *F*. In particular, such a B-structure would not carry the information that *s* is *F* even during the learning phase, when the semantic content of the cognitive structure type which is supposed to correspond to the belief that *s* is *F* is supposed to be fixed.

If B-structures, structures of the cognitive type supposed to correspond to the belief that *s* is *F*, never have the informational content that *s* is *F*, then since semantic content is, on Dretske's model, developed from, or a special sort of, informational content, it follows that structures of that cognitive (B) type can never have the information that *s* is *F* as their semantic content. If such structures cannot have the information that *s* is *F* as their semantic content, it is not clear how such a content could be acquired during the learning phase by the cognitive (B-type) structure and "conferred on subsequent tokens" in the manner Dretske describes. Dretske's analysis of error conflicts with his characterization of the probability relations involved in informational content, since if the resulting structure type can have tokens generated when *s* is not *F*, that cognitive (B) structure type cannot carry that information. Dretske's solution to the Problem of Error is inconsistent with the information-theoretic account of content to which it was intended to apply.

It is interesting to note that this result holds even if B-structures arise not as a result of any information-carrying structures at all, spontaneously, as it were, from the information theoretic point of view. As long as a the occurrence of a B-structure is consistent with *s* not being *F*, B-structures cannot carry that information however they might be caused.

There are many other problems or questions that can be raised for Dretske's account of false belief, such as exactly how a structure qualifies as a B-type cognitive structure (clearly, carrying the information that *s* is *F* cannot be the defining feature since the account of error requires tokens of such structure types which do not carry that information); how are semantic or cognitive structure types formed; is there any way to handle necessarily false contents which would have no information-carrying tokens; what is the status of the "semantic content" which is supposedly shared between a structure token corresponding to a true belief that *s* is *F* and a structure token corresponding to a false belief that *s* is *F* (the common view of semantic content would preclude one and the same semantic content being shared by two beliefs, one of which is true and the other false). None of these problems, however, can come into play unless the B-type cognitive structure can carry the information that *s* is *F*, and that is the most serious problem of Dretske's attempt to deal with the Problem of Error within an information-theoretic account of content.

These problems indicate that although Dretske makes a fair degree of progress toward giving a qualitative analysis of information and an account of informational content, his model founders in its attempt to account for false belief and the Problem of Error. Perhaps it is possible to distinguish the learning phase from the application phase of a cognitive structures life in some non-arbitrary, non-circular way, but it is not at all clear how this might be done and Dretske certainly has not done it. It might be possible to relax his "perfect indicator" criteria for information carrying, allowing the information carrying relation defined in formula (4) to hold if the conditional probability is less than one, but that would require a major re-working the notion of informational content as well as other parts of Dretske's model which have not been discussed here and rejecting some arguments central to his overall model which Dretske himself advances for why the criteria has to be as strict as it does.[12] Alternatively, he could try to move to a notion of informational content which is defined for tokens of a signal or structure, rather that for types, but, again, this would require major re-working of his model to accommodate such a change, if it were

---

12 See [Dretske '81,pp. 94-106].

even possible. While it is instructive to learn from these problems, Dretske's project cannot be judged a success from this perspective.

2.4. Dretske and Shannon

It is worth digressing a little at this point before leaving the consideration of Dretske to consider how Dretske's model of informational content relates to its foundations in Shannon's communication theory in order to clarify their differences and to see if the foundations offer any solutions to some of the problems regarding the relativization of information encountered in Dretske's model.

Dretske's theory diverges rather significantly from its foundations in Shannon's communication theory by relativizing information to receivers (or, more precisely, their background knowledge). Recall that, according to Shannon, a certain amount of information is generated when, in a situation where there is a number of events or states of affairs possible, one of those possibilities actually occurs. The amount of information associated with any particular event or state of affairs is a function of the number of alternatives that were available. This is an entirely objective and independent of any epistemic considerations regarding a receiver, or even any observers. The amount of information generated by a source (symbolized as $I(s)$ [13]), measures a reduction of

---

[13] $I(s)$ is not to be confused with the information carrying relation defined earlier. The amount of information associated with or generated by some

possibilities. For instance, a signal generated by a source in one state out of 6 other possible states the source could be in carries less information that a signal generated by a different source whose current state is a reduction of 10 other possibilities.

Dretske suggests, however, that both the content and the amount of information depends on the knowledge of the receiver. He suggests,

> a *relativization* of the information contained in a signal because *how much* information a signal contains, and hence *what* information it carries, depends on what the potential receiver knows about the various possibilities that exist at the source.[Dretske '81, p.79]

and,

> a receiver's background knowledge is relevant to the information he receives (both *how much* information and *what* information) ... to the extent that it affects the value of $I(s)$ — the amount of information generated at the source by the existence of a specific state of affairs. [Dretske '81, p.81]

When Dretske claims that the *amount* of information[14] carried by a signal is relative to what the receiver knows, he is claiming that in the generation of information (when the possible states of affairs at

---

source s , which reduces n equally likely states to a single actual state is typically given by the following formula: $I(s) = \log n$, where log is the logarithm to the base 2 (the power to which 2 must be raised to get n).

14 Dretske accepts Shannon's definition of the amount of information, or at least offers no alternative.

the source are reduced to one) the number of states of affairs possible at the source is affected by what the receiver knows about those possibilities. This definitely sounds odd, if not simply mistaken. At a minimum it is a conflation of two different interpretations of what it means to be a possible state of affair at the source: the logically or physically possible alternative states which could have obtained at the source and those alternative states which have not, as far as the receiver knows, been eliminated as possibly the actual state which obtained at the source.

It is possible to talk of a roll of a die as a reduction of 6 possible states (faces showing on top) to a single state which actually obtained. It is also possible to talk of someone who knows that neither a 1, nor a 2, nor a 3, nor a 4 was rolled, upon learning that a 5 was not rolled, having the possible states of the die reduced to a single possibility from two, knowing what she does about the number of faces on a die. Both seem to be, in their own ways, legitimate ways of talking about and counting possibilities which may obtain at the die (the source). What is clear, however, is that they are not two ways of talking about the same thing.

Shannon defined the amount of information associated with a source in terms of the former sort of possible states of the source, as evidenced by the absence of any consideration of any effect a receiver's knowledge on the transmission of information and the fact that the calculations of $I(s)$ depend of the probabilities of each of the possible states of the source obtaining. It seems, however, to be over the latter sense of possible states which Dretske wishes to describe the amount and content of information as relative to the receiver's background knowledge (the example he uses to motivate the need for relativization is similar to the die case[15]). Clearly, the knowledge of the receiver (or even the existence of a knowledgeable receiver at all) is irrelevant to the former method of counting logically or physically possible states of the source and thus cannot affect the value of $I(s)$ as defined by Shannon and (tacitly) adopted by Dretske.

Neither method of counting possible states of the source immediately presents itself as being more appropriate for discussions of the informational content of signals. What is necessary, however, is that the distinction between the two is kept clearly in mind and equivocation between the two avoided since they are not equivalent ways of carving up a common subject. Dretske appears to have confused this point in his explanation of the relativization of information.

Dretske's relativization of the informational content of a signal may strike some as being problematic for another reason. One of the consequences of defining informational content of a signal in terms of the background knowledge of the receiver about the source is that it is not clear what can be said, if anything, about the informational content of a signal independent of any knowledgeable receiver. Dretske's formulation would suggest that there is no informational content independent of a knowledgeable receiver. Even if this is too

---

15 See [Dretske '81, pp.79-80].

extreme a characterization, at a minimum Dretske is silent about the informational content of signals independent of knowledgeable receivers and how it would relate (since it would not be unreasonable to expect some relation) to the conception of informational content for knowledgeable receivers that is provided.

Recall, as well, one of the problems encountered with Dretske's model which was left in abeyance. Dretske's account of the informational content of a signal which relativizes the information carrying relation to the knowledge of the receiver (as in formula (4)) yielded problems for his account of the nested information. A signal might not carry all of the information which is nested in information which is carried by the signal, potentially causing problems for Dretske's account of semantic content. Clearly, if a piece of information is not carried in the signal, it is not available to be extracted by the digitalization process which generates cognitive structures.

Perhaps the sort of relativization of the informational content of a signal to a receiver's knowledge which Dretske describes and which is captured in formula (4) is too strong, leading to the sort of problems identified above. Perhaps a weaker form of relativization could capture the intuition that a signal may provide one person, knowing certain things about the source, with a particular content which may not be conveyed to another person who had a significantly different background knowledge. The foundations laid by Shannon's communication theory, which Dretske, himself, outlines in his first chapter, provide a possible suggestion in this regard. Communication theory distinguishes the amount of information generated or associated with the source, symbolized as *I(s)*, from the amount of information associated with the receiver, symbolized as *I(r)*. The details of this distinction are not particularly relevant to this discussion but the distinction itself is useful. If the informational content generated by or associated with a source were likewise distinguished from the informational content associated with a receiver, the details of a theory of informational content might be drawn out in such a way that it avoids the problems noted above. In what follows a quick sketch is presented of how such a theory might work. It makes no pretence of being complete or more than a brief indication of the directions which might be explored in order to produce a substantial alternative theory of informational content.

The informational content associated with a receiver (call it the information received or the R-content of a signal) can be straightforwardly relativized to the knowledge of the receiver, much as described in formula (4). If a receiver does not have the appropriate background knowledge then the signal may fail to convey the information that *s* is *F* for that receiver. That piece of information is not part of the R-content of the signal for that receiver, although another receiver may receive the information that *s* is *F* as part of the R-content from the same signal. For example if a signal r was generated by *s* 's being a square, but a receiver already knows that *s* is square (it is included in the receiver's *k*) then the information that *s* is a square would not be in the R-content of signal r for that receiver, even though it would, presumably, be in the content of the information generated by *s* and potentially in the R-content of r for other receivers. Similarly, in the shell game example, person X who already knows there is no pea under shell A receives from a signal on seeing that shell B does not hide a pea an R-content that is different from the R-content for person Y who has no knowledge about what is (or is not) under shell A. Information received could operate much in the relativized manner as described in section 2.1.1 above.

The informational content of a signal generated by the source (the G-content) could then be defined in a non-relativized manner, independent of any background knowledge which a receiver might have, independent of the existence of any knowledgeable receiver at all. This would yield a reconciliation of *prima facie* conflicting intuitions, with a G-content of an information bearing signal independent of receivers as well as an R-content of the signal which is relativized to knowledgeable receivers.

The problems noted above associated with nested information and whether a signal carries all the information nested in its informational content could be addressed by restricting the discussion of nested information to the information generated, rather than the information received. The receiver may or may not receive all the nested information in a signal, depending on the background knowledge possessed, but the difficulties identified above would not apply to the content of the information generated, which is entirely independent of receiver's background knowledge.

Since the digitalization process which Dretske describes in his account of semantic content and belief is not one which is necessarily carried out by a "knowledgeable" processor, the problems noted for the digitalization process and the missing nested information could likely be avoided by amending the description in accord with this distinction. It may even turn out that Dretske's talk of the informational content of a "structure" as opposed to a signal with a particular receiver, which is central to this portion of his theory, is better captured by the receiver independent view of informational content than the formulation he actually uses.

The variety of ways of counting possible states and, in turn, the amount of information carried by a signal roughly corresponds to the distinction between information generated and information received. It would seem reasonable to retain Shannon's objective calculation of the amount of information to describe the information generated by a source. It may be possible to utilize some measure of the amount of information received which depends, in part, on what the receiver knows about the possibilities at the source, although such a characterization is in need of clarification and elaboration which Dretske's attempt fails to provide.

As was noted above, this sketch is not intended to provide a complete theory of informational content. The responses to the problems raised above for Dretske are not extensive nor complete, and there are no doubt other problems that arise for this approach. For instance, a characterization of the informational content generated by a source, independent of receivers is required. It cannot be as simple as dropping the conjunct which relativizes the content to background knowledge of formula (1) and all other reference to background knowledge. That would result in every signal carrying all information that holds with a probability of 1, including all laws of science and logic, too much information for every signal to carry.

Once a characterization of the receiver-independent content of the information generated by a source is provided, it must be linked to the informational content for a particular receiver from a signal. At a minimum, it would seem unlikely that a successful theory would allow for someone to receive information from a signal that was not part of the information generated by the source. More detail of the relation between these two sorts of informational content needs to be spelled out before the theory can be considered complete.

This distinction between the information content for a knowledgeable receiver and the content of a signal as generated, independent of background knowledge, is not necessarily in conflict with the bulk of Dretske's theory. It adds a refinement which would require some amendment of the details, but that need not be too serious. In fact, there are at least two indications that something along these lines is what Dretske has in mind. The first occurs in the introductory overview of Shannon's theory of information and additionally provides some indication as to how the relation between generated and received informational content might be worked out. Dretske points out that while the orthodox application of the theory selects

> the set of possibilities at the source and receiver so as to make $I(s) = I(r)$, ... [he] envisage[s] changes in the output ensemble [the set of possibilities defining $I(r)$] without corresponding changes to the input ensemble [those possibilities defining $I(s)$] and vice versa.[Dretske '81, p.239, note 13]

Consider a case such as Dretske envisages, where the number of possibilities at the receiver defining $I(r)$ is reduced without any change to the "input ensemble." It would not be an unreasonable reading of Dretske's general discussion of informational content, although he does not deal with this point specifically, to suggest that since the input ensemble has not changed, in addition to the amount of information associated with the source $I(s)$ remaining constant, the content of the signal generated by the source also remains unchanged, unaffected by any change to $I(r)$.

Presumably, if $I(r)$ is changed, the informational content of the received signal also changes, at least to the extent that the amount of information that is received constrains what information can be contained in the signal. If, as Dretske suggests, the background knowledge of the receiver affects the possibilities that may obtain defining $I(r)$, reflected in a corresponding change to the informational content received, this in itself is not enough, as Dretske points out, to suggest that there has been any change to $I(s)$. If, as described above, this constancy, in turn, applies to the informational content of the signal generated by the source, a mechanism has been described whereby the content received from a signal can diverge from the content of the signal as generated.

Dretske's "unorthodox" treatment of the amount of information is in line with the sort of distinction between the content of a signal as generated and as received suggested above.

The second indication that this distinction might not be totally at odds with Dretske's theory is found in his discussion of nested information. When explaining the reason why all nested information should be carried by a signal, Dretske says, "[t]his is so because if the conditional probability (given *r*) of *s* 's being a square is 1, then the conditional probability (given *r*) of *s*' s being a rectangle is also 1."[Dretske '81, p.70] To be consistent with his own definition of the informational content of a signal (in either the original or amended form) the conditional probability should include reference to the background knowledge of the receiver, relativizing the nested information to that background knowledge. The lack of any reference to background knowledge would be consistent with the suggestion described above, that the nesting relation (particularly the claim that a signal carries all information nested in its content) applies to the information generated by a source rather than the information received which is relativized to the background knowledge of the receiver.

There is room to improve and clarify Dretske's theory of informational content along these lines which, as indicated, may not be entirely contrary to what was originally intended although there is no explicit discussion of such a distinction. Clearly a distinction between the informational content generated and the informational content received, such as has been suggested, would involve a great deal of tinkering with aspects of Dretske's overall model in order to solve some of the problems that arise for his conception of informational content. Unfortunately, such an endeavour, no matter how successful at dealing with the issues and the problems raised here for his theory of informational content, would not be able to avoid the more fundamental problems relating to the Problem of Error that were described in the previous section. Whether the fruits of such a project would justify the effort required is for someone else to decide.

Chapter 3. Fodor

This chapter examines Jerry Fodor's approach to the Problem of Error. It begins by summarizing Fodor's Representational Theory of Mind as characterized in [Fodor '87]. The theory of content which Fodor proposes is described highlighting how it deals with the Problem of Error and a related, more general problem. The chapter concludes with a critique of Fodor's theory of content and his solution to the Problem of Error.

3.1. Fodor's Representational Theory of Mind

Fodor sets out to give a naturalistic, that is, a non-intentional and non-semantic account of the mind in terms of his Representational Theory of Mind (RTM). Such an account would provide a reduction of intentional mental phenomena to physical phenomena which can be brought into a scientific model of psychology. RTM consists of a central hypothesis, called the Language of Thought hypothesis (LOT), which is focussed and elaborated by the conjunction of two claims, one about the nature of propositional attitudes and the other about the nature of mental processes. LOT postulates "an infinite set of 'mental representations' which function both as the immediate objects of propositional attitudes and as the domains of mental processes."[Fodor'87, pp.16-17] These mental representations have intentional contents and their tokens are physically realized (presumably as neural entities) exhibiting causal powers and roles and interacting within the system of brain states of their possessors.

Mental representations are both semantically evaluable, yielding the content of the representation, and causally efficacious.

In addition, "LOT claims that *mental states* [1] — and not just their propositional objects — *typically have constituent structure.*"[Fodor '87, p.136] Not only are the intentional contents of mental representations (propositions, for instance) structured entities but the mental representations, the physical symbols, themselves are structured, possessing a syntax as well as a semantics. Sentences which are the symbols which English uses to express propositions are complex, syntactically structured entities, made up of simpler symbolic components, such as phrase units, and words, combining in an orderly fashion which reflects the semantic relations between the contents of the sentence components. Similarly, according to LOT, mental states, the physical symbols themselves, "constitute a language; roughly, the syntactic structure of mental states mirrors the semantic relations among their intentional objects."[Fodor '87, p.138] This is to be contrasted with alternative views which are similar to LOT (and RTM) in that they are realist with respect to intentional contents of propositional attitudes such as belief and desire, within a framework of physicalism, but which maintain that only the intentional contents of mental states are complex structured

---

1 Fodor appears to use the terms 'mental representation', (mental) 'symbol' and 'mental state' more or less interchangeably at various points of his discussion.

entities; mental states or symbols are simple entities that have no syntactic structure.[2]

The two claims of RTM which supplement and focus LOT are:

> *Claim 1* (the nature of propositional attitudes)
> For any organism *O*, and any propositional attitude *A* toward the proposition *P*, there is a ('computational'/'functional') relation *R* and a mental representation *MP* such that
>
> *MP* means that *P*, and
> *O* has *A* iff *O* bears *R* to *MP*.

and

> *Claim 2* (the nature of mental processes)
> Mental processes are causal sequences of tokenings of mental representations.[Fodor '87, p.17]

Claim 1 is intended to capture the core thesis of *representational* theories of propositional attitudes: to have a propositional attitude such as a belief is to bear an appropriate relation to a mental representation which expresses or has as its content the proposition believed. It is in virtue of *O* bearing this appropriate relation to a mental representation which expresses *P* or means that *P* that *O* stands in the believing relation to the proposition *P*, that *O* believes that *P*. Another way of putting this point is to say that the believing relation between *O* and *P* is mediated by the more direct relation

---

2 For a more complete discussion of LOT and Fodor's argument in favour of this hypothesis, see [Fodor '87, pp.136-154].

between *O* and a mental representation which expresses *P*. Differences in the causal role of beliefs and other propositional attitudes are reflected in differences in the ('computational'/'functional') relations which *O* might bear to a mental representation which has the content toward which the attitude is directed.

Claim 2 describes the RTM account of the nature of mental processes.

> A train of thoughts, for example, is a causal sequence of tokenings of mental representations which express the propositions that are the objects of the thoughts. To a first approximation, to think 'It's going to rain; so I'll go indoors' is to have a tokening of a mental representation that means *I'll go indoors* caused, in a certain way, by a tokening of a mental representation that means *It's going to rain.*[Fodor '87, p.17]

The issues surrounding RTM's account of the nature of mental processes generally do not impact on the concerns of Fodor's theory of content and the Problem of Error and will be only be raised below as required.

There are some problems with claim 1 as it is stated. To more clearly see the difficulties, claim 1 can be formalized as follows:

(1) $\forall p\ \forall o\ \forall A\ \exists R\ \exists m\ [\mathrm{M}(m, p) \,\&\, (A\ (o, p) \text{ iff } R\ (o, m))]$

where the variable p ranges over propositions, o over organisms, A over propositional attitudes, R over ('computational'/'functional') relations, m over mental representations and where 'M(m , p)'

stands for '*m* means *p* ', '*A* (*o* ,*p*)' stands for '*o* bears *A* toward *p* ', and '*R* (*o* , *m*)' stands for '*o* bears *R* to *m* '.

Formula (1) implies that for every proposition *p* and organism *o* there exists a mental representation *m* . But this means that there must be enough mental representations to express all propositions. This seems too stringent a requirement. Formula (1) is also too weak in that it permits the relation *R* to vary across propositions and individuals. It is intuitively undesirable to allow that a single attitude, say, belief, be realized in virtue of one relation to mental representations for some propositions that are believed and a different relation for other propositions that are believed. It is, similarly, undesirable to allow *R* to vary across individuals.

There is a further difficulty with formula (1). As was pointed out above, the key feature of representational theories of propositional attitudes is that having an attitude toward a proposition amounts to bearing an appropriate relation to a representation of that proposition. Such a theory cannot permit the possibility of a situation where a person bearing the appropriate relation to a representation of a proposition fails to have the corresponding attitude to that proposition. A theory which allowed such a situation could not be a representational theory of propositional attitudes.

According to formula (1), if *o* bears *R* to <u>all</u> mental representations which mean that *p* , then *o* has *A* to *p* ; bearing *R* to at least one mental representation which means that *p* is necessary for bearing *A* to *p* . But, unless it is assumed that a proposition can be expressed by no more that one mental representation, it is possible for an individual to bear *R* to some mental representations which mean that *p* but to fail to bear *R* to others. A problem arises for formula (1) in just such cases where *o* bears *R* to some but not all mental representations which express *p*. Consider (2), an instantiation of (1) with a proposition p, organism o, belief (attitude) B, and appropriate relation R:

(2) $\exists m\ [M(m, p) \& (B(o, p) \text{ iff } R(o, m))]$

Suppose that there are just two mental representations that express p, m_1 and m_2, and suppose further, that o bears R to m_2 but does not bear R to m_1. One way in which formula (2) is true is for formula (3) to be true, where m has been instantiated with m_1.

(3) $M(m_1, p) \& (B(o, p) \text{ iff } R(o, m_1))$

Since, by hypothesis, $R(o, m_1)$ is false, B(o, p) must be false in order for the biconditional, and hence formula (3), to be true; if formula (3) is true, o does not believe that p. But, also by hypothesis, o does bear R to another representation of p, namely, m_2. Thus, formula (1) is consistent with o failing to believe that p (implied by (3) and hence by (1)) while bearing the required relation corresponding to belief to a representation of p. Formula (1) does not preclude the possibility of a situation where $R(o, m_2)$ & $M(m_2, p)$ & $\sim B(o, p)$ is true. Such a situation cannot be allowed by a representational theory of propositional attitudes and since it is permitted by formula (1), claim (1) is not acceptable as it has been stated.

Rather than attempt to modify RTM's claim 1 quoted above and symbolized as formula (1), it is useful to note that following that quoted passage, Fodor goes on to say,

> A cruder but much more intelligible way of putting claim 1 would be this: To believe that such and such is to have a mental symbol that means that such and such tokened in your head in a certain way; it's to have such a token 'in your belief box,' as I'll sometimes say.[Fodor '87, p.17]

This remark suggests the following characterization of claim 1:

Claim 1'
For any attitude A there is a ('computational'/'functional') relation R such that for any organism O and proposition P,

O has A towards P iff

there is a mental representation MP such that

MP means that P and O bears R to MP.

Claim 1' may be formalized as:

(4) $\forall A\ \exists R\ \ \forall o\ \forall p\ \ [A\ (o\,,p\,)\ \text{iff}\ \exists m\ (\mathrm{M}(m\,,p\,)\&\ R\,(o\,,m\,))]$

This version of RTM's claim about the nature of propositional attitudes avoids the problems noted above. There is no implication that there is a mental representation for every proposition; only those toward which someone has an attitude. R varies across attitudes, but not individuals nor propositions, in accord with intuitions. And finally, it is clear that o has A toward p if there is <u>any</u> mental representation which means that p and which is related to o

by *R*. When required below, claim 1' and its formal version in formula (4) will be assumed when discussing RTM.[3]

RTM is built upon the Language of Thought hypothesis postulating the existence of a set of syntactically structured mental representations with intentional contents, which are the immediate objects of propositional attitudes. Mental representations form the domain over which mental processes are characterized. RTM claims that to have a propositional attitude such as a belief is to bear a relation to a mental representation which expresses or has as its content the proposition which is believed and also claims that mental processes are causal sequences of tokenings of mental representations. With an outline of RTM in hand, it is now possible to consider the theory of content which Fodor proposes for RTM.

3.2. A Causal Theory of Content

Fodor's project requires a "*naturalized* theory of meaning; a theory that articulates, in nonsemantic and nonintentional terms, sufficient conditions for one bit of the world to *be about* (to express, represent, or be true of) another bit."[Fodor '87 p.98] Such a theory would explain the content of mental representations, how mental symbols and their physical tokens are related to the what the symbols mean. Fodor builds his theory of content on an intuition regarding

---

3 See [Kazmi, forthcoming] for a discussion of these difficulties with claim 1 and see [Soames, forthcoming] for an extended discussion of matters relating to alternative formulations of claim 1.

the role of causality in meaning and proposes an elaboration to avoid critical problems which arise for the raw intuition.

Although Fodor is committed to LOT and believes the symbols of a natural language such as English derive their semantical properties from the mental representations which they are normally used to express, he often constructs his examples and explanations in terms of linguistic symbols. This is usually only to aid in presenting a less complicated discussion and in those cases where the difference between mental and linguistic symbols is significant, the distinction is made explicit.

3.2.1. A Causal Intuition

The intuition on which Fodor builds is quite straight forward. Cases of predication, where a predicative expression is applied to some object, form the core of this intuition. According to this causal intuition, "[i]n such cases the symbol tokenings denote their causes, and the symbol types express the property whose instantiations reliably cause their tokenings. So, in the paradigm case, my utterance of 'horse'[4] says *of* a horse that it *is* one."[Fodor '87 p.99] The symbol 'horse' means *horse* because horses reliably cause that symbol to be tokened.

---

[4] Fodor's orthographic convention that names of words and mental symbols in quotes and names of properties will appear in italics will be followed below. See [Fodor '87, p.160, note 5].

The notion of 'reliable causation' is intended to avoid situations where a symbol tokening results from an occurrence of a predicative object by, as it were, happenstance. For instance, if it were said of a horse, "there is a cow", where the symbol 'cow' was tokened due to a slip of the tongue, this may not be an instance of reliable causation.

> 'Reliable causation' requires that the causal dependence of the tokening of the symbol upon the instancing of the corresponding property be counterfactual supporting: either instances of the property actually do cause tokenings of the symbol, or instances of the property *would* cause tokenings of the symbol *were they to occur*, or both ... [I]t is necessary and sufficient for such reliable causation that there be a nomological – lawful – relation between certain (higher-order) properties of events; in the present case, between the property of being an instance of the property *horse* and the property of being a tokening of the symbol 'horse'.[Fodor '87, p.99][5]

On this intuitive model, reliable causation connects properties with symbol types via nomic laws. A token of the symbol 'horse' denotes a horse, that is, denotes an instance of the property *horse*, in virtue of a nomological relation between the property type *horse* and the symbol type 'horse', between horses and 'horse's. It is in virtue of this same nomological relation which the symbol type 'horse' expresses the property *horse* . As Fodor puts it, "the intuition

---

5 Although this analysis of causation may not be universally accepted, when referring to causation below, Fodor's notion of reliable causation as a nomic relation between higher order properties will be assumed.

that underlies [this theory] is that the semantic interpretations of mental symbols are determined by, and only by, such nomological relations."[Fodor '87, p. 99][6]

3.2.2. Problems: Error and Disjunction

There are some immediate problems that arise for this intuitive model of content. It does not appear possible, within such a model, for a symbol token to misrepresent. This is the Problem of Error; how can the content of a symbol be fixed so that it is possible for it to be falsely tokened.

> Suppose, for example, that tokenings of the symbol 'A' are nomologically dependent upon instantiations of the property *A* ; viz., upon A's. Then, according to the theory, the tokens of the symbol denote A's (since tokens denote their causes) and they represent them *as* A's (since symbols express the property whose instantiations cause them to be tokened). But symbol tokenings that represent A's as A's are ipso facto veridical. So it seems the condition for an 'A'-token meaning *A* is identical to the condition for such a token being true. How, then, do you get *un*veridical 'A' tokens into the causal picture?[Fodor '87, p.101]

Consider an unveridical 'A' tokening in contrast with a veridical A-caused 'A' tokening, that is, a case where some instantiation of a

---

6 It has been argued, that contrary to this intuition, semantics involves in addition consideration of how symbols are used and the social and instutional contexts in which they are employed. See [Martin,'89]. This view will not be pursued here.

property B distinct from property A, causes an 'A' tokening. On this intuitive causal account of content, the property which 'A' expresses is the property which reliably causes 'A' tokenings. Since both A's and B's are sufficient to cause 'A' tokenings, the property expressed must be the disjunctive property of being ($A \vee B$). But if the property expressed by 'A' is ($A \vee B$), then, contrary to assumption, a B-caused 'A' tokening is veridical. It appears that there is no way, on this intuitive model, for a symbol tokening to misrepresent.

Fodor sets up the following example to illustrate the Problem of Error for this intuitive theory of content:

> I see a cow which, stupidly, I misidentify. I take it, say, to be a horse. So taking it causes me to effect the tokening of a symbol; viz., I say 'horse.' ... [O]n the one hand, we want it to be that my utterance of 'horse' means *horse* in virtue of the causal relation between (some) 'horse' tokenings and horses; and, on the other hand, we *don't* want it to be that my utterance of 'horse' means *cow* in virtue of the causal relation between (some) 'horse' tokenings and cows. But if the causal relations are the same, and if causation makes representation, how can the semantic connections not be the same too?[Fodor '87, p.107]

Since the property *cow* reliably causes tokenings of the symbol 'horse', presumably under appropriate circumstances such as poor lighting or at a great distance, according to the intuitive causal model of content, 'horse' must include the property *cow* in its meaning. If only horses and cows (under appropriate circumstances) reliably cause tokenings of the symbol 'horse', the intuitive causal model then assigns the property of being ($horse \vee cow$) to the meaning of the symbol 'horse'.

> A viable causal theory of content has to acknowledge *two* kinds of cases where there are disjoint causally sufficient conditions for the tokenings of a symbol: the case where the content of the symbol is disjunctive ('A' expresses the property of *being* ($A \vee B$)) and the case where the content of the symbol is *not* disjunctive and some of the tokenings are false ('A' expresses the property of *being* A, and B-caused 'A' tokenings misrepresent). The problem with the crude causal theory is that it's unable to distinguish between those cases; it always assigns disjunctive content to symbols whose causally sufficient conditions are themselves disjoint.[Fodor '87, p.102]

A theory of content should match up symbols and contents in such a way that conforms to independent intuitions about the contents of the symbols, or, where departure from those independent intuitions is necessary, provide a strong argument for the need to revise or reject the independent semantic intuitions. The intuitive causal theory holds that symbol tokenings denote their reliable causes and symbols express the property which reliably cause their tokenings. The problem for this theory is that for a symbol, in addition to the property which intuitively is what the symbol expresses, other properties might reliably cause tokenings of the symbol. As was seen above, cows mistaken for horses can reliably cause 'horse' tokenings. According to the intuitive causal theory, 'horse' must therefore express the disjunctive property ($horse \vee cow$). But this conflicts with independent intuitions which maintain that 'horse' expresses the non-disjunctive property *horse* . Fodor calls this an example of the Problem of Disjunction.

The Problem of Disjunction is more far reaching and general, however, than is apparent just from considerations of the Problem of Error. In a more recent work, Fodor makes the important observation that the Problem of Error is just a sub-species of the much larger Problem of Disjunction for causal theories of content. There are circumstances under which 'horse' is tokened as a result of neither an instance of the property *horse,* nor some other property mistaken for *horse.* Consider a situation where someone is discussing animals and happens to say or think,

(5) A horse is a beautiful animal.

This perfectly legitimate occurrence of the symbol token 'horse' in (5) is not caused by any particular instance of the property *horse* nor by an instance of another property mistaken for *horse* . Perhaps the occurrence of the symbol token 'horse' was caused by a preceding thought about other beautiful animals or a thought about animals which are used for racing or something else entirely. The point is that the symbol token need not have been caused by anything which is in the extension the symbol-type 'horse' nor by something mistaken to be within that extension. The sort of "train of thoughts" envisaged by RTM's second claim that mental processes are causal sequences of tokenings of mental representations provides a case in point of this sort of symbolic causation. The general Problem of Disjunction is that if a symbol type expresses the property which reliably causes its tokenings, why does the symbol type 'horse' not express the property of (*horse* $\vee$ *cow* $\vee$ *preceding mental representation* $\vee$ ...) and so on for each property which whose tokens reliably cause 'horse' tokenings under appropriate circumstances?

Fodor notes that occurrences of symbol tokens such as 'horse' in (5) are used to represent the extension of the term rather than to apply to any particular member of that extension. Such "representing" occurrences of symbol tokens are to be distinguished from "labelling" occurrences where, for instance, a token of 'horse' says of a horse that it is one. Representational tokenings of 'horse' need not be caused by any particular horse, in contrast with labelling tokens of 'horse'. The token of 'horse' in (5) is used representationally to stand for the things that it applies to, rather than to apply to any particular instance of the property *horse* .

These problems for the intuitive causal theory of content can be summarized as follows. Independent intuitions indicate that tokenings of 'horse' express the property *horse* . Tokenings of 'horse' are sometimes caused by instances of the property *horse* , that is, things in the extension of 'horse' (by horses); these are the so-called labelling occurrences of 'horse' tokenings. Tokenings of 'horse', however, are also sometimes caused by things which are not instances of the property *horse,* things that are not in the extension of 'horse'. An erroneous tokening, where, say, a cow is mistaken for a horse, is one such sort of 'horse' tokening. Representational occurrences of 'horse' tokenings such as in (5) are another sort of 'horse' tokenings that are not caused by instances of the property *horse*. The intuitive causal theory of content then yields the result that 'horse' expresses the property (*horse* $\vee$ *cow* $\vee$ *preceding mental representation* $\vee$ ...). This intuitive account is in serious conflict with the independent intuition that 'horse' expresses the property *horse* and not some disjunctive property.

3.2.3. Proposed Solutions

In cases of labelling tokenings, where 'A' tokenings are caused by instances of the property *A* , the intuition that a symbol expresses the property that reliably causes its tokenings coincides with independent intuitions about symbol meanings. The causal intuition diverges, however, from those independent intuitions about symbol meanings when 'A' tokenings are caused by non-A's, for instance, in cases of representational tokenings and mistaken tokenings. One way to save the intuition at the core of the causal theory of content discussed above is to find some feature which can differentiate between these two cases of symbol tokenings: A-caused 'A' tokenings and non-A-caused 'A' tokenings. With such a distinction in hand, a more sophisticated causal theory of content could then assign a meaning to a symbol on the basis of the causal intuition described above for labelling tokenings, and maintain that the content of mistaken and representational tokenings of a symbol is derivative from the content causally established by labelling tokenings.

The problem, then, is to find some difference between A-caused 'A' tokenings and non-A-caused 'A' tokenings which allows for a distinction to be drawn between causally established symbol content and derivative symbol content. Since Fodor wishes to use this causal theory of content to naturalize and explain semantic and intentional properties, he cannot cannot utilize such properties in making this distinction; the difference must be spelled out in terms of non-intentional and non-semantic properties of causal relations.

3.2.3.1. Qualified Tokening

One approach to solving the Problem of Error which has been tried in the past (including by Fodor, at one time) is to modify or qualify the conditions under which a tokening occurs. Consider the example where sometimes cows are mistaken for horses and cause 'horse' tokenings. Perhaps the lighting is poor or the distance is great. Presumably, however, there are conditions which could be specified under which only horses, things which are truly instances of the property *horse* , cause 'horse' tokenings. Under these specified conditions, sources of error would be excluded and nothing which did not have the property *horse* would cause a 'horse' tokening. These conditions might variously be described as optimal, ideal, ecologically valid, normal, counterfactual supporting, etcetera. On such an account, erroneous tokenings could be distinguished from true or veridical tokenings on the basis of counterfactual properties. If the tokening had taken place under the specified conditions, only horses would cause 'horse' tokenings; cows, for instance, would not.

There are some problems involved in choosing what the specified conditions should be, from among the various options available and, more importantly, whether a characterization of such conditions could be spelled out in non-semantic and non-intentional terms that would apply only to veridical tokenings. For instance, Fodor points out that "it looks as though error is a perfectly normal feature of the use of symbols, and there appears to be no reason why the statement that such-and-such circumstances regularly cause errors shouldn't be counterfactual supporting,"[Fodor, forthcoming, p.8] making 'normal conditions' or 'conditions which support counterfactual causal statements' unlikely qualification candidates.

There is, however, a more serious problem for this approach, which Fodor noted after further consideration of the Problem of Disjunction. Even if it were possible to set up a characterization of specified conditions under which the disjunction problem is handled for the case of error, such a qualification will not help fix the content of a symbol in the case of representational occurrences of symbol tokenings. Since each of these cases are instances of symbol tokenings caused by things that are not in the symbol's extension and, thus, varieties of the general disjunction problem, ideally a single unifying approach to a solution would be applicable to both.[7] But, "idealizing away from sources of error won't work for representation because representational occurrences of ['horse'] don't covary with [horses] <u>even when they are true</u>. ... [M]aybe you

---

7 Of course, if no single unifying approach were available to solve the more general disjunction problem, a solution for the Problem of Error which did not address the disjunctive content results of representational tokenings would be better than no solution at all. Fodor, however, thinks that he has such a unifying solution.

can idealize away from mislabelling; but surely you can't idealize away from thinking.[Fodor, forthcoming, p.9] Since qualifying the content fixing tokenings to avoid sources of error is not sufficiently general to handle the Problem of Disjunction for representational tokenings, Fodor rejects this approach as a viable solution to the Problem of Error. The solution he requires must be general enough to apply to the other cases of the more general Problem of Disjunction, such as representational tokenings, that arise, in part, due to the causal sequences of tokenings of mental representations involved in RTM's account mental processes.

3.2.3.2. Asymmetrical Dependence

Fodor's alternative solution, which is developed to handle the Problem of Error and extends to the more general Problem of Disjunction, seizes on an idea he attributes to Plato: that error is ontologically dependent on truth.

> The mechanisms that deliver falsehoods are somehow *parasitic on* the ones that deliver truths. In consequence, you can only have false beliefs about what you can have true beliefs about (whereas you can have true beliefs about anything that you can have beliefs about at all).[Fodor '87, p.107]

This observation suggests to Fodor that it should be possible to distinguish true or veridical tokenings from erroneous ones by considering the dependence relations between the nomic laws that govern the corresponding tokenings. This might be accomplished by looking for differences in the counterfactual properties of the tokenings. To the extent that it considers the counterfactual properties of tokenings, Fodor's current suggestion is similar the qualified tokening approach examined above. It differs, however, in which counterfactual properties it considers.

Consider the example examined earlier. The fact that a cow mistaken for a horse causes 'horse' tokening depends on the fact that horses cause 'horse' tokenings.

> [M]isidentifying a cow as a horse wouldn't have led me to say 'horse' *except that there was independently a semantic relation between 'horse' tokenings and horses.* But for the fact that the word 'horse' expresses the property of *being a horse* (ie., but for the fact that one calls *horses* 'horses', it would not have been *that* word that taking a cow to be a horse would have caused me to utter.) Whereas, by contrast, since 'horse' does mean *horse* , the fact that horses cause me to say 'horse' does not depend upon there being a semantic — or, indeed any — connection between 'horse' tokenings and cows.[Fodor '87, pp.107-108]

Thus, the nomic relation between instances of the property *cow* and instances of the symbol type 'horse' is dependent on the nomic relation between instances of the property *horse* and instances of the symbol type 'horse'. But the reverse dependence does not hold; the nomic relation between instances of the property *horse* and instances of the symbol type 'horse' does not depend a nomic relation between instances of the property *cow* and instances of the symbol type 'horse'. Fodor describes this situation by saying that "the causal connection between cows and 'horse' tokenings is ... *asymmetrically* *dependent* upon the causal connection between horses and 'horse' tokenings."[Fodor '87, p.108]

The content of a symbol can now be fixed by the veridical tokenings and the erroneous tokenings can derive their content from the veridical tokenings on which they are asymmetrically dependent. The theory of content which Fodor proposes then makes the following claim about the content of mental symbols or mental representations:

(C) Tokens of symbol type 'A' express property *P* iff

(i) some tokens of 'A' are caused by instances of property *P* , and

(ii) tokens of 'A' that are caused by instances of a property other than *P* are asymmetrically dependent on those that are caused by instances of property *P*.

(C) deals with the Problem of Error as follows: in the the situation where an instance of property *Q*, distinct from property *P*, is misidentified and causes an 'A' tokening, the nomic relation between instances of *Q* and tokens of 'A' is asymmetrically dependent on the nomic relation between 'A' tokenings and instances of property *P* ; if instances of *P* did not cause 'A' tokenings, the instance of property *Q* would not have caused an 'A' tokening. The nomic relation between instances of the property *P* and 'A' tokenings does not, however, depend on a nomic relation between the property *Q* and 'A' tokenings. Thus, tokens of the symbol type 'A' express the property *P*, including those 'A' tokenings which result from causal connections with instances of property *Q* . In this way (C) provides an account of how the content of a symbol can be fixed in such a way that erroneous tokenings are possible.

The Problem of Disjunction is also handled by (C). In the case of erroneous tokenings, "cow-occasioned tokenings of ['horse'] do *not* express the property *[horse] or cow* since it is [horse] occasioned ['horse'] tokenings (and not [horse]-or-cow occasioned ['horse'] tokenings) on which they asymmetrically depend."[Fodor, forthcoming,p.11] Claim (C) as it is stated is intended to be general enough in its approach to the disjunction problem that it can handle the case of representational tokenings as well as erroneous tokenings. "Representational tokenings of ['horse'] express the property *horse*, and so do (mis)applications of ['horse'] to cows; both are asymmetrically dependent upon there being (actual or possible) tokenings of ['horse'] that [are caused by horses.]"[Fodor, forthcoming, p.11] The content of a symbol type is fixed by the causal relations involved in veridical labelling occasions of its tokenings. Alternative causal routes to the symbol tokenings, as found in cases of representational or erroneous tokenings, are asymmetrically dependent on the nomic relations involved in the veridical labelling tokening, and such "alternative causal route" tokenings derive their content from the veridical labelling tokenings of that symbol type.

3.3. Critique of Fodor's Theory of Content

There are two general varieties of concerns about Fodor's theory of content as characterized by claim (C). There are concerns about the adequacy of (C) to provide meanings for the entire range of terms (or more precisely, in the context of RTM, of mental representations from which linguistic symbols derive their intentional contents) that need to have contents assigned and distinguished. If (C) only applies to a limited subset of the symbols and meanings which are in fact matched or if it fails to appropriately distinguish the contents of symbols, then it cannot be a completely general theory of content. If (C) must be restricted in its application, that restriction should not be ad hoc and based soley on the need to make the theory work. The other sort of concern focusses on whether the asymmetrical dependence criterion actually provides the needed reduction of intentional semantic properties of symbols to naturalistic properties. These concerns will each be examined in turn.

3.3.1. Adequacy

Consider claim (C) and the conditions under which tokens of symbol type 'A' do not express property P . If tokens of 'A' are never caused by instances of property P , or if some non-P -caused 'A' tokenings are not asymmetrically dependent on P -caused 'A' tokenings, then the right-hand side of (C)'s bi-conditional is false, so, to maintain the truth of the whole, the left-hand side of the bi-conditional must also be false; tokens of symbol type 'A' do not express property P. The concerns about (C)'s adequacy can be categorized as problems with cases which lack the required causal connections, cases which lack asymmetrical dependence, and a third problem distinguishing necessarily coextensive properties.

3.3.1.1. Lack of Causal Connection

There are a number of examples of cases where there is no causal connection between a symbol and the property which according to independent semantic intuitions it expresses. Such is the case for the symbol 'integer' and the property *integer*. Instances of the property *integer*, like all numbers, are abstract objects and hence do not cause any tokenings, *a fortiori* they do not cause any tokenings of the symbol 'integer'. If tokens of 'integer' are never caused by instances of the property *integer*, according to (C) tokens of 'integer' cannot express the property *integer* . Clearly, this is a problem for (C), since presumably all abstract objects are similarly causally impotent, and thus excluded from providing content for symbols according to (C).

Other symbols which lack the causal connections to their content required by (C) include symbols which are never truly applied, that is, symbols which express vacuous or uninstantiated properties. For instance, tokens of a symbol like 'unicorn', are never caused by an instance of the property *unicorn* . Fodor does make an attempt to deal with vacuous symbols. Of the 'unicorn' example he says, "presumably representational tokenings of 'unicorn', and its misapplication to cows, are dependent upon counterfactual applications to unicorns. More precisely, they're dependent on the fact that 'unicorns cause 'unicorns'' is counterfactual supporting and (hence) can be true in the absence of unicorns."[Fodor, forthcoming, p.16, note 7] It is not clear, however, just what changes to (C) would be required to make use of this suggestion. Even if such an amendment were to be spelled out, no minor adjustments to (C) along these lines will handle cases of necessarily vacuous terms, as Fodor, himself, is aware. "I suppose that [such a counterfactual] treatment would require unicorns to be at least *nomologically possible* ; so Heaven only knows what a causal theory of content ought to say about *any* symbol which expresses a property that can't be instantiated..."[Fodor '87, p.164 note 5] Fodor does not make it at all clear how (C) would be modified to utilize the suggestion of incorporating counterfactuals to handle the case of contingently vacuous symbols, and such a suggestion would provide no help in applying (C) to the case of symbols which necessarily have empty extensions.

In addition to the difficulties with abstract properties and vacuous symbols, (C) runs into problems with more commonplace symbols which lack the requisite causal connections. Consider the acquisition of a symbol such a 'platypus' or 'quark', which might be introduced entirely by description. No tokens of such a symbol need be ever caused by something in its extension, but nonetheless the symbol has a content (that may be shared by those with the appropriate causal connections). Establishing the use and hence content of a symbol via description in this way is incompatible with (C), yet it is a rather common way for symbols to actually secure their meanings.

Fodor built (C) from the base case of predicative expressions, but provides no indication how other types of symbols might be incorporated into his model of a causal theory of content. In order for complex predicative expressions such as 'the man in the corner' to be brought under his model, Fodor needs to provide a naturalistic account of the contents of the symbols 'the', and 'in', which do not have any apparent causal antecedents. A naturalistic account of how the content of the complex expression can be derived from the contents of its simple components would also be required. Other non-predicative expressions (verbs, for instance) need an account explaining how they fit into the causal or at least naturalistic model of content. It is not at all apparent how Fodor's theory of content which has been built on a base of (a subset of) predicative expressions and contents could be extended to include other sorts of expressions and components of propositions and non-propositional contents of symbols.

Perhaps Fodor's theory of content could be restricted to non-abstract, non-vacuous, non-predicative simple symbols which have not been descriptively introduced, and perhaps it is possible to generalize from these cases of symbols which do not have the requisite causal connections to satisfy claim (C) in a way that is more than simply an ad hoc appeal to save the theory. Fodor, however, provides no justification for such a restriction and there is no indication that these challenges can be reconciled to a framework of a causal theory of content.

3.3.1.2. Lack of Asymmetrical Dependence

What about violations of the second clause of (C), for instance where the nomic relations yielding cow-caused 'horse' tokenings but which lack the asymmetrical dependence suggested? The cases considered in the previous sub-section which lack the causal connections to satisfy the first clause also fail to satisfy the second, since there are no tokens caused by the property in question for the other tokenings to be asymmetrically dependent upon. It is difficult to offer an exhaustive account of how (C) might be qualified to avoid running afoul of any cases which violate its second clause. Fodor does, however, deal with one obvious case which violates the (ii) but not (i). "Imagine a case," he writes,

> where someone learns 'horse' entirely from noninstances. For example, from ostensions of cows, all of which happen to look a lot like horses. No doubt, once 'horse' has been mastered, wild (cow-caused) 'horse' tokenings would depend upon tame (horse-caused) 'horse' tokenings, exactly as required. But the dependence isn't, in this case asymmetric, since the speaker's current disposition to apply 'horse' to horses is a historical consequence of his previous disposition to apply it to cows. *Had he not previously applied 'horse' to cows, he would not now apply 'horse' to horses* . [Fodor '87, p.109]

Since the dependence between cow-caused 'horse' tokenings and horse-caused 'horse' tokenings is <u>not</u> asymmetrical, cow-caused 'horse' tokenings in this example violates clause (ii) and thus, according to (C), tokens of 'horse' do not express the property *horse.*

Fodor deals with this problem by stipulating that "the sort of asymmetrical dependence that's necessary for wildness [being an erroneous tokening] is *synchronic* ; and in the case imagined, my present disposition to apply 'horse' to horses does *not* depend on any corresponding *current* disposition to apply it to cows."[Fodor '87, p.109] Historical dependencies are irrelevant to the content of a symbol if they do not currently apply. Since the current nomological relations are asymmetrically dependent, (C) fixes the correct meaning to the symbol 'horse'.

As Cummins points out,[8] this stipulation is somewhat ad hoc; Fodor offers no justification for this synchronic condition and it is mentioned only in regard to solving this problem. There is no argument put forward to support the claim that the relevant dependencies exclude considerations about the history of a symbol and particularly how it was learned or acquired. Such a constraint is only invoked to solve the problem raised by the example. Additionally, it is not at all clear from Fodor's discussion what changes in the example between the point when "someone learns 'horse' entirely from noninstances," when, presumably, there is no synchronic asymmetrical dependence which would fix the content of 'horse' as *horse,* and the point described by the phrase "once 'horse' has been mastered" where the asymmetrical dependence is synchonically established and the content *horse* fixed. Are tokenings of 'horse' as a result of instances of *cow* in the former case not in error? Do tokens of 'horse' have any content at this point? How can 'horse'-once-it-has-been-mastered be distinguished from 'horse'-that-has-been-learned? In order to adequately deal with this

---

8 See [Cummins '89, pp.56-62].

problem, Fodor needs to provide a fuller explanation of how his theory of content can handle such situations rather than simply stipulating them away.

There may be other examples of circumstances where 'A' tokenings not caused by instances of property *P* are not asymmetrically dependent on 'A' tokenings which are. The sort of causal sequences of tokenings of mental representations which RTM claims provide an account of mental processes provide a rich range of examples of non-*P*-caused 'A' tokenings which would each have to be examined for dependencies on veridical labelling 'A' tokenings. Fodor has not provided any reason to accept his tacit assumption that every tokening which is not a veridical labelling tokening is asymmetrically dependent upon such tokenings.

3.3.1.3. Distinguishing Necessarily Coextensive Properties

Another serious problem which Fodor's causal theory of content faces is that (C) is not adequate to distinguish between necessarily coextensive properties. For instance, the properties *triangle* and *trilateral,* in the context of closed geometric figures are necessarily coextensive. Any instance of the property *triangle* is an instance of the property *trilateral,* and vice versa. It is not possible to be an instance of one but not the other. Consider (C) as it applies to two necessarily coextensive properties *P* and *Q* whose instances cause tokenings of 'A' satisfying clause (i) and (ii). In such a situation, tokens of the symbol type 'A' express the property *P,* because tokens of 'A' are caused by instances of the property *P*. But, any instance of *P* is also an instance of *Q*, since *P* and *Q* are necessarily coextensive. Any nomological relation involving instances of property *P* will also apply to instances of property *Q*, since they are exactly (and necessarily) the same collection of tokens. In any nomological relation which involves instances of property *P*, property *P* can be replaced by property *Q* and the relation will still hold. Thus, tokens of the symbol type 'A' also express the property *Q* , because tokens of 'A' are caused by instances of the property *Q*. Tokens of 'A' express both the property *P* and the property *Q* ; *P* and *Q* are synonymous.

To see the seriousness of this result, consider again the example of *triangle* and *trilateral* . According to (C) tokens of the symbol 'triangle' express the property *triangle* because, by hypothesis, tokens of 'triangle' are caused by instances of the property *triangle* (and because clause (ii) of (C) is satisfied). Since any instance of the property *triangle* is an instance of the property *trilateral*, tokens of the symbol 'triangle' also express the property *trilateral* because tokens of 'triangle' are caused by instances of the property *trilateral.* Therefore, according to (C), tokens of the symbol 'triangle' express both the property *triangle* and the property *trilateral* . But 'triangle' means *having three <u>angles</u>* and not *having three <u>sides</u>* (and not a conjunction of these properties). The causal foundation of (C) is not sufficiently fine-grained to make such a distinction. Such a foundation precludes the identification of some specific feature or causal power as the source or origin of the tokening, and hence the distinguishable content.[9]

It should be clear that this problem cannot be avoided through appeal to the second clause of (C). The difficulty at hand is not a case where 'triangle' is caused by an instance of a property other than *triangle* ; an instance of *triangle* just is necessarily an instance of *trilateral.* This is not a case of an instance of some other property being mistaken for an instance of *triangle* .

These challenges to the adequacy of Fodor's theory of content that arise for examples of cases which lack the appropriate causal connections, where there is a lack of asymmetrical dependence and for necessarily coextensive properties, alone constitute a serious problem for the theory. It may be possible to restrict his theory in some non-ad hoc way to avoid some of these problems, but Fodor's project would then still be in need of a general theory of content to handle the exceptions. It is time to move on to consider an additional problem which relates specifically to the notion of asymmetrical dependence.[10]

---

9 It might be possible for Fodor to re-work the conception of causality which underlies his theory to permit a sufficiently fine-grained notion of causal power which could distinguish between necessarily co-extensive properties. As it stands, however, Fodor's theory cannot make such a distinction.

10 See [Kazmi, forthcoming] for a presentation of concerns regarding the adequacy of (C) raised in this and the penultimate sub-section.

3.3.2. Problem with Asymmetrical Dependence

Fodor's major contribution to the discussion of the Problem of Error is his suggestion that there is an asymmetrical dependence of the nomic relations involved in erroneous tokenings on the nomic relations involved in veridical labelling tokenings. If instances of property A did not cause 'A' tokenings then instances of property B would not cause 'A' tokenings, but if instances of property B did not cause 'A' tokenings, instances of property A would still cause 'A' tokenings. By considering these counterfactuals, the Problem of Error can be handled and the content of a symbol fixed as spelled out in (C). Or, so the story goes.

3.3.2.1. Semantic Dependence

It is important to carefully examine Fodor's notion of asymmetrical dependence of the causal relations involved in the relevant tokenings with an eye to ensuring that it meets his naturalistic criteria. Recall that Fodor wants "a *naturalized* theory of meaning; a theory that articulates, in nonsemantic and nonintentional terms, sufficient conditions for one bit of the world to *be about* (to express, represent, or be true of) another bit."[Fodor '87 p.98] Unless asymmetrical dependence can be spelled out in nonintentional and nonsemantic terms, it is of little use to Fodor's project.

It can be argued that Fodor has failed to show how the relation of asymmetrical dependence satisfies his naturalistic criteria. Fodor introduces this notion of asymmetrical dependence in terms of a semantic description of misidentification and erroneous symbol tokening. Recall his example: "I see a cow which, stupidly, I misidentify. I take it, say, to be a horse. So taking it causes me to effect the tokening of a symbol; viz., I say 'horse.'"[Fodor '87, p.107] Fodor notes that "misidentifying a cow as a horse wouldn't have led me to say 'horse' except that there was independently a semantic relation between 'horse' tokenings and horses ... Whereas, by contrast, since 'horse' does mean *horse* , the fact that horses cause me to say 'horse' does not depend upon there being a semantic — or, indeed, any — connection between 'horse' tokenings and cows."[Fodor '87, pp.107-108] So, erroneous symbol tokenings depend on accurate symbol tokenings in a way that accurate symbol tokenings do not depend on erroneous symbol tokenings. This asymmetrical dependence is in virtue of the semantic relation between 'horse' tokenings and the property *horse,* depends on the semantic fact that 'horse' means *horse* and is evidenced in the intentional process of (mis)identification. In the semantic mode of description, erroneous symbol tokenings, where the object of the tokening (that to which the symbol token is applied) does not match the symbol's meaning are easily distinguished from accurate symbol tokenings, where the object of the tokening does match the symbol's meaning. Fodor seems quite safe in his observation that at the semantic level of description, there is an asymmetrical dependence relation between correct and incorrect identifications, between accurate and erroneous symbol tokenings. But since this asymmetrical dependence depends on, and is given in terms of, semantic facts and descriptions, it is not terribly useful to Fodor's project of providing a naturalistic reduction of symbolic content. Given the non-naturalistic nature of the entities involved in this notion of dependence, the relation of asymmetrical dependence in the semantic mode of description of misidentification on accurate identification does not satisfy Fodor's naturalistic criteria and cannot form part of his naturalized theory of meaning.

This observation would be useful, however, if Fodor could show that a similar asymmetrical dependence relation holds between and in virtue of naturalistic entities involved in his reduction. He claims that when the above description of misidentification is cast in terms of a causal theory of content (call it a causal mode of description in contrast with the earlier semantic mode of description), "we have it that the fact that cows cause one to say 'horse' depends on the fact that horses do; but the fact that horses cause one to say 'horse' does *not* depend on the fact that cows do."[Fodor '87, p.108] It does not follow, however, that simply because the tokenings are now described in causal terms, that the dependence relation which exists between the two causal chains no longer depends on semantic notions. The asymmetrical dependence which Fodor identifies, albeit described in causal mode, still depends on the semantic fact that 'horse' means *horse* ; but for that semantic fact, there would be no reason to suppose that the causal relations posited by the causal theory of content were asymmetrically dependent.

Consider this problem from the point of view of a solely causal description of an erroneous tokening involving a symbol 'A' and two properties P and Q. A relation of asymmetrical dependence is supposed to hold between two nomological generalizations of the sort: (some) tokens of 'A' are caused by instances of property Q, and (some) tokens of 'A' are caused by instances of property P. As far as this characterization goes, it is perfectly naturalistic. But what is the nature of the supposed dependence relation? Why is one nomological generalization dependent on the other? Which generalization is dependent on the other? These questions have no apparent answers which do not involve semantic or intentional notions. Is the dependence a result of some naturalistic property or properties that are manifested by the entities involved in the nomological generalizations? Perhaps. The problem is that Fodor has offered no justification or explanation of this crucial claim. The only route that Fodor uses to establish existence of this dependence relation between the nomological generalizations is via the dependencies between semantic and intentional entities that are apparent at the semantic, and not, necessarily, in the naturalistic mode of description. It is in virtue of semantic fact that 'horse' means *horse* that the dependence relation is supposed to exist between erroneous and accurate 'horse' tokenings.

Why should it be supposed that a relation of asymmetrical dependence which holds in the semantic mode of description between generalizations that describe (and provide an account of) mistaken or erroneous symbol contents, is reflected by a similar asymmetrical dependence which holds between the nomological generalizations which are supposed to provide a different (naturalistic) mode of description for the same situation? Such a supposition assumes that the dependence at the semantic level translates or reduces to a dependence at the naturalistic level, but it is just such a reduction of semantic properties to naturalistic properties that Fodor's project is intended to supply; Fodor can assume such a reduction only at the risk of begging the very question at issue. In arguing for, and in working out the problems that arise from his causal theory of content, it is not illegitimate for Fodor to utilize his assumed reduction of content to causal relations along the way. He cannot, however, assume that all the semantic and intentional properties of tokens, the content relations between symbols, likewise reduce to causal relations without doing the work to demonstrate or spell out the reduction involved.

Unless Fodor provides an account of such supporting reductions or unless he can give an account in naturalistic terms, which does not depend on intentional or semantic features of the tokenings, of the asymmetrical naturalistic dependence of the causal relations which eventuate in erroneous symbol tokenings on the causal relations which eventuate in accurate labelling tokenings, his notion of asymmetrical dependence does not meet his own naturalistic criteria. Fodor provides no such account, so his notion of asymmetrical dependence depends on semantic notions in a way that makes it unacceptable to the naturalistic project of providing a reduction of content.

3.3.2.2. A Naturalistic Formulation

It might be argued, however, that the line of reasoning outlined above fails to distinguish between the way in which a property is established or identified and the property itself; even if semantic or intentional notions are used to establish the existence of a property, the property itself may still be perfectly naturalistic. Recall that to handle the Problem of Error, all Fodor requires is some naturalistic property which can be used to distinguish accurate tokenings, which fix a symbol's content, from erroneous tokenings, which derive their content. If some property of the nomological relations between symbol tokens and the properties that cause them can provide such a distinction, then since nomological relations are legitimate in a naturalistic theory, that distinguishing property will be legitimate in a naturalistic theory.

Consider an accurate tokening in which property *P* and symbol token 'A' are nomically related and an erroneous tokening of 'A' due to property *Q*. The property which distinguishes the accurate from erroneous tokening can be described by (a conjunction of) counterfactual statements which involve no non-naturalistic terms:

(6) If instances of property *P* did not cause 'A' tokenings then instances of property *Q* would not cause 'A' tokenings, but if instances of property *Q* did not cause 'A' tokenings then instances of property *P* would still cause 'A' tokenings.

Fodor suggests a possible world analysis of counterfactuals[11] which would provide conditions under which the distinguishing property is satisfied and(6) is true; conditions which do not refer to the relation of asymmetrical dependence or other (potentially) non-naturalistic terms. Such an analysis defines possible worlds which generally share common physical laws as "near" each other, as opposed to "distant" possible worlds where some of the natural laws which hold in one world fail to hold in the other. A counterfactual is true, in a world w_i, then, if and only if in possible worlds near to w_i where the antecedent of the counterfactual is true, the consequent is also true.

The conditions under which the counterfactual property identified by (6) is satisfied can then be utilized to reformulate claim (C) so that it does not refer to asymmetrical dependence:

> (C') Tokens of symbol type 'A' express property *P* in a world w_i, iff
>
> (i) some tokens of 'A' are caused by instances of property *P* in w_i, and
>
> (ii) tokens of 'A' are <u>not</u> caused by instances of a property other than *P* in worlds near to w_i in which instances of property *P* do <u>not</u> cause 'A' tokenings, and

---

11 See [Lewis '73].

(iii) tokens of 'A' are caused by instances of property *P* in worlds near to w_i where instances of a property other than *P* do <u>not</u> cause 'A' tokenings.[12]

Consider how (C') handles the 'horse', *horse* and *cow* example. Tokens of 'horse' express the property *horse* in this world just in case three conditions are satisfied. First, some tokens of 'horse' must be caused by instances of *horse* in this world. Secondly, in nearby possible worlds in which instances of *horse* do not cause tokens of 'horse' (perhaps, horses are called something else or there are no horses) no instances of other properties such as *cow* cause tokens of 'horse'. If 'horse' means *horse,* a break in the causal link between instances of *horse* and tokens of 'horse' breaks any causal link between instances of *cow* and tokens of 'horse'. Thirdly, in nearby possible worlds where instances of other properties such as *cow* fail to cause tokens of 'horse', instances of *horse* still cause tokens of 'horse'. If 'horse' means *horse* , a break in the causal link between instances of *cow* and tokens of 'horse' will not affect the link between instances of *horse* and tokens of 'horse'.

Tokens of the symbol 'horse' express the property *horse* in the actual world if and only if the three conditions that make up the right-hand side of the biconditional are true. The reduction of semantic and intentional terms such as meaning or, in this case,

---

12 These conditions generalize a characterization made by Fodor; see [Fodor '87, p.109].

expression, can then be carried out by translating (replacing) talk of the symbol 'horse' expressing or meaning the property *horse* with a conjunction of the three naturalistic conditions on the right-hand side of (C')s bi-conditional, substituting appropriately for 'horse' and *horse* . Such a translation will be truth preserving as a result of the bi-conditional and will contain no semantic terms. In this way Fodor's causal theory can provide a naturalistic account of content which allows the content of a symbol to be fixed and permits erroneous tokenings.

3.3.2.3. Semantic Dependence Again

There are a couple of problems, however, with this attempt to spell out Fodor's property of asymmetrical dependence in naturalistic terms. It is not clear that this approach does indeed avoid a dependence on semantic notions. Even if this asymmetrical dependence approach did provide a naturalistic reduction of symbolic content, the sort of reduction that it would provide is of limited practical value.

The question at issue is whether (C') is truly independent of semantic or other non-naturalistic considerations. Consider how (C') applies to the following example: It is observed that tokens of symbol type 'A' are caused by instances of two distinct properties: *P* and *Q* .[13] Therefore condition (i) is satisfied for both properties in the

---

[13] For the purposes of this example assume that "A causes B" amounts to "the presence of A is a reliable predictor of the presence of B" or "A

actual world. What about condition (ii)? In nearby worlds in which instances of P do not cause tokens of 'A', do instances of Q cause tokens of 'A'? The answer to this question requires and depends on a prior determination of whether 'A' means P or not. If 'A' does express the property P then it would seem reasonable to suppose that in possible worlds where the causal link between instances of P and tokens of 'A' is broken the causal link between instances of Q and tokens of 'A' would also be broken. On the other hand, if 'A' expresses the property Q, breaking the causal link between P and 'A' should have no effect on the causal link between Q and 'A'. Without that prior semantic assignment, there appears to be no reason to suppose that in those nearby worlds in which tokens of 'A' are not caused by instances of the property P, that instances of property Q also fail to cause 'A' tokenings.

The possible world analysis of the counterfactual property that is supposed to distinguish accurate from erroneous tokenings cannot get off the ground unless a semantic assignment of meaning is assumed Only when the intuition that 'A' means or expresses the property P is added to the assessment of nearby possible worlds, does (ii) turn out to have a determinate truth value of TRUE. Similarly for (iii); unless it is already assumed that 'A' expresses the property P there appears to be no reason to suppose that in nearby

---

reliably covaries with B" or some other similar notion of empirical observation which would justify the postulation of the relevant nomological law.

worlds in which instances of property *Q* do not cause 'A' tokenings, instances of property *P* do cause 'A' tokenings. Claim (C') depends on a semantic assignment of content and requires that such an assignment be carried out prior to evaluating the naturalistic conditions which do the job for which (C') was designed: distinguishing accurate from erroneous tokenings. Unless the very notion of symbolic content assignment for which the claim is intended to provide a reduction is assumed, (C') cannot distinguish between accurate and erroneous tokenings. If it cannot make that distinction, it cannot fix the content of the symbol. This dependence on semantic facts of content assignment cannot be avoided simply by recasting asymmetrical dependence in terms of possible worlds. Claim (C') hardly looks like a favourable candidate for reduction of meaning which does not depend on semantic or other non-naturalistic considerations.

Even if (C') or some variation on its theme could be shown to truly independent of semantical or other non-intentional considerations, the sort of reduction which it would provide is of limited practical value. It would be possible to translate talk of the symbol 'horse' expressing or meaning the property *horse* with a conjunction of the appropriate conditions involving claims about the persistence of causal relations in various possible world scenarios and to preserve truth in the translation. Such a translation would be possible because the translator presumably already knows that 'horse' does express the property *horse* and thus the reduced translation which includes the corresponding condition (i), (ii) and (iii) is true just in case the original, non-naturalistic statement is true.

What this reduction will not do, however, is provide any clue about symbolic content when the prior semantic assignment is not available. To see this, consider again the example used above. Suppose it is observed that tokens of symbol type 'A' are caused by (reliably covary with) instances of two distinct properties: *P* and *Q*. There is no pre-established connection between the form of the symbol type and the labels used for the properties involved (as has been inherent in Fodor's use of *horse* to name the property expressed by the symbol type 'horse'). This feature of the example corresponds to the situation where the symbols involved are the sort of mental representations or physical brain states that RTM claims have intentional contents. In such a situation, there is no pre-established clues to semantic assignment based on similarities of form between the symbol and property which it expresses. Claim (C') (and (C), for that matter) are of no use in fixing the content of the symbol 'A' in such a situation. As was seen above, there is no more reason to suppose that in possible worlds in which 'A' tokens do not covary with instances of *P* , 'A' tokens also fail to covary with instances of *Q* , rather than vice versa.

Fodor's property of asymmetrical dependence and its characterization in (C') is only useful in situations where the semantic assignment which is to be reduced is known in advance. It does not help fix the content of a symbol from naturalistic first principles, it might be said. In practice, the best that (C') can provide is a means of translating the semantic left-hand side of its bi-conditional into an (allegedly) naturalistic right-hand side equivalent. It provides little help when going from a naturalistic mode of description corresponding to the right-hand side to fix a symbolic content as described by its left-hand side, unless the result of that semantic assignment is known in advance.

Fodor's causal theory of content faces serious challenges to its adequacy as a general semantic theory. Examples have been considered where the theory breaks down for symbols which lack appropriate causal connections to their contents, tokenings which lack the required asymmetrical dependence and when dealing with necessarily coextensive properties. If it is to work at all, a non-ad hoc restriction which avoids these conditions is required. Concerns have also been raised about the naturalistic footing of his asymmetrical dependence property which is intended to distinguish content fixing tokenings from tokenings which derive their content from other tokenings. Even if Fodor's asymmetrical dependence property could be given a sound naturalistic foundation, since his theory requires that for any reduction of symbolic content, an assignment of that content to the symbol be made prior to carrying out the reduction, its success and practical value is questionable.

Concluding Remarks

The Problem of Error remains in need of a satisfactory solution. It has been argued that the leading information-based theories of content proposed by Dretske and Fodor have, each in their own way, fallen short of the goal of providing a naturalized theory of content which solves the Problem of Error.

Dretske's model is a fairly straight-forward extension to the foundations of information theory as laid out by Shannon, incorporating a notion of informational content. This addition moves information theory beyond a strictly quantitative analysis of information. While he relativizes information to the knowledge of a receiver, at the heart of Dretske's conception of informational content is the idea that a signal or state carries information about what must be the case when that signal or state is present. This central point is both the source of the Problem of Error as it arises for Dretske's information-theoretic model of content and downfall of the solution he proposes.

Information, in the sense embodied by Dretske's model, is, by definition, true. There can be more of it or less of it, but it makes no sense to speak of information, in this technical sense, as being false. Beliefs and their associated contents are, on the other hand, frequently false. To solve this version of the Problem of Error, Dretske must show how the content of a belief can be based on information bearing structures but has the added feature of possibly being false.

Dretske attempts to give an account of false content by defining a different sort of content, a semantic content, which has its origins in informational content. A belief structure type is said to develop a semantic content during a learning period where tokens of the structure type are produced in response the some piece of information that *s* is *F*. The semantic content of the belief structure type is the most specific piece of information carried by tokens of that structure type which occurred during the learning period, supposedly, the information that *s* is *F*. Once the learning period is over, tokens of that same belief structure type can acquire the semantic content which is associated with the structure type, even if they do not themselves carry the information that *s* is *F*, even if *s* is not *F*. This is the way in which, on Dretske's model, some tokens of that belief type can have the semantic content that *s* is *F*, and have that content falsely.

The problems with this approach are quite straight-forward. There is no indication how tokens which are supposed to fix the semantic content of the structure type during the learning period are to be distinguished from later tokens which may be false. Dretske provides no suggestion as to how the learning period might be delimited from the rest of a structure type's occurrences.

More importantly, this account of false content conflicts with the model Dretske has set up for the carrying of information. A structure can carry the information that *s* is *F* only if it is not possible for tokens of that structure type to occur when *s* is not *F*. The account of false belief requires that some tokens of the belief structure type occur when *s* is not *F*, so tokens of that structure type cannot carry the information that *s* is *F*. If tokens of that structure type cannot carry the information that *s* is *F* during the learning period (assuming that such a period could be delimited) that piece of information cannot be the most specific piece of information the structure carries and hence identified as the semantic content of the type. This approach to accommodating false beliefs is not compatible with Dretske's formulation of informational content.

Fodor tries a different tack, framing his theory of content within the context of his Representational Theory of Mind. He wants to capture the intuition that symbols are about what causes them. Since symbols often are mistakenly tokened in response to things other that their intuitive content, the Problem of Error arises. Why do symbols not have everything that might cause their tokenings included in their content?

Fodor's proposal is to distinguish between the tokenings that fix the content of a symbol and others that derive their content from those content fixing tokenings. His suggests that there is an asymmetrical dependence of the content deriving tokenings of symbols on the content fixing tokenings tokenings. The mistaken tokening of a symbol "horse" in response to a cow is dependent on tokenings of "horse" in response to horses, but the reverse dependence does not hold.

There are a number of problems relating to whether Fodor's theory of content is adequate to deal with a full range of the cases expected to be handled by such a theory. His theory provides no account of the content for symbols which lack the requisite causal connections such as abstract, vacuous, descriptively introduced, and non-predicative symbols. It is not at all clear that such sorts of symbols could be accommodated by a causal analysis of content. Historical dependencies between symbols and their alternative causes appear to be another potential problem which Fodor deals with only by stipulating them away. His causal analysis is also not sufficiently fine-grained to distinguish between necessarily coextensive properties, yielding false results such as the synonymy of terms like "triangle" and "trilateral". In addition, there are concerns regarding whether the asymmetrical dependence relation can be spelled out avoiding any direct or indirect appeals to semantic or intentional notions, such as a pre-established meaning for the symbol in question, as would be required for a naturalistic theory of content. Without a fully naturalistic solution to the Problem of Error, Fodor's project cannot be deemed successful, even if the challenges to his theory's adequacy were met.

Without a solution to the Problem of Error, information-based semantics cannot be judged to have succeeded in providing a model of the sort of content required by cognitive science. Until such a model is developed, cognitive science will lack a necessary insight into the qualitative aspects of the structures it describes. It may be able to describe and understand to some extent the form involved in the processes and structures of cognition, but it will lack an understanding of the content which brings that form to life.

While it has been argued that the theories of content which Dretske and Fodor have proposed and their different approaches to dealing with the Problem of Error have not been successful, it should be apparent that this by no means shows that the problem is intractable. Each time a proposal is tried and fails it adds to the understanding of the complexities involved in symbolic content and its relation to error. It is clear, however, that a full understanding of the nature of meaning requires first an understanding of how to make a mistake.

Bibliography

Baker, Lynne Rudder. "What is a Mental Representation." Unpublished manuscript. Middlebury College, 1989.

Cummins, Robert.*Meaning and Mental Representation.* Cambridge, Mass.: MIT Press,1989.

Dennett, Daniel C. *The Intentional Stance.* Cambridge, Mass.: Harvard University Press,1987.

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

Fodor, Jerry A. *Psychosemantics.* Cambridge, Mass.: MIT Press,1987.

_______________ "Information and Representation." In *Information, Language, and Cognition.* Ed. Philip Hanson. Vancouver Studies in Cognitive Science. Vancouver, B.C.: UBC Press, forthcoming.

Kazmi, Ali Akhtar. "Remarks on Information and Representation." In *Information, Language, and Cognition.* Ed. Philip Hanson. Vancouver Studies in Cognitive Science. Vancouver, B.C.: UBC Press, forthcoming.

Lewis, David K. *Counterfactuals.* Cambridge, Mass.: Harvard University Press,1973.

Martin, C.B. "A New View of the Mind." Unpublished manuscript. University of Calgary, 1989.

Pylyshyn, Zenon W. *Computation and Cognition*. Cambridge, Mass.: Harvard University Press,1984.

Shannon, C.E. "The Mathematical Theory of Communication." In *The Mathematical Theory of Communication.*, by C.E. Shannon and W. Weaver. Urbana, Ill.: Univ. of Illinois Press, 1949.

Soames, Scott. "Belief and Mental Representation." In *Information, Language, and Cognition*. Ed. Philip Hanson. Vancouver Studies in Cognitive Science. Vancouver, B.C.: UBC Press, forthcoming.