A NEW THEORY OF CONTENT

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF SOCIAL SCIENCES

OF

MIDDLE EAST TECHNICAL UNIVERSITY

BY

TEVFİK AYTEKİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF PHILOSOPHY

SEPTEMBER 2003

Approval of the Graduate School of Social Sciences

_____

Prof. Dr. Bahattin Akşit
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. Ahmet İnam
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assoc. Prof. Dr. Erdinç Sayan
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Erdinç Sayan _____

Assist. Prof. Dr. John Bolender _____

Assist. Prof. Dr. Bilge Say _____

ABSTRACT

A NEW THEORY OF CONTENT

Aytekin, Tevfik

M.S., Department of Philosophy

Supervisor: Assoc. Prof. Erdinç Sayan

September 2003, 52 pages

Naturalistic philosophers of meaning try to define the recalcitrant concept of reference in terms respected by the empirical science, such as causality or teleology. In this thesis, after a brief introduction to these trials is given, Fodors theory of content in terms of asymmetric dependence is examined in some depth. I claim that although this theory involves an important insight, it is an unsatisfactory attempt at reduction of the notion of reference. I develop a new theory of content, which does not have the defects of the analyses in terms of asymmetric dependence, and more successfully deals with notorious cases, such as pansemanticism and the possibility of misrepresentation.

Keywords: Meaning, Naturalism, Reference, Asymmetric Dependence.

ÖZ

YENİ BİR İÇERİK TEORİSİ

Aytekin, Tevfik

Yüksek Lisans, Felsefe Bölümü

Tez Yoneticisi: Doç. Dr. Erdinç Sayan

Eylül 2003, 52 sayfa

Doğalcı anlam felsefecileri çözüme direnç gösteren gönderim kavramını nedensellik veya ereksellik gibi empirik bilimlerce saygın kavramlarla tanımlamaya çalışırlar. Bu tezde bu çalışmalar kısaca tanıtıldıktan sonra, Fodor'un asimetrik bağımlılık kavramına dayanan içerik teorisi belli bir derinlikte incelenmiştir. Asimetrik bağımlılık teorisinin önemli sezişleri olmakla birlikte, gönderim kavramının indirgenmesinde yetersiz kaldığı tezde iddia edilmiştir. Asimetrik bağımlılık analizlerinin sorunlarına sahip olmayan ve tümanlamlıcılık veya yanlış-temsil gibi, çetin olmalarıyla bilinen konularla daha başarılı bir şekilde başa çıkan yeni bir içerik teorisi geliştirilmiştir.

Anahtar Kelimeler: Anlam, Doğalcılık, Gönderim, Asimetrik Bağımlılık

To My Family

ACKNOWLEDGEMENTS

I would like to thank my supervisor Erdinç Sayan for his guidance and support throughout my thesis study. I am especially grateful to him for the extended discussions in our meetings. His comments and criticisms during these meetings were invaluable not only for this thesis but also for my entire philosophical development.

I would also like to thank my thesis committee members, John Bolender and Bilge Say, for their helpful comments and criticisms.

Finally I would like to thank Ayşenur Bay, Ayşegül Özkul, and Emre Akbaş for their brilliant intuitions which helped me to solve some hard problems.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

CHAPTER I

INTRODUCTION

This work is centered on the notion of meaning. The sequence of letters ‘Aristotle’ means, at least to philosophers, something, but the sequence of letters ‘Tunglig’ does not mean anything as you might accept. It is perfectly reasonable to ask in virtue of what ‘Aristotle’ has a meaning while ‘Tunglig’ does not? Even if one can answer this question one might still wonder what meaning is. When we assert that something has a meaning we implicitly assume that we know what meaning is. But a moment’s consideration will reveal that it is extremely hard to state precisely what we assume to know. As one would expect the literature on meaning is vast and sophisticated but we will not deal with all those issues. So let me try to locate our problem at the very beginning.

One primary line of view, beginning with Frege, claims that meaning has two components: sense and reference. What reference is relatively easier to grasp. The reference of the word ‘Aristotle’ is the last of the three great philosophers who had lived in ancient Greek times. The reference of ‘the morning star’ is the thing that we see in the mornings in the north direction. It is claimed that to understand the meaning of something is to grasp what it refers to. But it was Frege who first saw the problem with this claim. He said that both ‘the morning star’ and ‘the evening star’

refer to the same thing, however their meanings are not the same. So meaning cannot be exhausted by reference alone, i.e, meaning is more than reference.

Frege called the part of meaning of an expression that is not covered by reference the sense of that expression. Accordingly, although both 'the morning star' and 'the evening star' have the same reference they don't have the same meaning, since they have different senses. However what exactly the term 'sense' means is hard to state and there is no consensus on this matter. But what is generally accepted, independently of the discussions on the notion of sense, is that reference constitutes the core of meaning. Our primary task in this thesis is to give a naturalistic analysis of reference. We will try to give *sufficient* conditions for reference and see whether it satisfies our intuitive notion of reference that we have and if it does not, i.e., if there are counter examples, we will revise our conditions. Note that our task is to give only the sufficient conditions for reference, not the necessary ones. Giving necessary and sufficient conditions is much harder and to give sufficient conditions has proved to be hard enough. However even if we manage to give only sufficient conditions it will be a great step towards understanding the notion of reference. One crucial rule that we shall adopt is that our analysis should be naturalistic. What this requirement of naturalism amounts to is that we should not use any intentional or semantic notions, such as "believes that", "is about", "is true", "represents," etc. Note that this requirement is different from the general requirement that no definition should use in its definiens what it tries to define; that is, the definition shouldn't be circular. Explications of meaning can use both intentional and semantic notions, as in the work

of Grice (1957), and still be illuminating. The naturalism requirement allows us to use only notions which are used in contemporary scientific fields like physics, biology, chemistry, etc. Philosophical analyses that are not naturalistic can be illuminative and improve our understanding of the subject but naturalistic analyses have the additional virtue of allowing physical implementations. In this respect naturalistic analyses are of utmost importance for fields like artificial intelligence and cognitive science.

Now what we want to do is clear but why we want to do it might be less clear. So let us give our motivations for giving a naturalistic account of reference. I will talk about two such motivations: The first is meeting the challenge of Brentano and the second is related with Fodor's Representational Theory of Mind. Now we will look at each in turn.

1.1 Intentionality

The challenge of Brentano which is also known as Brentano's Thesis is summarized in the following quotations:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something in it as an object . . . (Brentano, 1995, p.88).
>
> No physical phenomenon exhibits anything similar (Brentano, 1995, p.97).

Brentano talks about two important features of the mind: Its ability to refer to a content and its ability to refer to an inexistent object. For example the sentence "Tom believes that ghosts exist" attributes to Tom's mind a direction toward something that does not exist. Although the capacity to refer to nonexistent objects is an important feature of the mind, what we shall be primarily interested in is the more fundamental capacity of our minds to refer to a content whether existent or not. Although Brentano did not use the term, this property of our minds is called "intentionality" in contemporary philosophy of mind.

The other main claim of Brentano is that intentionality is the demarcating property of minds that no physical thing can have this property. When you look at the nature, you see rocks, trees, dogs, mountains, etc., but none of these are about anything. That is, a rock is a rock and is not about anything. But a moment's thought will reveal that in fact there are many physical things that are about other things. For example, signs, sentences, maps, books are physical things and indeed they are about other things. A sign might be about a traffic rule, a sentence might be about a person, and a map might be about a city. The proponents of intentionality as the mark of the mental hold that the intentonality of these latter things come from the concerns of humans who create them; in other words they have *derived* intentionality. For those philosophers minds' intentionality is not derived but it is *intrinsic* or *original*. The key claim here is that original intentionality is independent of any interpretation. In other words there is something intrinsic in mental states that links those states to other things or fixes their reference.

As you can see this is a great challenge to any philosopher who has a materialistic standpoint. It is also a great challenge to the artificial intelligence researchers since Brentano claims that no matter how "smart" the machines they build, since they will be physical things they won't possess intentionality, and so the project of artificial intelligence is doomed to fail. However if you can manage to show that physical can things have original intentionality, that will be an answer to Brentano's challenge. Many philosophers like Fodor, Dretske, Millikan, and others have tried to meet Brentano's challenge in this way. We will see whether they have been successful in the following section.

One philosopher, John Searle, has a unique position in that although he has a materialistic attitude, he believes that original intentionality can not be achieved by the methods of artificial intelligence, especially by symbol manipulation.

> "Could a machine think?" My own view is that *only* a machine could think, and indeed only very special kinds of machines, namely brains and machines that has the same causal powers as brains. And that is the main reason strong AI has had little to tell us about thinking, since it has nothing to tell us about machines. By its own definition, it is about programs, and programs are not machines. Whatever else intentionality is, it is a biological phenomenon, and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena. No one would suppose that we could produce milk and sugar by running a computer simulation of the formal sequences in lactation and photosynthesis, but where the mind is concerned many people are willing to believe in such a miracle because of a deep and abiding dualism: the mind they suppose is a matter of formal processes and is independent of quite specific material causes in the way that milk and sugar are not (Searle, 1980, p.456).

Although Searle's arguments against strong AI are powerful he has given us no analysis of intentionality by just referring to the causal powers of the brain. What are

the characteristics of these causal powers? To have intentionality do we have to duplicate a brain molecule by molecule? Can't there be machines that have intentionality but not as sophisticated as our brains? Yes it is very probable that the causal powers of our brains produce intentionality but we need to find out the nature of these causal powers. This is another motivation what we need a naturalistic account of intentionality.

Some terminological remarks are appropriate at this point. In this thesis you will often come across with the terms 'meaning,' 'intentionality,' 'reference,' 'representation,' and 'aboutness.' I assume that meaning and intentionality are both more than reference. And as I stated before my primary purpose is to give an account of reference, not meaning or intentionality. But since I assume that reference is the main component of both meaning and intentionality, giving an account of reference would be a major achievement if it can be accomplished. Also, I will use the words 'reference,' 'representation,' 'content,' and 'aboutness' interchangeably, i.e., I assume that they amount to the same thing.

1.2 Representational Theory of Mind (RTM)

Fodor's RTM is the general framework which he had developed as a realist explanation of folk psychology. Folk psychology can be defined roughly as the set of statements ordinary people use to describe, explain, and predict human behavior. These statements contain familiar terms like "belief," "desire," "pain," "anger," etc. Usually each statement relates external stimulus with mental states (e.g., "The

challenging words of her rival increased her desire to win"), mental states with other mental states (e.g., "He relaxed after realizing the trick in the question"), or mental states with behavior (e.g., "Becoming angry she shouted at her"). We use folk psychology in our everyday lives to explain and predict other people's behavior. The crucial question is whether folk psychology is a true theory of human behavior. Of course by a true theory I do not mean that folk psychology needs to have the same status as other scientific disciplines such as physics, chemistry, etc. For example folk agriculture or folk geology are not genuine scientific theories but they are not totally false as they have been vindicated by scientific agriculture and geology. However some of the theories used by the folk in the past, instead of being vindicated, had been eliminated. So the crucial question is whether folk psychology will be vindicated or eliminated by the future science. There are at least three positions regarding the fate of folk psychology. For the eliminativist, folk psychological relations are false and its ontology is an illusion (Churchland, 1981). For the instrumentalist, folk psychology works and hence it is a kind of theory capable of producing predictions, but it is totally nonsense to anticipate any similarity between the inner structure of the brain and the structure of folk psychological theory (Dennett, 1975). The third position is known as "Representational Theory of Mind" which asserts that folk psychology will be vindicated by scientific psychology and has a realist attitude towards its ontology.

According to Fodor (1987) RTM consists of two claims:

> Claim 1 (The nature of propositional attitudes):
> For any organism O, and any 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. (p. 17)
>
> Claim 2 (the nature of mental processes):
> Mental processes are causal sequences of tokenings of mental representations. (p. 17)

To understand what Claim 1 offers and especially to understand the nature of the relation R, we should distinguish two aspects of propositional attitudes. Every propositional attitude has two essential parts: *state* and *content.* States are characterized by the words belief, desire, etc. Contents are what the complement clause in propositional attitudes refers to. For example in the sentence "She believes that the war will end", the state of the propositional attitude is "believing" and the content is "the war will end." Claim 1 asserts that the contents of propositional attitudes are represented in the brain in some form and R is the relation of this mental representation to the organism (or brain). So what distinguishes a belief from a desire when both of them have the same content is the relation of their content to the organism. The functional role played by the same content is different in believing versus desiring.

The second claim is proposed to explain how (rational) thinking is possible mechanically as follows. According to Fodor, the explanation of thinking was made

possible by two great achievements in the twentieth century, namely, the development of symbolic logic and Turing machines. Development of symbolic logic showed us that semantic aspects of reasoning could be captured by syntax alone and what syntax amounts to is nothing but the shape of the symbols used in the formalization. The development of Turing machines (effective procedures), on the other hand, made it possible to carry out the formalism mechanically. According to Fodor computation is just causation that preserves semantic values. According to RTM mental representations are symbols and thinking is just those causal relations in the brain which are truth preserving in the sense of formal logic.

The reader will probably notice one important gap in the foregoing account of RTM. RTM does not explain how mental representations get their contents. This is exactly the same problem of reference that we considered in discussing intentionality. For RTM to be a complete theory of folk psychology and thinking in general, it must solve the problem of representation. This is the second motivation for looking for a naturalistic account of reference/representation.

CHAPTER II

MEANING NATURALIZED

In this section we will look at recent attempts to naturalize meaning. Before we start let me point out that what naturalism maintains is not very clear to state.[1] Naturalism in general is the view that whatever exists has a place in natural order and can be explained by the empirical sciences. Any notion employed by any *respectable* empirical science is also assumed to be respectable as a naturalistic notion. Since semantic predicates are not used by the empirical sciences they are not counted as naturalistic. As we will see the two notions that are heavily used by naturalistic philosophers of meaning are *laws* and *causation*. But note that although contemporary philosophers generally assume that laws and causation are respectable notions, from a naturalistic point of view, this is not without controversy. That is, it is not implausible to argue that the notions of law and causation themselves involve semantic notions. If this is the case, a successful explanation of reference in terms of laws and causation would amount to a reduction of meaning which is not a naturalistic one, but such a reduction could have a great value.

[1] In fact naturalism very much overlaps with materialism and physicalism. In whatever ways they are different, if they really are, do not concern us here.

The naturalistic theories that we will discuss in what follows try to explain what it is for a *mental representation* to have content. Discussions of in virtue of what something has a reference have turned into the question of in virtue of what a mental representation has a content. These two questions have no fundamental difference. The discussions in the literature are centered on mental representations because it is generally thought that it is the right place to look for original intentionality. The roots meaning must be somewhere in the brain and mental representations are promising candidates. It is supposed that if we can explain how mental representations have content, then we can explain how other things, like signs, words, etc., have the content they do.

2.1 The Similarity Account

The oldest account of meaning employs the notion of similarity but today nobody is an advocate of it. I include this account merely for illustrative purposes. According to the similarity view a mental representation has content in virtue of its similarity to what it represents.

Fodor (1990c) gives two arguments against the similarity account of representation. Representation is an asymmetric relation but similarity is a symmetric relation; so similarity can't be representation. The word 'Aristotle' represents a certain philosopher but that philosopher does not represent 'Aristotle.' Whereas whenever something X is similar to Y it follows that Y is also similar to X.

Fodor's second argument is through the singularity of representation. While the concept of *tiger* represents *all tigers*; the concept *this tiger* represents only this particular one. For example, think of the *clone* of an individual tiger which is a molecule-by-molecule duplicate of the other one. How can you have two representations one of which represents the clone and the other represents the real one? Each of the representations would have to represent either one of the tigers since both representations are equally similar to what they represent, viz. the two identical tigers.

Another problem with the similarity account is that it makes *misrepresentation* impossible. As rightly pointed out by Cummins (1989, p.27), "The similarity view seems to allow for misrepresentation only when the dissimilarity is relatively small." For example a representation of cats is a misrepresentation when its dissimilarity with real cats is small relative to other representations which do not represent cats. But it is still not the representation of cats since it is not similar enough to cats. It seems hard to have a workable notion of "relatively small," however the difficulty is a real one. How can a representation be about something but still be false or mistaken? Misrepresentation is a crucial notion for all naturalistic theories of representation to capture. It is widely accepted that a correct theory of representation has to solve the problem of the possibility of misrepresentation. As we shall see, misrepresentation turns out to be a really tough problem, and especially the theories that rely on the notion of causation (including the theory proposed in this thesis) will have much to say about it.

There are other problems with the similarity account mentioned by Cummins (1989). One is that it is incompatible with physicalism and the other is that it is unable to deal with abstraction. But enough is said about this account. Let's proceed with the other theories.

2.1 Conceptual Role Theories

The conceptual role theories, also known as the causal role or inferential role theories, are a species of internalist theories which seek to explain the content of a mental representation by looking at the causal or inferential relations between that mental representation and the others. One of its prominent supporters explain what the conceptual role theory is in the following words:

> The doctrine has its roots in positivism and pragmatism and in the Wittgensteinian idea of meaning is use.… A crucial component of a sentence's conceptual role is a matter of how it participates in inductive and deductive inferences.… Conceptual role is total causal role, abstractly described.… Conceptual role abstracts away from all causal relations except the ones that mediate inferences, inductive or deductive, decision making, and the like (Block, 1994, pp. 93, 94).

Fodor and LePore (1994) have two objections to conceptual role theories. One is that conceptual roles do not seem to be compositional and the other is that they lead to semantic holism. Now let us look at each in turn.

Compositionality refers to the fact that the meaning of a sentence is a function of the meaning of the words in it together with its syntax. According to Fodor and LePore (1994) the two non-negotiable features of language, namely productivity and

systematicity, cannot be explained without assuming compositionality. Productivity is the fact that natural languages can potentially produce an infinite number of expressions, and systematicity is the fact that if a language can express an expression P then it can also express an expression Q, if Q is semantically close to P. For example, if a sentence can express aRb then it can also express bRa. Fodor & Lepore claim that conceptual roles are not compositional and they show this by giving an example:

> Consider the meaning of the phrase "brown cow"; it depends on the meanings of "brown" and "cow" together with its syntax, just as compositionality requires.… But now, *prima facie*, the inferential role of "brown cow" depends not only on the inferential role of "brown" and the inferential role of "cow," *but also on what you happen to believe about cows.* So, unlike meaning, inferential role is, in the general case, *not* compositional (Fodor and Lepore, 1994).

On the other hand the concept of conceptual role leads to meaning holism, which indeed leads to undesirable results. If meaning is determined by the totality of inferential roles, i.e., by the so called conceptual networks, then no two persons can ever mean the same thing by an expression since any two persons will have different conceptual networks, unless they are exact duplicates. Any local change in the network should affect all the meanings of the expressions in the network. I think it is clear that this kind of holism is better avoided

.

2.2 The Crude Causal Theory and The Problem of Misrepresentation

One intuitively plausible candidate for a naturalistic relation between a symbol and its referent is causation. Causal theories, in general, try to explain meaning by relying on properties of causation. According to the most basic form of causal theories, known as the crude causal account, a symbol S represents F if F lawfully causes S. For example, if flies cause a certain neural activation in the frog's brain whenever they pass by, then, according to the crude causal account that particular neural activation in the frog's brain represents flies.

Although causal accounts of representation are appealing approaches, they nevertheless pose a major difficulty known as the disjunction problem (Fodor, 1990a). Suppose that horses cause "horse" tokens in (human) brains, then, according to the causal account "horse" tokens represent horses. However, a cow-on-a-dark-night might also cause "horse" mental tokens and indeed it might be the case that the causation relation between "horse" mental tokens and cows-on-dark-night is not a coincidence but a systematic causal relation, i.e., a causal law. Therefore the crude causal theory has to admit that "horse" tokens represent not horses but horses *or* cows-on-dark-night. But surely we want "horse" tokens to represent only horses and not horses *or* cows-on-dark-night.

The disjunction problem, in fact, overlaps with another problem: the problem of misrepresentation. It is widely assumed that any theory of meaning should account for misrepresentation, i.e., should answer the question, "How can a representation

how a representation be falsely tokened?" The crude causal theory is unable to solve the disjunction problem; it is not possible for the crude causal theory to exclude some of the causes from the extension of a representation, as it should. In order for misrepresentation to be possible, a causal theory of representation needs to find a principled way to exclude semantically irrelevant causes from the extension of a representation.

Both the problem of misrepresentation and the disjunction problem are related to a prominent characteristic of meaning. This characteristic is what Grice (1957), about a half century ago, tried to explain by making the distinction between natural and non-natural meaning. Fodor stresses the same point by introducing another name, "robustness of meaning:" "In actual fact, "cow" tokens get caused in *all sorts* of ways, and they all mean *cow* for all that." (Fodor, 1990b, p.90) According to Fodor, misrepresentation is only one aspect of robustness, that is a representation can get caused even though there is no misrepresentation and the cause is semantically irrelevant. For example, in certain situations milk causes "cow" tokens.

Different proposals have been offered to solve the disjunction problem. Many of them try to differentiate in a principled way between two types of situations. The first type of situations are supposed to be those in which all the causes of a symbol are meaning forming, and the second type of situations are supposed to be those in which the causes should not be included in the extension of the symbol.

An idea naturally comes to mind is to assert that the first type of situation is one where there are "normal" or "optimal" conditions. Misrepresentation usually occurs under non-optimal conditions. A cow can cause a "horse" token only on a dark night. Under plenty of sunlight a cow nearby always causes "cow" tokens in a observer's brain. But it proves very hard to define what are to be taken as "normal" conditions. For example, should we include the condition "under plenty of sunlight" in our optimal conditions list? If the answer is yes, then what should we do with fireflies? The optimal conditions for fireflies to cause "firefly" tokens should not include presence of sunlight but instead presence of darkness.

An early theory proposed by Dretske (1981), which he later abandoned, proposes that symbol (or mental token) "S" represents P if "S" is tokened by P during the "learning period." Dretske's insight was that a symbol gets its meaning during the learning period of an organism and is fixed at the end of that period. The first difficulty with the learning-period theory is that it is not possible to define a learning period, a period of time in which we learn a concept fully. If we accept that our concepts change then how can we differentiate the causes that change our concepts and causes which are occasional misrepresentations? Fodor (1990c) points out another difficulty. According to Fodor, any causal account, such as Dretske's, has to look for counterfactual situations for assigning meaning. For example consider a child whose "dog" tokens were caused only by dogs. And assume that somehow the time has come and the learning period has ended. Now consider this question: What would have happened if the child had seen a cat on a dark night during the learning

period? Assume that, under certain circumstances, cats on dark nights lawfully cause "dog" tokens. Then it is obvious that a cat on a dark night would have caused "dog" tokens within the learning period. So, in terms of counterfactual relations, even if all "dog" tokens as a matter of fact caused by dogs in the learning period, the meaning of "dog" would still be dog or cats on dark nights or whatever counterfactually causes "dog" tokens.

2.3 Teleological Theories

Another group of theories seek to locate genuine meaning by utilizing the theory of natural selection developed by Darwin and Wallace. Below is a short description of such a theory by Papineau (1997):

> The central thesis of the teleological theory is that the representational contents of the mental states can be explained in terms of the biological functions of those states.… Defenders of the teleological theory standardly cash out their references to biological purposes *etiologically*, in terms of past processes of natural selection. According to the etiological account of biological function, an item X has a biological purpose Y if and only if X is now present because previous versions of X were selected in virtue of doing Y (Papineau, 1997, p1).

Now let us see some possible problems raised in the literature this appealing account may create. The first one is the case of "swampman" (the man with no history). Imagine that a molecule-by-molecule duplicate of a person is created. Our intuitions tell us that this clone's mental representations should have the same content as its duplicate. However, this clone has no evolutionary history, so according to what shall we assign content to its mental representations? The mental representations of

the clone did not function for any biological purposes, hence according to the teleological theory their content must be empty.

The biggest problem for teleological theories is that in fact they do not solve the disjunction problem. Let us again consider the frog example. Suppose that whenever a fly passes by, a certain neural mechanism in the frog's brain is activated. Should we say that the activated neurons in the frog's brain represent flies or small moving black things? At first it seems that the teleological theory has an answer, since the biological function of the neural mechanism is to represent flies, i.e., the teleological theory seems to assign flies as the content of the neural mechanism. But this can't be true, for whether or not the biological purpose of the neural mechanism is detecting flies *or* small moving black things frogs will continue to catch flies and hence will be selected for evolution.

CHAPTER III

MISREPRESENTATION AND ROBUSTNESS: NEW SETTING

The issues which are labeled under "disjunction problem," "misrepresentation," and "robustness" have significant roles in the project of naturalized semantics. It is widely accepted that any theory of naturalized semantics should have a solution/explanation for these notions. In this section I will consider them again, examine whether they are different names for the very same problem or have different meanings, and criticize the currently accepted formulations of them. In fact since the disjunction problem and the robustness of meaning are originally put forward by Fodor (1990a, p59), our discussion will be largely a criticism of Fodor. Finally I will introduce new definitions which I think are more precise and will be vital for the theory that will be developed later.

Let us begin with the disjunction problem. The disjunction problem was formulated by Fodor to show the inadequacy of the crude causal theory of content. It is usually held that there can be many different causes of a symbol (mental representation) and that some of these causes are wild (not meaning forming, i.e., not in the extension of the symbol). For example the mental symbol "cow" means cow but a horse on a dark night can also cause "cow." However since they both cause "cow," according to the crude causal theory the content of "cow" will be *cow* or

horse on a dark night. This can be generalized; not all causes of a symbol are in its extension so we want to exclude some of the causes of a symbol "S" from the extension of "S". Else we will get a disjunctive meaning for "S", namely, $cause_1 \vee cause_2 \vee cause_3 \vee \ldots \vee cause_n$; hence the disjunction problem.

Misrepresentation and robustness, on the other hand, are considered to be two important features of meaning that any naturalistic theory should account for. These two properties are usually discussed in a rather mixed way so that their differences go unnoticed. Are they one and the same property? Is one of them a special case of the other? Or are they related but different properties? And what is their relation to the disjunction problem? Fodor answers some of these questions:

> Errors raise the disjunction problem, but the disjunction problem isn't really, deep down, a problem about error. What the disjunction problem is really about deep down is the difference between *meaning* and *information* (Fodor, 1990a, p.90).
>
> [The following might be called the *robustness* of meaning]: In actual fact, "cow" tokens get caused in *all sorts* of ways, and they all mean *cow* for all of that. Solving the disjunction problem and making clear how a symbol's meaning could be so insensitive to variability in the causes of its tokenings are really two ways of describing the same undertaking. If there's going to be a causal theory of content, there has to be some way of picking out *semantically relevant* causal relations from all the other kinds of causal relations that the tokens of a symbol can enter into (Fodor, 1990b, p.91).

It is clear that according to Fodor robustness of meaning and disjunction problem are different aspects for the same issue. Another thing which is also clear is the definition of "robustness". According to Fodor robustness of meaning refers to the fact that sometimes tokenings of symbol "S" are caused by instantiations of

properties which are not in the extension of "S". What is less clear is the relationship between misrepresentation and robustness. Fodor did not explicitly state this relationship, but according to his writings it seems that misrepresentation, for him, is the result of a special case of robustness. If you token "cow" after seeing some milk, this is a case of robust tokening[2] which is not a misrepresentation. But if you token "cow" after seeing a horse on a dark night this is a case of robust tokening which is also a misrepresentation. Misrepresentation is a false tokening of a symbol, i.e., its truth conditions are not satisfied.

Problems begin to emerge when we consider Fodor's examples more carefully. The first, and relatively minor, problem is related with horses on dark nights. According to Fodor, the crude causal theory assigns the content *cow* or *horse-on-a-dark-night* to "cow" tokenings as explained before. However, I think, according to the crude causal theory the meaning of "cow" should be *cow* or *horse,* not *cow* or *horse-on-a-dark-night.* If we qualify the circumstances in the horse case then why don't we qualify them similarly in the cow case and say *cow-under-daylight*. Fodor insists, rightly I think, that laws of special science all contain *ceteris paribus* clauses and this does not invalidate their being genuine laws. But then we should formulate the law as a statement about horses and "cows" and move the qualifications like "on a dark night" into the *ceteris paribus* clause. Note that this minor modification does not affect Fodor's theory but is necessary for a clearer exposition of the matter.

[2] I use the expression 'robust tokening' to refer to mental tokenings of a symbol "S" which are caused by instantiations of properties which are not in the extension of "S". For example the tokening of "cow" by milk is a robust tokening whereas the tokening of "cow" by cows is not.

Now let us look at the laws in Fodor's cow-horse example. According to Fodor there are two laws involved here: the first one is the nomic connection between cows and the mental symbol "cow" and the second one is between horses and the mental symbol "cow". If we formulate the laws in terms of properties[3], "cow" tokens are caused by the instantiations of the complex property of *being a cow* or the property of *being a horse*. Let us consider whether the "cow" tokenings caused by horses are a kind of robust tokening or not. If we consider *coarse-grained* properties like the property of *being a horse*[4] it clear that this is a robust tokening since the property of *being a horse* is not in the extension of "cow". However my intuitions say that there is a difference between this kind of robust tokenings and the "cow" tokenings caused by the property of being milk. In fact Fodor is wrong in formulating the law as stating a causal relation between "cow"s and horses (or horses-on-dark-nights). Horses cause "cow"s because of the *properties* they share with cows: like having four legs, having a certain size, and so on. The situation here is different from the situation in robust tokenings that Fodor describes. In fact this type of tokenings are very common, as in, causation of "snake" tokenings by ropes, causation of "water" tokenings by mirages, or the illusions performed by a magician. Ropes cause "snake"s because of the properties they share with snakes; similarly mirages cause "water" because of the properties they share with water. In other words, contrary to what Fodor thinks, causation of "cow"s by horses (that is to say, by the instantiations of properties of having four legs, having a certain size, etc.) is not wild but meaning forming!

[3] What is the relata of causation is one of the major questions of metaphysics of causation. There are several answers, such as; they are events, facts, or properties. I assume, similar to Fodor, that causation can best be analyzed as a relation between properties.
[4] Being a horse is a coarse-grained property relative to having four legs, having a tail, etc.

But if we formulate causal laws with finer-grained properties then one might complain that we have the disjunction problem back again. But this is not so. This will be clear in the following sections. Let me first make some terminological clarifications. At the beginning of this section I have summarized Fodor's definition of the disjunction problem and robustness. And I have noted that Fodor does not give a precise treatment of misrepresentation. I agree with Fodor's definition of robustness: tokening of a symbol "S" by a property which is not in the extension of "S", e.g., tokening of "horse"s by horse shoes. I also agree that robustness of meaning and the disjunction problem are different ways of looking at the same phenomena. What I disagree with is his analysis of misrepresentation. Let me try to explain. I maintain that there are basically two types of tokenings: robust and meaning forming. Robust tokenings are exactly the same as in Fodor's definition of robustness. Meaning forming tokenings are also not new: They are caused by properties which are in the extension of the symbol. What is interesting is that both types of tokenings allow misrepresentation. This is exactly what Fodor seems to have missed. For example when a person wrongly believes that it's raining because of the darkness in her room, this is a case of a misrepresentation which is caused by a robust tokening. Darkness causes a "rain" tokening in her brain. However when a person tokens "water" by seeing a mirage this is a case of a misrepresentation which is meaning-forming. For some properties of mirages (e.g., the property of reflecting light), which they share with water cause "water" tokens. In sum we have four cases of tokenings:

(i) Meaning forming tokening, true-representation.

E.g.: causation of “water” mental tokens by water.

(ii) Meaning forming tokening, misrepresentation.

E.g.: causation of “water” mental tokens by mirages.

(iii) Robust tokening, true-representation.

E.g.: causation of “cow” mental tokens by milk.

(iv) Robust tokening, misrepresentation.

E.g.: causation of “rain” mental tokens by darkness.

I claim that for the robust tokenings the problem is to find a principled way to exclude their causes from the extension of the symbol they cause, and for the meaning forming tokenings the problem is to show how misrepresentation is possible. I believe that the two problems require different theories and that Fodor’s theory of content offers a solution only to the second problem, and moreover it is insufficient to solve it. In the rest of this thesis I will discuss these claims in detail and offer two theories enabling us to solve the two problems that I have sketched.

3.1 Fodor’s Theory of Asymmetric Dependence

Fodor’s own solution to the disjunction problem is dubbed by him as the asymmetric dependence theory (Fodor, 1990b). The notion of asymmetric dependence should be understood as a component of his naturalistic theory of representation. Although Fodor’s theory of representation has gone through several revisions the following three statements, I believe, capture the core of his theory:

A symbol "S" represents the property P if,

(i) Instances of P lawfully cause "S" tokens.

(ii) Sometimes tokens of "S" are lawfully caused by non-Ps.

(iii) Non-P caused "S" tokenings asymmetrically depend on P caused "S" tokenings .

The statement (i) is used to reduce the representation relation to the naturalistic relation of causation. Although objections can be raised to (i), for the purposes of this thesis, I assume it to be non-problematic and concentrate on the statements (ii) and (iii). These two statements are closely connected and hence need to be analyzed together. But for ease of exposition I will start with the notion of asymmetric dependence. I claim that although this notion contains substantial insight, it affords us to solve the problem of misrepresentation only *partially*. Let us start by Fodor's own description of his theory:

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

To grasp Fodor's insight here is not an easy matter. A number of philosophers have complained that the notion of asymmetric dependence as described by Fodor is vague (Boghossian, 1991), non-naturalistic (Loewer, 1996) and even not a solution

but a redescription of the problem (Hutto, 1999). Because of the common complaint of vagueness against his theory, Fodor sometimes feels forced to give a possible world interpretation (e.g., his reply to Block in Fodor, 1990b). According to this interpretation "cow" means cow and not *cat* or *cow or cat,* because there are nomologically possible worlds where cows cause "cow" tokens and noncows don't; and that they are nearer to our world than any other world in which some noncows cause "cow" tokens and no cows do. Although the possible world interpretation of asymmetric dependence might be technically the most precise one, Fodor's important insight can be best understood in still another way. Fodor often talks about breaking the nomic connection between cows and "cow" tokens in place of talking about the world in which no cows cause "cow" tokens. According to this interpretation of asymmetric dependence, "cow" means cow and not *cat* or *cow or cat* because if you break the connection between cows and "cow"s then the connection between cats and "cow"s gets broken too but not the other way around, i.e., if you break the connection between cats and "cow"s the connection between cows and "cow"s remains. One might still complain that the notion of breaking nomic connections is still vague and can be interpreted in at least two ways. Firstly, one may think that cows do not cause "cow" tokens because the properties in virtue of which cows cause "cow" tokens cease to exist. This first interpretation is presumably not what is intended. The other interpretation is that cows do not cause "cow" tokens because the properties in virtue of which cows cause "cow" tokens, somehow, stop causing "cow" tokens. We believe that this second interpretation is on the right track. From a Humean viewpoint since causation does not involve a necessary connection, it is not hard to think of a world in

which the properties in question still exist but there is no more a causal relation between them.

I believe that illustrating the asymmetric dependence on the mirage example will help to strengthen our feelings about the importance of asymmetric dependence in discriminating between what is misrepresentation and what is not. Both a mirage on a desert and a pool of water cause lawfully “water” tokens in humans. We know that mirage-caused “water” tokens are misrepresentations while water-caused “water” tokens are not. Asymmetric dependence relation says that mirage-caused “water” tokens are misrepresentations since there is a way to prevent mirages from causing “water” tokens without destroying water to “water” connection, but there is no way to prevent water from causing “water” tokens without destroying the mirage to “water” connection. Imagine that we close the eyes of a person to block the causation of “water” tokens. At first one might reply that this will block both the water to “water” and mirage to “water” causal connections. However “water” tokens might be caused not only visually but also by other sense modalities. In other words, a person whose eyes are closed might token “water” when she drinks some or hears the sound of water coming from the tap. But it is not easy to find a similar way that will block water to “water” connection but not the mirage to “water” connection. Accordingly “water” means water and “water” tokens caused by mirages are misrepresentations.

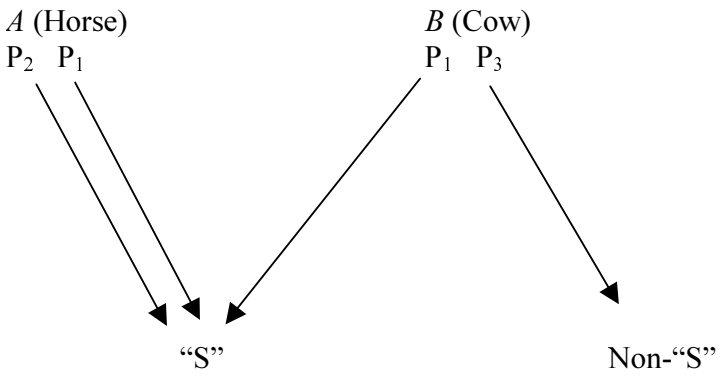
The primary insight in the idea of asymmetric dependence is that you can’t have false representations unless you have true ones. In other words false representations

depend on the true representations but not vice versa. This much I think is correct. But it is too vague and uses semantic notions. It needs a more formal interpretation which does not use semantic terms and is more intelligible. However the formulations given by Fodor, in terms of possible worlds or breaking the nomic connections, deviates from the original insight and makes it look incorrect as can be seen clearly in Hans Robert Cram's interpretation that will be discussed next.

3.2 Cram's Interpretation of Asymmetric Dependence

I have claimed that contrary to what he thinks, Fodor's theory of representation can only be a theory of misrepresentation in meaning forming tokenings and does not explain robustness. Moreover, Fodor's theory can only *partially* explain misrepresentation in relation to meaning forming tokenings. My argument to show the truth of those claims will depend on Cram's (1992) interpretation of Fodor's idea of asymmetric dependence. So it might be thought that my argument can at best show that Cram's interpretation is false, not Fodor's theory. However, I believe that Cram's version of asymmetric dependence is one of the most intelligible ever proposed, and what is more important is that it lays bare Fodor's insight even if it is not sufficient for accounting for misrepresentation in meaning forming tokenings.

Cram illustrates his interpretation with the following kind of figure:



1. Cram's Interpretation of Asymmetric Dependence

Figure 1 illustrates a situation in which two individuals, A and B, cause a token "S". The token "S" is supposed to be a mental representation which represents horses. The individual A, which is a horse, has two properties, P_1 and P_2, while the individual B, which is a cow, also has two properties, P_1 and P_3, one of which is identical with and the other is different from the properties of individual A. The important point about these properties is that each of the properties P_1 and P_2 is *sufficient* to cause the token "S" *independently*.

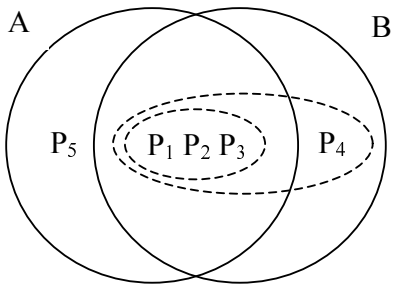
Now we can easily see how asymmetric dependence solves the disjunction problem. According to Cram's interpretation, breaking the connection A $\rightarrow$ "S" means that the properties P_1 and P_2 no longer cause the token "S". Since the only property in virtue of which the individual B causes the token "S" is P_1, the connection

B → “S” also breaks down. But the reverse is not true: if we break the connection B → “S”, which is to say that P_1 does not cause “S” any more, since the causal route from property P_2 to “S” remains, the connection A → “S”, through P_2 is still there. This is why the connection B → “S” is asymmetrically dependent on the connection A → “S” and why the token “S” represents only As and tokening of “S” by Bs is a misrepresentation.

While Cram thinks that this is a precise and intelligible account of asymmetric dependence he is pessimistic about its truth. He does not give any reasons for his pessimism, saying that this is a topic of another paper. We also think that Cram’s account of asymmetric dependence, though an intelligible one, can only partially explain misrepresentation. Suppose that “S” represents only the individual X. Suppose further that individual Y sometimes causes “S” tokens, i.e., they are misrepresentations. Let’s denote the properties of individual X which are not shared by the individual Y by P. Asymmetric dependence works only when the properties in set P are sufficient to cause “S” independently of the other properties of the individual X. For example in Cram’s example the property P_1 is sufficient to cause “horse” tokens. However when this is not the case, i.e., the properties in the set P are not sufficient to cause “S”, asymmetric dependence does not work. I believe that this is not an exception but a common case. For example in his criticism of Fodor’s asymmetric dependence, although Cummins (1989) does not explicitly state it, the kind of interpretation in his mind is very similar to Cram’s. Cummins’ example tests the asymmetric dependence theory with mice and shrews. The true theory of

representation should say that the token “mouse” represents only mice and if a shrew causes the token “mouse” this should be a misrepresentation. However, as Cummins states, in the case of shrews and mice, instead of an asymmetric dependence, there seems to be a symmetric dependence. The reason behind this is that *most* of the properties in virtue of which an individual causes a “mouse” or “shrew” token are common to both shrews and mice. Breaking one connection will break the other.

Let us illustrate the situation one more time by using a somewhat more formal notation. Suppose there are two individuals, A and B, and five properties, P_1, P_2, P_3, P_4, and P_5, as shown in Figure 2.



2. Why Asymmetric Dependence Does Not Work?

Suppose further that individual A instantiates the properties P_1, P_2, P_3, and P_5, and individual B instantiates the properties P_1, P_2, P_3, and P_4. There are two sets of properties circumscribed by the dashed ellipses. Let $X_1 = \{P_1, P_2, P_3\}$ and $X_2 = \{P_1, P_2, P_3, P_4\}$. These sets indicate the properties that are *sufficient* to cause the token “S” when co-instantiated in an individual. For example, an individual which instantiates only the property P_1 or P_4 cannot cause the token “S”. Whereas an individual which

co-instantiates all of the properties P_1 and P_2 and P_3 or P_1 and P_2 and P_3 and P_4 can cause the token "S". Now according to Cram's interpretation of asymmetric dependency these two individuals are *symmetrically* dependent on each other. To see this note that there are only two ways to break the connection between the individuals and the token "S": You can either break $X_1 \rightarrow$ "S" or $X_2 \rightarrow$ "S". Whichever connection you break neither of the individuals can cause the token "S" any more.

To give a flesh-and-blood example, let's consider a rat and a mouse in place of the individuals A and B respectively. Suppose that P_1 is the property of having a pointed snout, P_2 is the property of having small ears, P_3 is the property of having a thin and long tail, P_4 is the property of having small size, and P_5 is the property of having large size. (Note that we use relative terms for the values of the size attribute but one can instead give more quantitative measures of size here.) Asymmetric dependency theory fails in this example because the properties of rats and mice overlap so much that the properties of mice that are not shared by rats (e.g., having small size) are insufficient by themselves to cause the token "mouse".

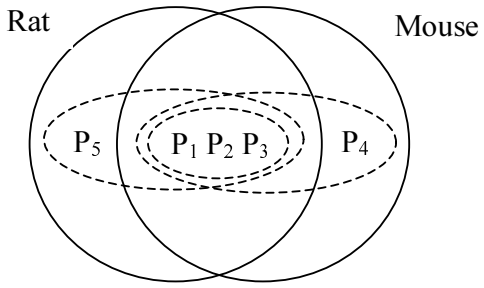
Having seen the inadequacy of the criterion of asymmetric dependence in explaining misrepresentation, let us now see why it cannot explain, even partially, robustness. In fact in light of Cram's interpretation it is easy to see this. Suppose that the extension of "S" is the property P and "S" can occasionally be caused by the property R. That is, "S" is caused by a property which is not in the extension of "S" (e.g., causation of "cow" tokenings by milk.) If we break the P $\rightarrow$ "S" connection,

since R and P are different properties, R $\rightarrow$ “S” connection will remain intact. Similarly if we break the R $\rightarrow$ “S” connection P $\rightarrow$ “S” connection will also remain intact. It follows that “S” represents the disjunctive property P $\vee$ R. Disjunction problem is back again.

CHAPTER IV

MISREPRESENTATION IN MEANING FORMING TOKENINGS

As I stated before the crucial task is to explain misrepresentation in meaning forming tokenings. The critical observation is that, although in meaning forming tokenings all the properties are in the extension of the symbol that is tokened, *different number of properties can cause the same symbol.* Let's consider again the mice and rats example. Figure 3 shows two individuals, a rat and a mouse.



3. Maximal Conjunctive Properties.

Once again the dashed ellipses indicate properties together sufficient to cause a token. The properties P_i are the same properties (e.g., pointed snout, etc.) of rats and mice as given for Figure 2. Instead of two, now there are three causal relations that are in action:

(i) $P_1 \wedge P_2 \wedge P_3 \rightarrow$ "mouse"

(ii) $P_1 \wedge P_2 \wedge P_3 \wedge P_4 \rightarrow$ "mouse"

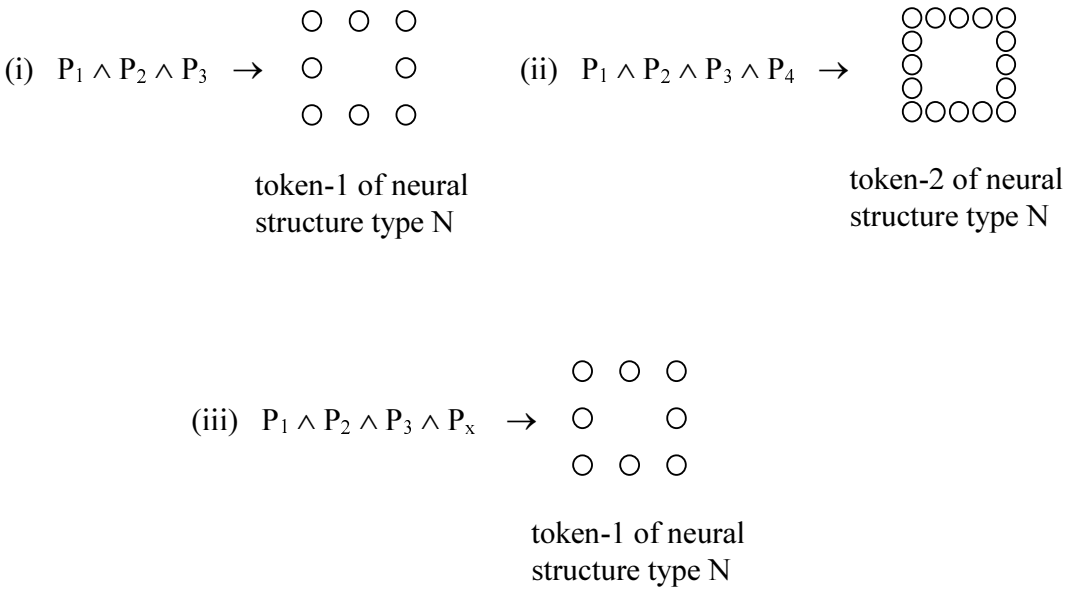
(iii) $P_1 \wedge P_2 \wedge P_3 \wedge P_5 \rightarrow$ "rat"

Misrepresentation occurs when rats in virtue of having the properties P_1, P_2, and, P_3 cause "mouse"s. Assuming all the possible causal relations are the above three, then we will say that the extension of "rat" is the conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_5$, and the extension of "mouse" is the conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_4$. Then to a first approximation the content of a non-semantic type "S" is the *maximal conjunctive property* which is sufficient to cause "S".

Now let us look at a possible complication. One might object that if the conjunctive property $P_1 \wedge P_2 \wedge P_3$ is sufficient to cause "S"s then $P_1 \wedge P_2 \wedge P_3 \wedge P_x$, where P_x is any property, is also sufficient to cause "S"s. Note that the problem occurs when there is a subset relation between two conjunctive properties as in the first two causal relations (i) and (ii) above. If a property P is sufficient to cause a symbol then any conjunctive property including P also causes that symbol. We should have a principled way of including the relevant properties, e.g., property P_4, and excluding irrelevant or redundant properties. This complication is in fact terminological and can be solved by relying on the type/token distinction. We assume that causation is a relation between property tokens. On the other hand the mental representations which we have been indicating using quotes, e.g., "cow", and which we deliberately call "mental tokens" or *symbols*, are not really tokens but are types.

Accordingly, the conjunctive property $P_1 \wedge P_2 \wedge P_3$ and the conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_4$ cause the same type of symbol, namely "mouse", but different tokens of it. On the other hand, if we have replaced the property P_4 with some arbitrary property, say P_x, then the new conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_x$ would cause the same mental sign *token* as the conjunctive property $P_1 \wedge P_2 \wedge P_3$, i.e., the property P_x is causally irrelevant.

Let me make this type/token issue more precise by an illustration.



4. Redundant Properties in Causation.

Three causal relations are depicted in Figure 4. In the causal antecedents there are conjunctive properties which are sufficient to cause a certain neural structure as shown in the consequents (represented by a square made of dots for illustrative

purposes.) The neural structure is supposed to be a mental representation. In Figure 4 there are two tokens but only one type of neural structure N. The fact that these neural structure tokens belong to the same type is represented by their being square shaped. So the conjunctive properties in question cause the same mental process or structure types in that they all have square shape. On the other hand, more finely speaking, there are two tokens here because one of the squares is sparse and the other is thick.

Now let us see why P_x is a redundant property whereas P_3 is not. The conjunctive property $P_1 \wedge P_2 \wedge P_3$ causes a sparse square. When we add the property P_4 then the conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_4$ now causes a thick square, i.e., same type square but a different token. However addition of the property P_x does not make change, i.e., the conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_x$ causes the same mental token as does $P_1 \wedge P_2 \wedge P_3$ and therefore the same type. This is why P_x is a redundant or irrelevant property but P_3 is not.

Misrepresentation in meaning forming tokenings occurs when an individual A, in virtue of having some set (or conjunction) of properties P, causes a mental representation "S," and if A does not possess all the properties in R where P is a subset of R and R is the maximal conjunctive property which is sufficient to cause "S". A key factor in misrepresentation is the existence of conjunctive properties, which cause the same type of mental representation, and which have a subset relation with each other. A cow can be mistaken for a horse because cows and horses share the properties of having a certain size, having four legs, etc. But the extension of

"horse" includes individuals having not only these shared properties but also others which are absent in cows.

It might be complained that this analysis leads to the classical theory of concepts, which defines concepts by necessary and sufficient properties, and that should be avoided. After all Wittgenstein (1953) convincingly argued that some concepts, like the concept of game, cannot be defined by a set of necessary and sufficient properties. Fortunately the theory I propose naturally accommodates family resemblances. I have asserted that the extension of "S" is composed of the individuals possessing the maximal conjunctive property which is sufficient to cause "S". Now suppose that each one of the following three conjunctive properties is sufficient to cause "S" as indicated below:

(i) $P_1 \wedge P_2 \wedge P_3 \rightarrow$ "S"

(ii) $P_1 \wedge P_2 \wedge P_3 \wedge P_4 \rightarrow$ "S"

(iii) $P_1 \wedge P_2 \wedge P_3 \wedge P_5 \rightarrow$ "S"

If there were only the causal relations (i) and (ii), since $P_1 \wedge P_2 \wedge P_3$ is a subset of $P_1 \wedge P_2 \wedge P_3 \wedge P_4$ we will say that the content of "S" is the individuals with the maximal conjunctive property $P_1 \wedge P_2 \wedge P_3 \wedge P_4$. However there is no subset relation between the conjunctive properties in (ii) and (iii). If there is no other cause of "S", this means the extension of "S" is individuals satisfying the disjunction of them, namely, $(P_1 \wedge P_2 \wedge P_3 \wedge P_4) \vee (P_1 \wedge P_2 \wedge P_3 \wedge P_5)$. In other words there may be

disjunctive contents similar to the content of the word 'game.' In this case, misrepresentation occurs when an individual A, in virtue of having some properties, P, causes a mental representation "S" and if A does not possess all the properties in at least one of the disjuncts.

This kind of disjunctive meaning should not be confused with another kind. This other kind was raised by Baker's (1991) as a counter example to Fodor's theory of asymmetric dependence. Baker imagines a situation in which a person, young Sally, lives in an environment where there are equal number of ordinary cats and robot-cats. Robot cats can only be distinguished by experts, i.e., their appearances and behavior are the same as those of the ordinary cats. Sally has tokened "cat" only by seeing robot-cats and she has not seen any real cat. One day Sally sees a real cat. Now, asks Baker, what is the meaning of the "cat" tokening caused by this real cat? Baker thinks that there are three alternatives all of which are unacceptable: cat-caused "cat" tokens represent a cat, cat-caused "cat" tokens represent a robot-cat, or cat-caused cat tokens represent a cat *or* a robot-cat. In his reply Fodor opts for the third alternative, namely, cat-caused "cat" tokens represent a cat or a robot cat. I will not go through Baker's reasons why she finds all of the three alternatives unacceptable. What I want to say is that I would also choose the third alternative as Fodor does. Baker's example is similar to Putnam's (1975) twin-earth example. In that thought experiment Putnam imagined two persons who are molecule-by-molecule duplicates of each other. However one of them lives in our world but the other lives in another world which is exactly similar in all respects to our world, except that instead of H_2O

in our world, in the imagined world there is XYZ. Like the robot-cat, XYZ is very similar to H_2O in terms of appearance and behavior such that it fills the lakes, is drinkable, people call it water, etc. Now the question is whether the meanings of the "water" tokenings in our two persons are the same or not. My intuitions, contrary to Putnam, say that their meanings are the same. In my view symbols represent properties. The individuals that are in the extension of a symbol are all the individuals which instantiate these properties. For example, the content of the "cat" tokenings of Sally are those properties instantiated by both cats and robot-cats. A similar treatment can be given for H_2O and XYZ example. Accordingly in both cases the meaning is disjunctive and this is not a problem, as long as we make room for misrepresentation and robust tokenings.

Since I see the theory developed in this section as a correction and improvement on Fodor's theory of content let me reformulate Fodor's three conditions of meaning in light of the new theory:

A symbol "S" represents the property P if,

(i) Instances of P lawfully cause "S" tokens.

(ii) Instances of proper subsets of P lawfully can sometimes cause "S" tokens.

(iii) P is the maximal conjunctive property which is sufficient to cause "S"

The condition (i) is also found in Fodor's theory of content we gave on p.27. It is basically there to reduce the relation between a symbol and its reference to causation, and hence to make the theory naturalistic. But (i) is not sufficient by itself to allow for misrepresentation.

As I stated earlier different number of properties can cause the same symbol. This feature is not only possible but also a necessary condition for representation. Causation of a symbol by a certain property does not mean that the symbol represents only that property in fact it usually represents more. If the individual which instantiates the property that caused the symbol does not instantiate those other properties then we say that misrepresentation occurs. Note that the conditions (ii) and (iii) also prevent pan-semanticism: the view that meaning is everywhere. For a symbol to have meaning at least two different sets of properties should cause that symbol type and there should be a proper-subset relation between the two. There is plenty of causation in nature but only some of them satisfy this condition.

If there are different sets of properties, one of which is a proper subset of the other, that cause a certain symbol then one might naturally ask which one is the content of the symbol. The condition (iii) answers this question. The maximal conjunctive property among the causes of a symbol is the one which constitutes the content. As I mentioned before, there may be more than one maximal conjunctive property capable of causing the symbol in which case the content is the disjunction of these maximal conjunctive properties. Note that finding out what is(are) the maximal

property(ies) for a given symbol type is an empirical issue. When science discovers all possible properties that cause a certain symbol, it would easy to find its content: it is simply the maximal one or the disjunction of those maximal ones.

CHAPTER V

ROBUSTNESS

Even if the account in the previous sections given for content-forming tokenings is true, robust tokenings still create a problem. Robust tokenings, by definition, are not meaning forming and so we have to find a way to exclude them from the extension of a symbol. However the account thus far is unsuccessful in this respect. If empirical science made a list of all possible causes of a symbol the properties which cause robust tokenings would be also in the list. For example, since sometimes milk causes the mental symbol "cow," the property of being milk will also have to be included in the content of "cow," although it shouldn't be. How can we exclude the properties which cause robust tokenings? Our task is simpler then Fodor's. According to Fodor causation of "cow" by horses and milk are both robust. But we have carefully distinguished these two types: Horses share some properties with cows but milk does not. In fact it is not the horses but those properties of horses that they share with cows that do the causation. That is, the causation of "cow" by horses is not robust but meaning-forming as I have explained in section 3. Now our problem is more specific; we need to solve the problem posed by robust tokenings only.

A natural strategy to follow for a solution of this problem would be to look for something common between the varieties of robust rokenings. In fact when we look

closer, there does seem to be something they share. Suppose your friend always walks with his dog. If one day you see your friend walking without his dog most probably you will token "dog" although there is no dog around. Why is it that your friend but not another thing that caused a "dog" tokening in you? What is the relation between your friend and his dog? It is the fact that you always see them together, i.e., your mental symbol of dog and your mental symbol of your friend are associated in your mind. It is always such associations that makes robust tokenings possible. You associate milk concept with cows or horse shoe concepts with horses. That is why when you see some milk you token "cow". But how can this observation, even if it is true, help us? Association, by definition, is between concepts or mental representations. That is, a mental representation can only be tokened robustly by another mental representation. For example when you see some milk, you first token "milk" and then that "milk" token causes the "cow" token.

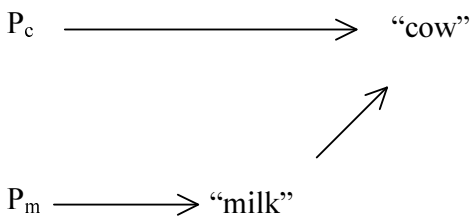
Now we have another (fourth) condition for representation:

(iv) We should exclude all properties whose instantiations cause a symbol only via activating (causing) another symbol.

If an instantiation of a property P causes a symbol "S" *directly*, i.e., if there is no other symbol in the causal path between P and "S", then that property is a meaning-forming one. Milk causes "cow" only after causing "milk" first but milk causes "milk" directly.

When we consider a causal law like P → "S", we always mean the property P causes "S" non-redundantly. In other words any other property which is *less complex than* P would not cause "S". Making this provision I can now develop a simple model for the things I have said in this section.

Let P_c be the property of being cow and P_m the property of being milk. Figure 5 shows a simple model of the causal relations among these properties and mental representations "milk" and "cow".



5. Robust Tokening of "Cow" by The Property of Being Milk.

One last problem remains. How do we know that "milk" is a mental representation? When trying to formulate the conditions for being a mental representation, did we use the notion of mental representation which leads to circularity? No, actually only the conditions (ii) and (iii) decide whether a brain state is a representation or not. Condition (iii) says what is its reference and we need the fourth condition mentioned above only to block properties, which cause robust tokenings, to enter into the extension of a symbol.

Let me finally discuss how an omniscient scientist can find out the mental representations and their reference in an organism. First the scientist finds out all possible the causal relations among properties and brain states. These brain states are candidates for being a mental representation. In order to be a mental representation they have to satisfy the condition (ii). In other words there should be different set of properties which cause them. Having found the mental representation now we exclude those properties which cause robust tokenings by applying the condition (iv). Finally the condition (iii) gives us the references of the mental representations.

CHAPTER VI

CONCLUDING REMARKS

In order to give sufficient conditions for reference I have offered two independent theories. It is almost always more preferable to solve a problem using as few theories as possible. However as I have tried to explain throughout the thesis, contrary to Fodor, there is not one but two independent problems and hence it is natural to offer two different theories for each of them. We saw that many cases of misrepresentation are not the result of robustness of meaning. Since Fodor thinks that there is only one kind of problem he tried to give an account using only the theory of asymmetric dependence. Nevertheless, as we have seen, his theory can only account for the misrepresentation cases in meaning-forming tokenings.

The basic idea in the theory that I have offered to explain misrepresentation in meaning-forming tokenings is that humans might token the same mental representation by being exposed to different amounts of instantiations of properties. In other words humans employ a kind of inference mechanism during tokening of mental representations. By seeing only the tail of a horse one might think that there is a horse over there. And inference may lead to error.

The other theory, which employed the notion of associations, is offered to account for robust tokenings. Since Fodor sees only one kind of problem, this solution was unavailable to him. If one maintains, as Fodor did, that being a horse on a dark night is a property, which shares nothing with being cows then one cannot use the associative kind of solution that I have offered. Horses on dark nights sometimes cause "cow" tokenings directly, i.e., without first causing other tokens.

I want to mention one last point, which is an important threat to all kinds of causal theories of reference. Usually there is more than one cause of an effect. Rubbing a match against a rough surface causes fire. However rubbing the match is not the only cause; in fact there are a plenty of other causes such as the presence of oxygen and absence of wetness of the match. Imagine an absurd world in which there is no oxygen but lots of match rubbings. In that world when one frees some oxygen near a rubbing match, fire is caused. But again freeing of oxygen is not the only cause of fire in that world. That is, selecting one of the causes and regarding the others "causal conditions" or "standing conditions" is interest driven and depends on the context. When Fodor talks about the law between cows and "cow" tokenings the same problem occurs. There is no objective basis for selecting cows as *the cause* of "cow" tokens. There are other causes of "cow" tokens like the presence of light that is reflected by cows and reaching out eyes. Though it is counter intuitive, it is perfectly legitimate to state the law as between presence of light and "cow" tokens rather than presence of cows and "cow" tokenings. So a causal theory has to include all the causes of "cow" tokens in the extension of "cow" mental representation. Note that

asymmetric dependence offers no help since all of the causes of “cow” (e.g., cows, presence of light, etc.) depend symmetrically on each other, i.e., without one of them causation cannot take place.

REFERENCES

Brentano, F., (1995), *Psychology from an Empirical Standpoint*, 2nd ed., trans. A. Rancurello, et. al., London: Routledge.

Baker, L. R., (1991), "Has content been naturalized?" In Loewer B., and Rey, G., (eds.), *Meaning in Mind*, Oxford: Blackwell.

Block, N., (1994), "Advertisement for a Semantics for Psychology," In Stich, S., and Warfield, T., (eds.), *Mental Representation: A Reader*, Oxford: Blackwell.

Boghossian, P., (1991), "Naturalizing Content," In Loewer B., and Rey, G., (eds.), *Meaning in Mind*, Oxford: Blackwell.

Churchland, P., (1981), "Eliminative Materialism and Propositional Attitudes," in Rosenthal, D.M. (1991), pp. 601-612.

Cummins, R., (1989), *Meaning and Mental Representation,* Cambridge, MA: MIT Press.

Cram, H., R., (1992), "Fodor's Causal Theory of Representation", *The Philosophical Quarterly*, 42, 56-70.

Dennett, D., (1975), "Three Kinds of Intentional Psychology," in Rosenthal, D.M. (1991), pp. 613- 626.

Dretske, F., (1981), *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.

Fodor, J., (1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, Massachusetts: MIT Press.

Fodor, J., (1990d), *A Theory of Content and Other Essays*, Cambridge, Massachusetts: MIT Press.

Fodor, J., (1990a), "A Theory of Content I: The problem," In Fodor, 1990d.

Fodor, J., (1990b), "A Theory of Content II: The theory," In Fodor, 1990d.

Fodor, J., (1990c), “Semantics, Wisconsin Style,” In Fodor, 1990d.

Fodor, J., (1990d), *A Theory of Content and Other Essays*, Cambridge, Massachusetts: MIT Press.

Fodor, J., Lepore, E., (1994), “Why meaning (probably) isn’t conceptual role,” In Stich, S., and Warfield, T., (eds.), *Mental Representation: A Reader*, Oxford: Blackwell.

Hutto, D. D., (1999). *The presence of mind*, Philadelphia, J. Benjamins Publishers.

Grice, H. P., (1957), “Meaning,” *Philosophical Review*, 66, 377-88.

Loewer, B., (1996), "Naturalizing Semantics," In Wright and Hale (eds.), *The Companion to the Philosophy of Language*, Oxford: Blackwell.

Papineau, D., (1998), “Teleosemantics and Indeterminacy”, Australasian Journal of Philosophy, 76, 1-14

Putnam, H., (1975), “The meaning of meaning." In H. Putnam, (Ed.), Mind, Language and Reality. New York: Cambridge University Press.

Searle, J. R., (1980), “Minds, brains, and programs,” *Behavioral and Brain Sciences*, 3 (3), 417-457.

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