This is to certify that the

dissertation entitled

Structural Semantics:
A New Picture Theory of Representation

presented by

Heather E. Johnson

has been accepted towards fulfillment
of the requirements for

Ph.S. degree in Philosophy

Richard J. Hall
Major professor

Date May 8, 2002

STRUCTURAL SEMANTICS: A NEW PICTURE THEORY OF REPRESENTATION

By

Heather Elizabeth Johnson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy

2002

ABSTRACT

STRUCTURAL SEMANTICS: A NEW PICTURE THEORY OF REPRESENTATION

By

Heather Elizabeth Johnson

Structural semantics is a theory that purports to explain how mental states come to represent or "be about" external objects and states of affairs. According to this view, both mental states and the things they represent can be described as sets of elements structured by relations. It is in this sense that mental states and external states of affairs are what we call "relational systems." A mental state, or "cognitive relational system" *represents* some external state of affairs, or "external relational system," when and only when there is a homomorphic function mapping the elements of the cognitive relational system to those of the external relational system.

Because the only necessary condition for representation is the existence of a homomorphism between the representation and thing represented, structural semantics may be criticized for failing to specify adequately which of any number of external relational systems a given representation is about. In other words, insofar as representations will always map homomorphically to more than one external relational system, structural semantics cannot support the assignment of representations to *unique* objects or states of affairs in the external world. This is referred to as "the uniqueness problem" for structural semantics.

This dissertation examines various ways of raising the uniqueness problem for structural semantics including a version of the problem expressed through Quine's

arguments regarding translational indeterminacy and a persistent version of the problem introduced by Quine in Ontological Relativity and Other Essays.[1] In addition to providing an exposition of the "mechanics" of structural semantics as a theory of mental representation and to formulating a response to the uniqueness problem, this dissertation attempts to evaluate the ability of structural semantics to satisfy a number of basic conditions commonly thought to be required for any adequate theory of mental representation.

[1] Quine, W.V.O. Ontological Relativity, and Other Essays. New York: Columbia University Press, 1969.

This dissertation is dedicated to my husband Serge, who made himself available on many occasions as an editor, mathematics consultant, and coach.

ACKNOWLEDGMENTS

The original idea for this topic arose in connection with a seminar on mental representation offered at Michigan State University in the Fall of 1998 by Professor Richard Hall. My experience in this seminar fostered an interest in mental representation that proved enough to sustain the completion of this work almost four years later. During that time, I've consulted a lot of people about the ideas that are described and defended here. Among the most significant contributors has been my husband, Serge Canizares, who has served as an invaluable resource and has been willing to drop in on the middle of a half-formed train of thought, doing his best to make sense of it and provide critical feedback.

I would also like to thank my mother, Jacqueline Johnson, who supports me in everything I do, and has been an enormous encouragement at the times when I most needed to be encouraged. Were it not for her love and support, I would not be in a position to accomplish the goals associated with this work.

Finally, I'd like to express my deepest appreciation for my advisor, Richard Hall, who has spent many precious hours evaluating my work and providing insightful criticisms and suggestions for improvement to it. Our discussions about the ideas defended here have always revealed aspects of the problem which I had not even imagined before, and evidence of his contributions is found in every page of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Images in this dissertation are presented in color.

KEY TO SYMBOLS OR ABBREVIATIONS

Σ - a cognitive system possessing mental representations
| X | - mental representation of X
$\mathfrak{C}$, $\mathfrak{R}$, $\mathfrak{I}$ - structured sets of elements, e.g.-relational systems or structures
$P^{\mathfrak{R}}$ - the predicate P in the structure $\mathfrak{R}$
$f^{\mathfrak{R}}$ - the function f in structure $\mathfrak{R}$
Γ - a set of formulas in a language

INTRODUCTION

If a person wants to know how to make a cheese soufflé, one of the things she/he might do is take a trip to the local library in search of relevant information. Today, one can also find such information on the internet, or perhaps by watching one of those Saturday afternoon cooking shows on cable. Suppose I want to learn the art of soufflé-making, and I choose to visit the library to begin my tutelage. How will I find the information that I need? Assuming that I am familiar with the library's indexing system, I should be able to narrow my search to the part of the library containing the book or books I want to read. When I arrive at the right shelf, I should be able to recognize my book by reading the titles. The words in the title, and later, the words contained in the text, *convey information* to me. The title of the book reveals that it is *about* the process of making soufflés and the words it contains collectively *represent* a set of instructions for making soufflés.

What is it that makes this book *about* soufflés and not chocolate mousse or basketball? The immediate answer is that the book contains sentences which *mean* propositions concerning the making of soufflés and not anything else. But this answer only defers the question: we can still ask what it is that makes these sentences mean what they do, and in general, in virtue of what it is that language has meaning.

But ultimately, the question of why a set of representations is *about* soufflés and not something else cannot be reduced to a question about the meanings of sentences anyway. This is because language is not the only way of representing the process of soufflé-making. The same information contained in the book might be accessible online, via a computer, or embedded in a series of digitized photographs. When representations are conveyed via the internet, they are not ultimately reducible to sets of sentences, but to sets of electronic states of a system which hosts them. This collection of system-states represents the process of soufflé-making too, and by hypothesis, represents the very same process as that represented by the library book. Or again, if I were to commit the contents of the entire library book to memory, the process of soufflé-making would be represented in yet another medium, namely, in my own head. In this scenario, my own mental states or states of the brain, somehow represent soufflés and soufflé-making.

The question we want to answer is this: in virtue of what can things like words, states of a machine, and states of mind *represent* objects, processes, or states of affairs in the external world? This dissertation concentrates on the last of these: it focuses on how states of mind can *represent*, or be *about*, another thing (e.g., an external state of affairs). I will propose and defend a theory called "structural semantics" which is an attempt to answer this question.

Chapter 1 of the dissertation is devoted to the critical exposition of a view very similar to the one which I will defend, namely, Robert Cummins' interpretational semantics. Careful consideration of Cummins' view will reveal a potentially serious problem with the way in which he explains mental representation, a problem that applies to my own view as well, and which I will refer to as "the problem of uniqueness." Much of the dissertation is an effort to explain the problem of uniqueness and respond to the threat it presents.

Interpretational semantics and structural semantics both bear a strong resemblance to a class of theories about mental representation known as *picture theories*. In Chapter 2, I describe some of the features of picture theories of representation and compare their strengths and weaknesses to those of interpretational and structural semantics. The examination of the picture theory reveals that the uniqueness problem is not a new concern. Opponents of the picture theory of mental representation have posed objections to picture theory which are strikingly similar to the challenges the uniqueness problem poses. In Chapters 2 and 3, I begin to suggest how to refine Cummins' view in such a way as to respond to the problem of uniqueness. This refinement is what I call structural semantics, and is the view I defend for the duration of the dissertation.

Chapters 4 and 5 are both devoted to explaining and responding to specific ways of raising the problem of uniqueness. Chapter 4 reveals how Quine's arguments concerning translational indeterminacy are really a version of the uniqueness problem. Consideration of the uniqueness problem in this context helps to refine the sort of response which I can offer to it on behalf of structural semantics. And Chapter 5 describes a persistent version of the problem which resists any of the responses which I have defended until that point.

Most of the dissertation is devoted to explaining the mechanics of structural semantics and to defending this view against the problem of uniqueness. However, many philosophers agree that a good theory of mental representation needs to meet at least three conditions in addition to providing a plausible account of how mental states come to have representational content. Therefore, in both Chapters 3 and 6, I will attempt to state how the theory I defend can meet these three basic conditions.

In Chapter 6, I attempt to explain how structural semantics can account for mental representations of things which do not really exist. Explaining representation of non-existent objects is the first of the three conditions which theories of mental representation are typically required to meet. Representations of unicorns, of the boogie man, hallucinations, or of your imaginary friend when you were five are all real enough as mental phenomena; a good account of mental representation ought not to characterize representation in such a way as to rule out these sorts of representations. Second, a theory of representation ought to allow for misrepresentations. For example, it ought to be possible for one to represent the state of affairs expressed by the sentence, 'The cat is on the mat' even when the cat is in fact nowhere near the mat and the sentence is therefore false. Since we are capable of representing things in a way that is inconsistent with the facts, we need an account of representation which is compatible with the possibility of that inconsistency. Finally, in Chapter 3, I discuss how a good theory of mental representation will be consistent with the widely held view that "substitutivity of identity" fails where the representational content of mental states is concerned. For example, in *Star Wars*, it is true that Luke Skywalker believed:

(*s*) I want to be a Jedi like my father.

Despite the fact that, as later movies revealed, Darth Vader and Luke's father are one and the same, we would not conclude that Luke therefore believed:

(*s*') I want to be a Jedi like Darth Vader.

The content of mental states does not necessarily remain the same when we substitute one description of a thing for another description of the same thing. A proposition that expresses the content of Luke's representation is true under one of these descriptions (*s*), but false under another (*s*'). This is a peculiar characteristic of the contents of mental states which one does not see in non-mental contexts. For example,

(*t*) Anakin Skywalker was Luke's father. and
(*t*') Darth Vader was Luke's father.

are the same statement save for the substitution of the co-referential terms, and this substitution does not change the veracity of the proposition. A successful account of mental representation will explain how the content of two representations can be different, even when the things which they represent in the external world are exactly the same.

I believe that structural semantics meets all three of the conditions for a successful account of mental representation and also, offers many additional advantages for an analysis of representational content. I also believe that it can be successfully defended against the uniqueness problem. In many ways, structural semantics is a new and somewhat unconventional approach to the problem of explaining mental content. However, the view has recognizable ties to traditional approaches, and in one notable case, to a recent theory offered by Robert Cummins called "interpretational semantics." I begin the analysis of mental representation with an examination of Cummins' view.

CHAPTER 1: *Robert Cummins' 'Interpretational Semantics'*

Robert Cummins' "interpretational semantics" is a close relative of structural semantics—the view that I will defend here. Cummins has written two books describing and defending interpretational semantics. The first, Meaning and Mental Representation[2] was published in 1989 and provided a basic overview of a number of traditional and modern theories about representation. After detailing the difficulties with each one, Cummins concludes the book with a characterization of his own view and the ways in which it avoids falling prey to the shortcomings of other accounts. In 1996 Cummins published Representations, Targets, and Attitudes.[3] This work went further in its defense of interepretational semantics by attempting to address a number of criticisms that had been raised against it since the release of the first book. In it, Cummins provides a detailed response to the problem of representational error and tries to address a number of concerns regarding whether or not his theory of representation can be fully naturalized.

In this chapter, I will describe Cummins' positive view of representation as it is set forth both in Meaning and Mental Representation and in Representations, Targets, and Attitudes. I'll also examine the way in which Cummins addresses the problem of representational error for interpretational semantics and will consider whether or not there are any other difficulties facing interpretational semantics which require attention. Ultimately I will argue that there are a number of important problems which

---

[2] Cummins, Robert. Meaning and Mental Representation. Cambridge: The MIT Press, 1989.

[3] Cummins, Robert. Representations, Targets and Attitudes. Cambridge: The MIT Press: 1996. p. 96.

interpretational semantics leaves unsolved. The uniqueness problem is arguably the most important of these as it provides much of the grounds for moving to a significantly different account of mental representation, namely structural semantics.

Cummins' positive account of mental representation changes very little from his 1989 work, Meaning and Mental Representation to his later work, Representations, Targets, and Attitudes. The major innovation of Representations, Targets, and Attitudes seems to be his account of cognitive error, although there is also some acknowledgment and discussion of the uniqueness problem in it not present in the former work. I will briefly recapitulate the positive account of mental representation that Cummins has called "interpretational semantics" here, concentrating especially on those elements of the account that allow him to articulate the uniqueness problem.

1.1 Cummins' Positive Account of Representation

1.1.1 The Picture Theory of Meaning

Cummins himself classifies his theory of mental representation as a version of the picture theory of meaning. He believes that picture theories advocate three basic claims about representation and mental representation in particular. These are:

> (1) The most basic form of representation is the form of representation used in mathematics—specifically, representation is a relation between two structures.

Supposing that $\mathfrak{R}$ and $\mathfrak{C}$ are both structures, $\mathfrak{R}$ will *represent* $\mathfrak{C}$ just in case $\mathfrak{R}$ and $\mathfrak{C}$ are isomorphic. Cummins makes what he calls the "radical" suggestion that the kind of representation described here is not only the kind which is most appropriate to mathematics and computer science, but that it is the *only kind capable* of accounting for meaning.

> (2) Only structures (or things which are mathematically equivalent to structures) represent, since only structures can be related by isomorphism.

Cummins isn't explicit about what he means by "structure." However, his discussion of structures and how they represent would suggest that he sees them simply as sets of elements which are related to one another. The nature of the relationships between elements is what can be characterized as the "structure of the set." This description of structures is sufficient for the purpose of understanding Cummins' theory of representation. However, I will give a more formal characterization of structures in Chapter 3.[4]

> (3) *Mental* representation is, likewise, representation which is grounded in isomorphism.

This is the third and final claim that Cummins attributes to picture theories of mental representation. In order to understand more about exactly how mental representation works on this view, let's turn to a more specific description of the theory Cummins calls

---

[4] Notice that if only structures represent, then the picture theory may encounter difficulty with explaining things which seem to acquire meaning *conventionally*. For example, a stop sign (conventionally) means 'stop' because we have, as speakers of a language, decreed that it mean 'stop.' It doesn't seem to be in virtue of any isomorphism that stop signs acquire the kind of meaning that they have. It is arguable that Cummins will need to handle this problem if interpretational semantics is going to provide a comprehensive account of representation. But perhaps Cummins isn't going for a comprehensive account of representation after all. Cummins could contend (and I believe he <u>should</u> contend) that conventional meanings are derivative or dependent on mental meanings in some way and that interpretational semantics is really a theory about mental meanings.

interpretational semantics.

1.1.2 Isomorphism and Multiple Levels of Representation

The crux of Cummins' representational theory is the claim that the representation relation is characterized by an isomorphism between representing and represented structures. He specifically describes the relation as follows:

> The idea is that there is a mapping between the two structures such that (1) for every object in the content structure *C*, there is exactly one corresponding object in the representing structure *R*; (2) for every relation defined in *C*, there is exactly one corresponding relation defined in *R*; and (3) whenever a relation defined in *C* holds of an *n*-tuple of objects in *C*, the corresponding relation in *R* holds on the corresponding *n*-tuple of objects in *R* .[5]

Although a more formal definition of isomorphism than Cummins offers here will be given later, it is useful to remark here that Cummins' definition implies that isomorphic functions are functions which (among other things) map a domain of elements *into* a range of elements. A function maps one set of elements *into* another when the range of the function is a subset of the elements belonging to the set that is the object of the mapping. Isomorphic functions are also one-to-one. Functions which are one-to-one map distinct elements of the domain into distinct elements of the range. Most standard definitions of isomorphism as given in abstract algebra and group theory include one addition condition

---

[5] Cummins (1996), p. 96.

on isomorphic mappings, namely, that the mapping be *onto*.[6] A mapping f between two sets A and B is *onto* when: $\forall b \in B, \exists a \in A$ such that $f(a) = b$.

Although Cummins' intuitive English rendition of the notion of isomorphism does not seem to imply that it requires an onto mapping, I will understand isomorphism as requiring that mappings be both one-to-one and onto.

Note that Cummins uses the phrase "content structure" to refer to the structure that gets represented, while the phrase "representation structure" is used to refer to the structure that is doing the representing. In addition to the content structure however, there are other things which get represented in virtue of the isomorphism between *R* and *C*. In particular: "If a structure *R* represents a structure *C*, then:

1. An object in *R* can represent an object in *C*.
2. A relation in *R* can represent a relation in *C*.
3. A state of affairs in *R*—a relation holding of an *n*-tuple of objects—can represent a state of affairs in *C*."[7]

In other words (in a move that may sound a lot like conceptual role semantics), the capacity of individual objects, relations, or states of affairs within a structure to represent objects, relations, or states of affairs in another is *derived from* the relation which holds between the structures themselves (i.e., the isomorphic relation)—not the other way around.

---

[6] See, for example, the following: Hungerford, Thomas. Graduate Texts in Mathematics: Algebra. New York: Springer-Verlag, 1974. p. 30; Fraleigh, John B. A First Course in Abstract Algebra. 5th ed. Reading: Addison-Wesley Publishing Company, 1992. p. 170; Jacobson, Nathan. Basic Algebra I. 2nd ed. New York: W.H. Freeman and Company, 1985. p. 37; Fletcher and Patty. Foundations of Higher Mathematics. 2nd ed. Boston: PWS-Kent Publishing Co., 1992. p. 170.

[7] Cummins (1996), p. 96.

> Note however, that *R* does not represent *C* because the objects, relations, and states of affairs in *R* represent objects, relations, and states of affairs in *C*, but the other way around: things in *R* represent things in *C* because *R* is isomorphic to *C*. According to [the Picture Theory of Meaning], there is no such thing as an unstructured representation, except in the derived sense just introduced: an unstructured element in a representing structure *R* may be said to represent its counterpart in a represented structure *C*. We must be careful, then, not to think of the objects, relations, and states of affairs in *R* as independent semantic constituents of *R*. [The Picture Theory of Meaning] entails that every genuinely representational scheme is molecular in that the meaningful constituents of a structure are meaningful only in the context of some structure or other.[8]

Although this *does* begin to sound like conceptual role semantics, Cummins firmly disassociates himself from such a view and I believe that he is correct in so doing. After all, the objects, relations, and states of affairs which depend upon their place in the representational structure for obtaining the content they have, do not, in so doing, depend on their role in a *conceptual* network. The relations and objects which collectively constitute the representational structure need to be conceived of as nothing more than a formalism, and so the claim that their semantic content is derived from the higher-level features of that formalism needs to be conceived of as nothing more than a statement about their formal relations to one another and to the structure they help compose. Cummins thinks it necessary to abandon conceptual role semantics primarily because it uses cognitive capacities to explain representation, where he believes one of the major functions of a successful theory of mental representation is to explain cognitive capacities.[9]

---

[8] Cummins (1996), p. 97.

[9] Cummins (1996), p. 90. Perhaps it is not obvious, despite Cummins' program, how this view is appreciably different from a kind of conceptual role semantics. Indeed, whether or not it really is seems to depend on how loosely you construe the notion of conceptual role. If you argue that a mental state is defined by its conceptual role just in case the relations it bears to other mental states (or *would* bear to them, if placed in the right circumstance) are what primarily or even exclusively determine its content, and that nothing much more is required for conceptual role semantics, then Cummin's view certainly qualifies. On the other hand, if you think that there is something more to conceptual role semantics than just a commitment to meaning holism, as do I, then Cummins' interpretational semantics may or may not

1.2 Cummins' Theory of Error

What is unusual about Cummins' approach to accounting for mental representation, besides its explicit endorsement of picture theory, is that it begins with an account of how such representations could be in error. The idea is that competing theories of mental representation have often performed well when it comes to providing an explanation of how cognitive entities represent, but that their very success in this endeavor has prohibited them from generating a robust explanation of representational error. Cummins admits that his approach may amount to little more than "squeezing the balloon in a different place" but claims that the new "bulge" that results may be an interesting one.[10]

1.2.1 An Example of What Can Go Wrong: Fodor's Causal Theory

An example of how a successful account of representational content can compromise one's ability to provide a successful positive account of representational error may be useful in order to illustrate the difficulty that Cummins confronts.

Jerry Fodor has defended a theory of mental representation which has it that the content of a mental representation is, by and large, the *cause* of that representation. He begins with two assumptions: (1) what Cummins calls the "Representational Theory of Intentionality" or the view that the content of mental states is derived from the contents of their constituent representational states,[11] and (2) that mental states are language-like

---

be a version of this view.

[10] Cummins (1996), p. 5

[11] Note that this approach is the opposite of Cummins' own holistic strategy.

symbols, or symbols in what Fodor refers to as the "language of thought."[12] Since for Fodor, mental representations are just meaningful symbols in the language of thought, explaining mental representation amounts to explaining how terms in the language of thought (sometimes called "Mentalese" by Fodor) can have semantic content. The answer Fodor gives, in a nutshell, is that token symbols of Mentalese denote their causes, while symboltypes denote properties whose instantiations reliably cause their tokenings.

The problem that arises for Fodor, and for similar causal accounts of mental representation, is that not all mental representations of X (hereafter denoted | X |) are caused by Xs.[13] In other words, sometimes we *misrepresent* the world around us, and the idea is that a good theory of representation ought to be able to explain failures like this one as well as it can explain the successful cases of representation.

Closely connected to the problems causal theories have with accounting for representational error is what is known as the "disjunction problem." It has been noted that anytime it appears that a non-X has caused the tokening of | X | in some subject, it may be possible to claim that in fact, the content of the subject's mental representation was not | X | at all, but | $X \vee Y$ | (where Y is the misperceived object responsible for the tokening of the subject's representation). For example, suppose I mistake an Oak tree for an Elm. Since | Elm | was really caused by the Oak tree outside my window, perhaps my mental representation ought not be thought of as merely an | Elm | but rather, as an | Elm ∨ Oak | (or, surely even worse, as a | Elm ∨ the Oak tree outside my window |). If

[12] Fodor, Jerry. Psychosemantics.. Cambridge: The MIT Press, 1987.

[13] Nor do all Xs reliably cause tokenings of | X |.

we think of representations as having disjunctive content of this sort, then the contents of representations will always match their causes exactly (or, should they appear not to, this is just as much a reason for thinking we should add another disjunct to our description of the representation's content as it is for thinking that a *mis*representation occurred).

This result is problematic for at least a couple of reasons. First and foremost, it seems to eliminate the phenomenon of misrepresentation rather than explain it. It seems obvious that misrepresentations *do* occur and therefore that any theory which is incapable of explaining them without explaining them away is a theory which does not match up well with the way we believe the world (and our mental representations in particular) to be. In addition, the notion of disjunctive content is epistemologically suspect. If disputes about the content of a representation can always be settled by appending another disjunct to the representation's description—and it would seem that nothing in this view as we have characterized it so far prohibits this—then there will never be any way of establishing the correctness or incorrectness of claims about content. If Jim swears that the content of Ella's representation was | Elm | and I swear that it was | Elephant | for example, then what, in this view, prevents us from compromising to the effect that Ella's representation was really about | Elm $\vee$ Elephant |?

But Fodor has a response to the disjunction problem. Indeed, like Cummins, Fodor starts out by trying to solve the problem of misrepresentation (and the disjunction problem in particular), but in the process, forms for himself a full-blown theory of content. Fodor's account is known as "asymmetric dependence theory." I'll briefly explain it, its advantages, and its drawbacks before moving on to consider the advantages of Cummins' approach.

Fodor argues that any given mental representation is consistently correlated with the instantiation of certain "psychophysical properties"[14] or properties which reliably—even lawfully—impinge on our senses whenever we come into contact with an object possessing them. Because psychophysical properties are causally responsible for the tokening of specific symbols in us, the causal theory of representation, together with an account of just how much exposure we have to have to such properties and in what manner, etc. (a task Fodor leaves to the science of "psychophysics"), "... provides a plausible sufficient condition for certain symbols to express certain properties ...".[15]

If it is an object's *psychophysical properties* which ultimately causes tokenings of mental symbols in us, then we may have a way of explaining how | X |s could be caused by non-| X |s after all. Suppose that my | Elm | representation is provoked by the impingement of some set of psychophysical properties instantiated in the Oak outside of my window (e.g., leafyness, brownish-colored tree trunk shapes, the smell of sap, the look of branches). It is arguable that the Oak outside my window can only cause | Elm |s in me because it shares certain psychophysical properties with Elms—that is, because it *looks like an Elm.* Indeed, if Oaks didn't look like Elms then I would not ever produce the symbol | Elm | in the presence of an Oak. For example, I would not produce the symbol | Elephant | in the presence of an Oak since the two do not share a sufficient number of

---

[14] Fodor, Jerry. Psychosemantics. Chapter 4: "Meaning and the World Order." Cambridge: The MIT Press, 1987. p. 120.

[15] Fodor (1987), p. 113.

psychophysical properties. This may seem like a simple, even trivial observation. But note that it explains what about my tokening of the symbol | Elm | on this particular occasion was a mistake (i.e., because it was caused by a non-Elm) while at the same time allowing that the content of my representation was in fact | Elm | in a way that does not compromise the causal theorist's positive account of representation. In particular, my tokening of | Elm | in the presence of an Oak is made possible only because of the existence of an "asymmetric dependency relation" between Elms and Oaks. And in fact, if the psychophysical properties had by Elms never produced the mental representation | Elm | in me, then presumably those same psychophysical properties instantiated in Oaks wouldn't produce the relevant representation. To put it plainly: if Elms didn't cause | Elm |s then Oaks wouldn't cause them either. After all, this is just what it means to say that such psychophysical properties covary lawfully with the production of mental symbols in us. Fodor's account of mental representation is therefore still a causal account, but one which allows for principled misrepresentation without the price of disjunctive contents.

Perhaps the biggest difficulty with asymmetric dependence is that its avoidance of the disjunction problem seems to come at the price of getting rid of *both* disjuncts in favor of more proximal stimuli. After all, if what makes Oaks asymmetrically dependent on Elms is the fact that they produce a "look" like that of an Elm (which, albeit they wouldn't produce were it not that Elms produced the same "look"), then why not make some description of that *"look"* the content of the mental representation rather than either the disjunction of the two distal objects or the more foundational of the two objects in the dependency relation? Fodor anticipates this criticism. He makes the same point during the course of considering how we form theoretical concepts such as that of "proton," the psychophysical properties of which might include things like the "look" of a photographic plate, for example.

> We've got something into the belief box for which instantiations of *proton* are causally responsible; but it's the wrong thing. It's not a token of the concept PROTON; rather, it's a token of some (probably complex) *psychophysical* concept, some concept whose tokening is lawfully connected with the look of the photographic plate .... Something needs to be done about this. Here's where the cheating starts.[16]

Fodor attempts to circumvent this problem by adding a restriction to the account of asymmetric dependence given above. In particular, he now requires that the psychophysical properties which are responsible for the reliable tokening of mental symbols in us are *in fact* causally connected to the presence of the objects which the symbol denotes. So, in the case of | Elm |s, | Elm | will represent Elms just in case the psychophysical properties which cause the tokening of | Elm | are themselves caused by Elms. The requirement of a causal connection between the set of psychophysical properties which provokes the mental representation and the object which instantiates the set of psychophysical properties reestablishes a reliable link between object and mental representation. However, in so doing, it seems to compromise the main advantage of Fodor's theory of error. For suppose, as we have been doing, that the psychophysical properties causally responsible for my tokening of | Elm | can themselves be caused by both Elms and Oaks. Which of these objects is the cause of my representation? Insofar as both produce the psychophysical properties which are causally responsible for my representation, either one ought to be an acceptable answer. It really is of no consequence

[16] Fodor (1987), p. 120.

that in any given case, only one *actually did* cause these psychophysical properties to be instantiated. For one thing, one would have absolutely no way of knowing this in cases where the set of psychophysical properties produced by either candidate were sufficiently similar. But more importantly, it isn't clear what we'd gain (apart from salvaging Fodor's version of causal theory) by reconnecting the more proximal stimuli with a distal object and subsequently assigning content to it. In fact, I think there is a strong case to be made for the claim that there are a number of advantages to avoiding this move and to deemphasizing the role of objects in fixing content in favor of the (specifically structural) properties which they exhibit. A closer consideration of Cummins' interpretational semantics will begin to lay the groundwork for this argument.

1.2.2 How Cummins Handles Error

The causal theorist's dilemma is not unique. It is common to encounter the problem of being unable to account for how representations acquire their content, while at the same time providing room for the possibility of representational error. Cummins will try to avoid the dilemma by making what he views as a crucial, but commonly ignored distinction between the *content* of a representation and its *target*.

1.2.2.1 Targets vs. Contents: An Example

While programming an application designed to average grades a few years ago, I encountered a troubling problem whose solution seemed to elude me. The program was designed to compute the average of 5 grades. It worked by prompting the user to enter a grade, storing the grade in a data structure called an array, and prompting for the next grade until the array was full. When the array was full, the application added the values stored in it together and divided the sum by the size of the array (i.e., the number of elements it contained). Although whole arrays may be treated as the objects of computations in a program, they are usually used to store values which are themselves the objects of computations, as was the case with my application. For this reason, the values stored in an array must be "indexed" or given a reference, so that these values may be retrieved by the program when they are needed. If the number of values to be stored in an array is known in advance, an array of that size can be created to store the values. The number corresponding to the size of the array is then a *representation* of the number of values stored in it and can be used accordingly. Knowing that arrays behaved this way, I created the following algorithm for my grading application:

```
1: Create a data object called SizeOfArray = 5.
2: Create an array of size SizeOfArray and call it ArrayOfGrades.
3: Create a data object for temporary storage of the current grade called CurGrade.
4:
5: Create a data object which keeps track of how many grades have been entered called
6:      NumGrades. Set NumGrades to an initial value of 0 to indicate that no grades
7:      have been entered.
8:
9: Create a data object called Sum which keeps track of the sum of the entered grades.
10: Set Sum to an initial value of 0.
11:
12:     while NumGrades < or = SizeOfArray
13:     {
14:        Prompt the user for the next CurGrade.
15:        Store CurGrade in the NumGrade index of the array.
16:        Sum = Sum + CurGrade
17:        Update NumGrade          //add one to the value of NumGrade
18:     }
19:
20: Create a data object called Average to store the average of the grades entered.
21: Average = Sum / NumGrade
22: Display Average
```

Figure 1-A: Grading Algorithm

Unfortunately, this grading algorithm didn't work very well. It seemed to produce lower averages than it should have produced and it prompted me for six grades rather than five. To see why the algorithm produced this error, study the trace of the program for the following set of 5 grades: {67, 98, 87, 56, 71}. Note that this trace begins with a representation of the values of the data objects *after* at least one cycle through the "while" loop (see line 12). The value of *Average* is given at the end of the trace.

SizeOfArray = 5
ArrayOfGrades =

67				

NumGrades = 1
CurGrade = 67
Sum = 67
Is *NumGrades* < or = *SizeOfArray*? Yes

SizeOfArray = 5
ArrayOfGrades =

67	98			

NumGrades = 2
CurGrade = 98
Sum = 165
Is *NumGrades* < or = *SizeOfArray*? Yes

SizeOfArray = 5
ArrayOfGrades =

67	98	87		

NumGrades = 3
CurGrade = 87
Sum = 252
Is *NumGrades* < or = *SizeOfArray*? Yes

SizeOfArray = 5
ArrayOfGrades =

67	98	87	56	

NumGrades = 4
CurGrade = 56
Sum = 308
Is *NumGrades* < or = *SizeOfArray*? Yes

SizeOfArray = 5
ArrayOfGrades =

67	98	87	56	71

NumGrades = 5
CurGrade = 71
Sum = 379
Is *NumGrades* < or = *SizeOfArray*? Yes

SizeOfArray = 5
ArrayOfGrades =

67	98	87	56	71

NumGrades = 6
CurGrade = ?
Sum = 379
Is *NumGrades* < or = *SizeOfArray*? No (terminate loop and calculate *Average*)

Average = 379/6 = 63.2

Table 1-A: Trace of Average Application

The problem here, as the trace demonstrates, is that the condition governing the number of times the application will traverse the loop (i.e., *NumGrades* < or =

SizeOfArray—line 12) allows the application to traverse the loop one too many times.[17] As a result, the value of *NumGrades* becomes 6 and fails to represent what we wanted it to represent, namely, the number of grades to be averaged by the application.

This example is useful for illustrating Cummins' conception of misrepresentation. For him, a cognitive system Σ is in error when "... there is a mismatch between the state of affairs Σ needs to represent when it tokens [a representation *r*] ... and the state of affairs ... [*r*] actually represents ...."[18] The state of affairs that Σ "needs" to represent is called its *target* while the state of affairs that Σ actually represents is called the *content* of Σ's representation.[19] In our example, the system Σ which computes the grading algorithm "needs" to represent the actual average of the five grades entered. The average of these grades is the *target* of a tokening of *Average*. Nonetheless, what *Average* represents *in fact* (i.e., the actual representational *content* of *Average*) is the sum of those grades divided by 6. The tokening of the representation *Average* is in error, "*...when the target of tokening it on that occasion fails to satisfy its content*"—that is, when the target of the token and the representational content of the token are not the same.[20] Cummins writes:

---

[17] Actually, this algorithm would probably cause the computer to crash if implemented in some common programming languages, since it would attempt to insert the value of *CurGrade* in the last loop traversal into a non-existent index of the array, but we will ignore this complication for the meantime, since the example can be used to illustrate another kind of error closer to misrepresentation. Also note that pointing to the loop condition as the source of the error is a little bit misleading since it is only one of at least two possible sources of the algorithm's failure. Had we indexed the array starting at 1 instead of 0 (see the initialization of *NumGrades*—line 6), this error would not have occurred.

[18] Cummins (1996), p. 6.

[19] In what sense a system can "need" to represent one thing rather than another is something which needs explaining. I'll show how Cummins attempts to naturalize this notion.

[20] Cummins (1996), p. 6.

> It is precisely the independence of targets from contents that makes error possible. If the content of a representation determined its target, or if targets determined contents, there could be no mismatch between target and content, hence no error. Error lives in the gap between target and content, a gap that exists only if targets and contents can vary independently. It is precisely the failure to allow for these two factors that has made misrepresentation the Achilles heel of current theories of representation.[21]

The distinction between targets and contents is a nice distinction particularly when attending to representation and error as they are understood in computer science. For computer scientists, a program can perform incorrectly when one or more of the data structures it utilizes fails to represent what the programmer intended it to represent. Moreover, it is possible to think of this failure in performance as consisting in a relationship between the *actual* representational content of a data structure and its *intended* representational content, or target. After all, the algorithm described above would perform correctly if the target of the representation *Average* was indeed the result of dividing *Sum* by the number 6.[22] It is because the target of the *Average* is *Sum* divided by 5 that we say the algorithm performs incorrectly or is in error.[23]

A distinction similar to the one between content and target can be found in

---

[21] Cummins (1996), p. 7.

[22] Note that the fact that the representation has been referred to as an "Average" would not prohibit it from having the target *Sum*/6 in this algorithm since referring to the data structure in this manner is merely the programmer's way of reminding herself of what *Average* represents and not a condition which controls its actual content.

[23] This kind of error may seem more like an error in the *performance* of the system, rather than an error in representation. In fact, Cummins solution to the problem of error *is* designed to explain performance error, not representational error. I will say more about this presently, but for now, note that Cummins himself is aware of the distinction between performance and representational error, makes the observation elsewhere (Cummins (1996), p. 99) that his account is more adequate as an explanation of the former, and is perfectly comfortable with the limitations which accompany this observation (Cummins, (1996), p. 102).

discussions of *perceptual* error. In the case of perception, we can distinguish the content of the perception from the object being perceived. In fact, it seems arguable that, as with representation, the very *possibility* of perceptual error turns on just such a distinction. To revisit an earlier example, when I misperceive an Oak as an Elm on a dark night, the content of my perception is | Elm |, but the object that is actually being perceived (read: the *target* of my perception) is an Oak tree. Indeed, there does seem to be a distinction between perceptual content and perceived object that is closely related to the distinction which Cummins attempts to articulate between representational content and target.

This observation raises a number of interesting issues for Cummins' discussion of the distinction between contents and targets. For example, how do we distinguish the content of a representation and its target? Or alternatively, how are targets specified? In the Elm-Oak example, it seems arguable that the target of a representation is equivalent to its *cause* while the content of a representation may have nothing to do with the cause at all, or may be related only asymmetrically to it. But unlike Fodor, Cummins does not avail himself of a causal account of target fixation so he will have to propose an alternative method for making the distinction. In addition, whatever theory of target fixation we adopt must imply or—at the least—be compatible with a completely naturalized conception of "target." And finally, we have intimated (two paragraphs back) that Cummins' notion of target is more useful for explaining performance error than for explaining representational error. What reasons are there for accepting an explanation of performance error as adequate for a theory of error? Do we really need a notion of representational error, or is performance error good enough?

1.2.2.2 Performance Error vs. Representational Error

Cummins acknowledges that when representations are misused (as in the case where a representation of *Sum*/6 is used to signify *Average*), error certainly occurs, but not representational error. For example, a map of the Upper Peninsula of Michigan can contribute to error if the map's user wrongly takes it to be a map of Florida, but the error is not representational—the map represents the U.P. accurately. Instead, the kind of error involved is a kind of "performance error": error which is the result of the misapplication of representations which accurately signify states of affairs external to the representing system.[24] Indeed, Cummins argues that we should "...resist any formulation that forces us to infer misrepresentation from a mere failure to perform, since misrepresentation is compatible with—sometimes essential to—successful performance, and failed performance is compatible with accurate representation."[25] But if misrepresentation and failure to perform (i.e., failure to apply representations correctly so that the system performs in the way expected) are not the same, then Cummins' account of error is not an account of how misrepresentation is possible, but rather, of how cognitive systems fail to perform when correct representations are misapplied. Cummins thinks that this kind of account of error is good enough. That is, any case where before it seemed we needed an account of representational error will now be explained by citing a mismatch of

---

[24] Even though Cummins terms such representations "accurate," the notion of an "accurate" representation is somewhat out of place in an account which does not allow for representational *in*accuracies. It would probably be better (and just as effective) to say, "... the misapplication of representations which map isomorphically to states of affairs external to the representing system."

[25] Cummins (1996), p. 99.

representational content with intended target. In short, *performance* error will do the job previously done by the concept of representational error. To be clear, Cummins himself does repeatedly describe his as a theory of *representational* error. However, as we shall presently see, it seems clear this his account is perfectly compatible with the view that the contents of one's representations are always "correct"—that is, that they will never fail to correspond with their content. The account simply explains the case in which a "correct" representation is misapplied.

In the end, it may make little sense to call a representation "correct" under these circumstances at all. More likely, representations are not properly thought of as correct or incorrect in the context of an account like Cummins'—instead, there simply is or is not a correspondence between some potentially representing structure[26] and another, potentially represented object. When there is such a correspondence, representation occurs. When there isn't representations aren't "incorrect," rather, they just aren't there at all.

There may be good reasons for allowing performance error to do the work that misrepresentation is normally thought to do. Let's examine the way in which Cummins makes use of targets a little more closely in order to understand what some of these reasons might be.

---

[26] If Cummins is right and a structure represents in virtue of bearing an isomorphic relation to another structure, then any given structure ought to *actually represent* any number of other structures. I use the phrases "potentially representing structure" and "potentially represented structure" here to refer to the possibility that some mental structure represents some *particular* external structure. In other words, while all mental structures will be isomorphic to some structure in the world and are thereby *actually representing structures*, few if any of them will bear this relation to everything. It may need to be established that a mental structure is in fact a representation of some specific external structure in which we are presently interested.

1.2.2.3 How Are Targets Fixed and Can They Be Naturalized?

Following Cummins, I have been speaking of the target of a representation as that which the system "needs" to represent in order to avoid performance error. As pointed out above, this is a decidedly unnaturalistic way of speaking about target fixation since it seems to imply that identifying error requires access to the system's intended goals. Later in his book, Cummins attempts to explain what he means by what a system "needs" to represent more naturalistically. Cummins claims that, "the target of tokening [a representation] *r* is what [a cognitive system] Σ expects to find when *r* is accessed."[27] What the system "needs" or "expects" is not a matter of the system's intended goals, nor is it a matter of the intended goals of the system's programmer. Rather, what the system "expects" to find is "... a matter of [the system's] architecture or design."[28] The idea is that representing systems, and cognitive systems in particular, incorporate mechanisms whose function it is to represent objects or sets of objects in the external world.[29] Cummins cites the example of a visual module, whose function it may be to represent sizes, shapes, and edges. Targets are determined, he claims, *by the function of the mechanism in question together with the current state of the world.* Mechanisms which have the kind of representational function described here are called *intenders.*

---

[27] Cummins (1996), p.18.

[28] Cummins (1996), p.18.

[29] Cummins (1996), p. 8.

> The conception of target fixation I want, then, is the conception ... that the target of tokening *r* is what *Σ* expects to find when *r* is accessed. Not literally 'expects' of course; the idea is that *Σ* incorporates a design assumption to the effect that a representation generated by a certain intender will be a representation of *t*. This is ... the conception that falls naturally out of thinking of intenders as mechanisms for binding programming variables. Just as programs are designed around assumptions about what will be accessed when a given variable is evaluated, cognitive systems are designed around assumptions about what will be represented by various intenders. Specifying a system's design or functional architecture involves specifying what intenders are possible, and hence, on the current conception, what targets are possible. The targets that are possible for a given system are thus fixed by its functional architecture.[30]

But is the concept of "function" as it is utilized here a naturalizable concept? And if so, will function together with the "current state of the world" be enough to explain why, for example, the Average program failed? More important, will it be enough to describe its failure as representational error rather than as performance error and is such an explanation even necessary?

If it is true that targets are determined, in part, by what functions the mechanisms in cognitive systems have, and if a system's functions are derivative of its "architecture or design," then we may be on our way to a more naturalistic account of target fixation. However, there seems to me to be a significant difference between saying that the function of a representing mechanism is determined by a system's *architecture* and saying that it is determined by a system's *design*. The difference lies in the distinction between two distinct and independent senses of the term "function." When Cummins claims that some representing mechanisms in a system (namely, intenders) have "functions" which specify their targets, he might mean that they actually compute formal mathematical functions. Although it is somewhat misleading to say in this circumstance that the mechanism "has"

[30] Cummins (1996), p. 18.

the function in question, we will suppose that the relationship the mechanism bears to the function it computes (in virtue of the fact that it computes it) is enough to make sense of this terminology. On the other hand, Cummins might mean that mechanisms acquire "functions" in a more teleological sense of the term. Perhaps systems subject to evolutionary development, for example, acquire mechanisms whose function is to promote survival. Let us signify that we use the term "function" in the former sense by writing it as "$function_1$" while signifying its use in the latter sense with the notation "$function_2$."

Now, if a system's mechanisms have $functions_1$ and they can be used to determine the targets of the system's representations, then target fixation does seem to be a matter of a system's architecture. This would be consistent with a fully naturalistic account of target fixation, since one has only to describe the formal properties of the system in order to understand its $functions_1$. The problem with understanding the function of a mechanism this way however, is that it cannot provide for a distinction between targets and contents—a distinction which Cummins needs if he is to retain his account of cognitive error. To see why this is so, consider the grading algorithm described above: Conceive of the algorithm as a simple system whose data structures are functioning mechanisms. In particular, consider the mechanism, *Average*. If we want to know the target of *Average*, then we need to know what its $function_1$ is (i.e., its formal description as implied by the system's architecture). To obtain this, we examine the algorithm and find that *Average* $functions_1$ to calculate the *Sum* divided by 6. When the program is executed, moreover, the content of *Average* is also the *Sum*/6—this is what *Average* in fact represents. The problem with conceiving of an intender's function as a function in the formal mathematical sense, as this example illustrates, is that this is the same strategy employed, at least on Cummins' view, for obtaining the representational content of an intender. Hence, doing so will not allow for a distinction between target and content.

Suppose, instead, that it is the $function_2$ of a mechanism which determines its target. If the content of the representing mechanism is determined by its $function_1$, but the target of the mechanism is determined by its $function_2$, then we can re-establish the distinction between target and content Cummins believes necessary for a successful account of error. However, unless the origins of $functions_2$ are explained without an appeal to the intended goals of the system and/or without appeal to the goals of the system's designer (a difficult task in the case of the grading algorithm, to say the least), this approach threatens the project of providing a fully naturalistic account of representational error, and by association, of representation itself.

Accounts which attempt to provide a naturalized account of a system's functions are readily available.[31] And although Cummins' notion of "function" turns out to be something close to $function_2$, he is not satisfied simply to help himself to one of them by way of offering up a solution to the dilemma illustrated above. As he sees it, there are at least two reasons why a typical rendering of the notion of "function" (as understood especially in contexts such as functional and adaptational role theory) will not suffice for the problem of describing how targets are fixed. First, Cummins contends that most understandings of function have it that only event *types* can have functions—that is, once

[31] See, for example Ruth Millikan's "Biosemantics" in Journal of Philosophy, Vol. 86, No. 6, 1989.

the function of a thing or event has been determined, all things or events of that general type must have the same function. The notion of function is not, as ordinarily understood, something which can vary from token to token:

> ...available theories of functions apply to events only as instances of types: 'the function of *x* is *f*' is defined in terms of what *x* does Normally (Millikan 1984) or typically or ideally, or what *x* does whenever some specified condition holds. But it is essential to the current project [i.e., the project of explaining target fixation in a way that is consistent with the positive theory of interpretational semantics] that different tokenings of the type "tokenings-of-*r*" can have different representational functions, for *r* can be applied to different targets on different occasions. Not every tokening of | elm | has as its function representing elms, for, if this were the case, there could be no error. Representational error arises when an | elm | is applied to something that isn't an elm, for example, when the function of a tokening of | elm | is to represent a beech or the number 9 or the proposition that roses are red. So we need a theory of functions that allows for the function of tokening *r* to differ from occasion to occasion.[32]

In other words, since targets *can* vary from token to token, either we cannot make target and function synonymous, or we must find an alternative notion of function which allows us to identify functions at the token-level. Second, Cummins believes that most "over-the-counter" analyses of function equate successful representation with correct functioning. But, as we have already seen, Cummins does not equate successful representation with a content-target match. Therefore, if the appropriate target of a representation is going to get determined by the function of that representation, we must have a notion of function which does not immediately imply a kind of success.

> ...we need an analysis of functions ...[which] allows for the fact that representational success is neither necessary nor sufficient for the nonrepresentational success of representational systems. Target Fixation will be incorrect if we expand it via a theory of functions that yields the

[32] Cummins (1996), pp. 113-14.

implication that cognitive processes require representational correctness to function properly. Representational error can be quite compatible with success.[33]

Cummins circumvents the first of these two difficulties by proposing that we give up the idea that representations themselves ever have the function of representing a particular target. Instead, Cummins argues that substructures of the representing system called *mechanisms* have the function of representing one or another particular target, and that representations *inherit* this function when they are tokened by the relevant mechanism. Mechanisms which function like this are referred to elsewhere by Cummins as "intenders" (see above) and when they function specifically to represent a target *t* he calls them "*t-intenders*." Since mechanisms having very different functions might token the same type of representation, the representation in question can have one function/target when tokened by one mechanism and a very different function/target when tokened by another. Similarly, the same mechanism might token different types of representations, leading to a case in which different types of representations have the same function/target.

This raises the question of how to understand a mechanism as having the function of representing some target or other. Cummins notes that if we are to satisfy his second criterion for the analysis of function, then we must not understand it in a way that equates

[33] Cummins (1996), p. 114. Cummins argues elsewhere at length for the point that representational error can be compatible with successful performance. (See pp. 27ff, 44-47, 50, 99, 116-118.) For example, on p. 27 (footnote) he states: "Less accurate representations are often tolerable because they are less costly to compute. Misrepresenting a crow as a hawk is a less serious error for a field mouse than misrepresenting a hawk as a crow. Given that recognition must occur quickly, a fast but inaccurate system may be better than a slower more accurate one. Since crows greatly outnumber hawks, a fast system that generates many false-positive hawk indentifications but no false negatives is less accurate but more effective than a slower system that generates fewer false positives while still avoiding false negatives."

the notion of function with anything ensuring the correctness of the representation on the part of the system which incorporates the mechanism.

> Our problem is this: how can it be the function, or a function, of a mechanism to represent *t* [the target] if neither it nor its ancestors have actually succeeded in accurately representing it? The key to understanding this issue is to realize that it is the (or a) function of [a mechanism] *N* to produce representations of *t* if it is the representational relation—the degree of fit—between *t* and the tokens *N* produces that underlies *N*'s contribution to the system that contains it. To get a feel for this, we may imagine the situation in which a map *M* is completely accurate for *m*1, but used by [a system] Σ to get around the somewhat different *m*2. To understand Σ's success and failures in negotiating *m*2, and *M*'s contributions to these, we have to know how accurately *M* represents *m*2, since Σ's errors are measured relative to *m*2, *M*'s target on the occasions in question. Hence Σ's capacity to negotiate *m*2 as well as it does cannot be understood in terms of its capacity to perfectly represent *m*1, even though any representation that represents *m*1 perfectly will in fact represent *m*2 accurately enough (and in just the right way) to explain Σ's performance in *m*2. For while it is true and lawlike that Σ performs as it does in *m*2 because it perfectly represents *m*1, this holds only because of the relation between *m*1 and *m*2. To assess Σ's performance in *m*2, we need somehow to assess Σ's representation of *m*2. We can do this by assessing Σ's representation of *m*1, but only if we are already in a position to assess how well *m*1 represents *m*2. Since *M* and *m*1 are isomorphic by hypothesis, assessing the representational relation between *m*1 and *m*2 is equivalent to assessing the representational relation between *M* and *m*2.[34]

What Cummins provides here is a case in which the system Σ exploits a representation in the service of negotiating some real-world problem—we might imagine that navigating the city of East Lansing is the problem which Σ wishes to handle. The representation *M* which Σ exploits bears a relationship to the city of East Lansing (*m*2), but it just so happens that it bears an even better (read: more accurate to the point of isomorphism) relation to the city of East Splansing (*m*1). Cummins' account of target fixation requires that East Lansing is the target of Σ's representation on this particular occasion, because it is the city of East Lansing that Σ's representation functions to navigate. But the problem is this: How can East Lansing be the target of *M* when *M* doesn't accurately represent East Lansing, and in fact, *does* accurately represent East Splansing? Why is the target of *M*

[34] Cummins (1996), p. 117.

East Lansing (*m*2) and *not* East Splansing (*m*1)? Cummins argues that *m*2 can be the target of *M* in virtue of the fact that there is some degree of fit between Σ's representation and the actual city of East Lansing that informs Σ's behavior. In Cummin's terms, it is the existence of some "fit" between representation and represented thing which "underlies [the mechanism which produced the representation] *N*'s contribution to the system that contains it."[35]

This move sounds a lot like Fodor's asymmetrical dependence to me, although I can't be certain how Cummins actually makes use of the similarity between *m*1 and *m*2 the way it seems obvious that Fodor does. Perhaps one way of describing Cummins' move is that he makes *m*1 (in virtue of the fact that is bears the closer (isomorphic) relation to *M*) the primary player in a Fodor-like asymmetric dependency relation. Something like, "*M* can only function to assist in the navigation of East Lansing (*m*2) because *m*2 *looks like* East Splansing (*m*1)." Translation: *M* can only have *m*2 as a target because this target *looks like* the content of *M*. If Cummins is attempting to make a move which employs a notion similar to asymmetric dependence, then his characterization of target fixation is subject to criticisms like those made against Fodor. I leave it to the reader to determine this.

Let's step back to consider what this way of understanding target fixation accomplishes and what it does not. We have already seen that the introduction of the notion of a *mechanism* or "*t-intender*" makes it possible for Cummins to avoid understanding functions as pertaining only to event and thing types. This latest move would also seem to lay the groundwork for the possibility of maintaining a distinction

---

[35] Cummins (1996), p. 117.

between accurate/successful representation and the proper functioning of a system's mechanisms, resulting in the conclusion that mechanisms can have functions/targets which do not require (totally) successful/accurate representation. For while *m*2 is the target in this example, it is not the most accurate (read: most closely mapped to *M*) of the available objects which stand in relation to *M*. So it seems that Cummins has met both of the conditions he set out to meet when seeking to fix targets via the identification of a system's proper function.

Notice that what still cannot be distinguished are the functions/targets of a system's mechanisms and what promotes the successes and/or goals of the system as a whole. Cummins has shown that "accurate" representation may not always be required for what maximizes a system's benefits, but this point necessarily involves the retention of the notion of what is beneficial to the system, and it remains unclear how *this* can be understood naturalistically. In Chapter 6, I'll revisit this issue in more detail. For now, it will have to be sufficient to note that there may be questions which a theory seeking to naturalize representation cannot *in principle* address naturalistically—which is to say that it cannot address these questions at all.

Cummins has made a significant contribution to the project of providing a naturalized and highly formalized account of mental representation. In addition to reviving what many thought to be a discarded form of representation theory namely, picture theory, Cummins has updated the view and shown how it can be used in the service of providing a solution to one of the most challenging problems of representation theory: the problem of error. However, the review of his work provided here has demonstrated, I believe, that there are still some significant issues to be addressed if a modern form of the picture theory of representation is to be taken seriously. Among these issues is the problem of uniqueness, which, as we shall see, influences everything from the fixation of representational content to the identification of the targets of our representations. Therefore in the remainder of this chapter, I will attempt to set out in greater detail the particular challenges which the uniqueness problem raises.

1.3 The Uniqueness Problem

Perhaps the biggest advantage of interpretational semantics rests in the sheer simplicity of its account of mental representation. A representation represents in virtue of bearing a relation to some object or state of affairs which shares its structure—plain and simple. But what makes interpretational semantics so attractive as an account, may also be the source of its greatest weakness. This is because the simplicity of the account may allow that mental structures represent all kinds of things—indeed, more than one thing at a time. If the content of a mental representation is just an external object with which it shares a structure, which external object should constitute the content of a mental representation when it is isomorphic to more than one? Shouldn't mental representations map onto external objects *uniquely* if they possess the content that they do exclusively in virtue of this mapping?

It is important to recognize that this is not the same problem that we have been recently examining when considering the manner in which targets are fixed. The problem of target fixation is a problem about which things a system *intends* or *functions* to represent in any given instance—not about which things a system *actually does* represent. When we examined the problem of target fixation, we assumed, for the moment, that there was no problem about which things the system actually did represent. We start out by assuming that this much is clear: Σ possesses a representation *r* with the content C. The problem of targets is the problem of figuring out whether it was C which Σ was *supposed* to represent on this occasion or not. In contrast, the problem of uniqueness is the problem of knowing what the content of *r* is in the first place.

There are a number of ways in which a "uniqueness" problem can arise for interpretational semantics and for picture theories of representation in general. I have provided diagrams of three varieties of the problem below. Only one of these versions of the problem is addressed explicitly by Cummins in Representations, Targets, and Attitudes. The others are acknowledged by other authors in other contexts such as measurement theory and abstract algebra. For the moment, I will primarily address the threat to picture-theory-style semantics which is posed by Cummins' own version of the problem and will comment only briefly on what I consider the "standard" version of the problem. Ultimately, I will argue that some of these versions of non-uniqueness are not legitimate concerns for the interpretational semanticist.

Cummins recognizes that there is a "uniqueness problem" for representational content when it is accounted for in the way he (and picture theory in general) propose to account for it. He describes the problem as follows:

> ...isomorphisms, where they exist at all, are not unique. A structure *R* may be said to represent another structure *C* wherever *R* is isomorphic to *C*. In general, however, there are bound to be many isomorphisms, that is, many one-to-one structure-preserving mappings, between two

structures if there are any. Hence, when *R* is isomorphic to *C*, there are many different ways in which each structure can represent the other.[36]

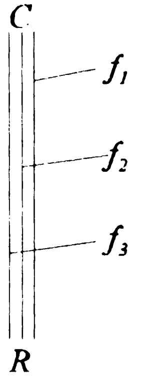

Figure 1-B: Non-Uniqueness—Cummins

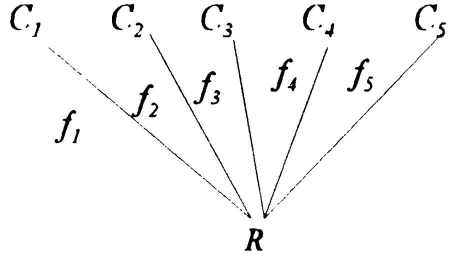

Figure 1-C: Non-Uniqueness—Standard

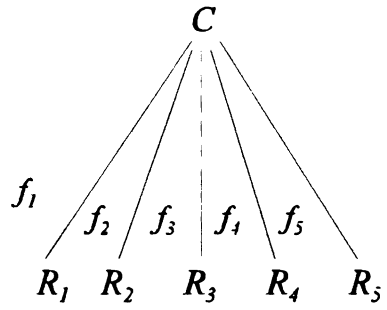

Figure 1-D: Non-Uniqueness—Measurement

---

[36] Cummins (1996), p. 97. Note that it is odd for Cummins to speak of isomorphisms "where they exist at all" since isomorphisms are actually quite common. For example, for any given structure, an isomorphism can be established between its elements and the set of natural numbers (more on the threat this type of isomorphism poses to a Cummins-like account in Chapter 5). Cummins seems to be focused on a very specific type of uniqueness problem wherein *if* any two objects can be said to be isomorphic in virtue of the existence of some function mapping the elements of one onto the elements of the other, then it seems that not only that function, but an indefinite number of other functions ought to exist which map the two objects to one another. In other words, for Cummins, the problem seems to be that there is more than one mapping between the same two objects. A more serious form of the problem arises when more than one external object can be mapped to the same mental structure (see **Figure 1-C**). This would occur whenever there is an isomorphic relationship between the mental structure and more than one external object. Although Cummins' interpretational semantics is threatened by this sort of non-uniqueness too, he only addresses the former variety.

For Cummins, the uniqueness problem arises because it is possible for there to be several isomorphic functions which constitute a mapping between two structures: *R,* a representing structure and *C*, a represented one. His version of the problem might therefore be represented graphically as shown in **Figure 1-B** (where each *fi* is an isomorphism of *R* onto *C*).

To illustrate this brand of non-uniqueness, Cummins imagines a mechanical car capable of navigating a maze called the "Autobot." which is. The Autobot is crafted with cog wheels on the rear axle which pull a card parallel to the ground through the car as it moves. In the card is a jagged slit. A pin on the tie rods controls the steering of the car as

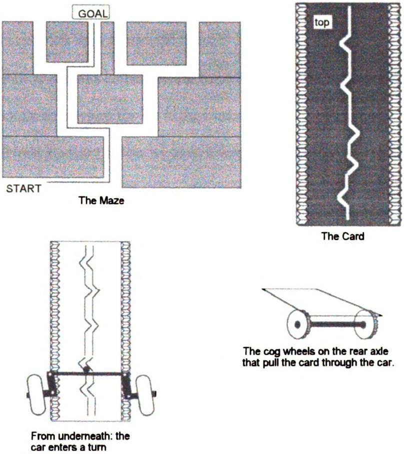

Figure 1-E: Autobot

it passes through the slit (see **Figure 1-E**).[37] The idea is that each left hand groove in the card represents a left hand turn in the maze, while right hand grooves represent right hand turns. Supposing that the card is designed in such a way as to represent the turns of the maze, the car should navigate the maze successfully when its motion is governed by the card. Of course, the card represents the maze if and only if there is an isomorphic relationship between the slit in the card and the path through the maze. But there is such a relationship—let us refer to this isomorphism as f_1.

The uniqueness problem arises when it is observed that f_1, which promotes a successful navigation of the maze by the Autobot, is not the only isomorphism which exists between card and maze. Imagine that the card was inserted into the Autobot upside down such that now left hand grooves in the card are mapped to right hand turns in the maze and right hand grooves in the card are mapped to left hand turns in the maze. This too is an isomorphism between the slit in the card and the path of the maze—let us call this isomorphic relationship f_2. Now we have a version of non-uniqueness which matches Cummins' description of the problem (see **Figure 1-F**).

---

[37] reprinted from Cummins (1996), p. 95 with permission from The MIT Press.

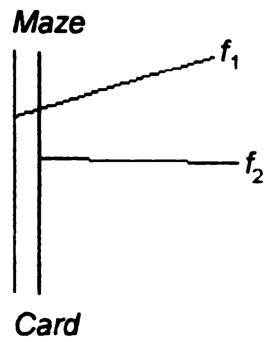

**Figure 1-F:
Non-Uniqueness in the
Autobot**

Obviously, f_1 promotes the successful navigation of the maze by the Autobot whereas f_2 does not. However, f_2 is certainly isomorphic to the correct path (just as f_1 was) since it maps the slit in the card to its mirror image. Since both functions establish an isomorphism between the slit in the card and the path through the maze, but only one promotes its successful navigation, it may seem that both functions could not serve to specify the representational content of the card. In fact, it seems (if the goal is ultimately to represent the maze) that f_1 produces an accurate representation, while f_2 promotes representational error.[38] This may make it seem like we need a criterion for choosing f_1 over f_2 rather than be satisfied with the mere existence of an isomorphic relationship between the card and the maze. Indeed, Cummins believes that it is tempting to see the *only* solution to this problem as consisting in the development of a criterion for choosing between competing isomorphic mappings.

---

[38] Or perhaps it is simply performance error that f_2 promotes—that is, behavior which is inconsistent with the overall goals of the system or of the system's programmer. This is the conclusion you would reach if you accept Cummins' strict distinction between targets and contents.

Of course, when considering which of these mappings identifies the *target* of the card's configuration—the one which maps left turns in the card to left turns in the maze, or the one that maps the card's left turns to the maze's right turns—we *do* have to make a choice. Since both mappings do not promote the Autobot's successful navigation of the maze equally well, and since we can presume for the sake of argument that successful maze-navigation is the goal of the system, one mapping must fail to match the representational content of the card with its target. But Cummins argues that no such choice is required in order to know the *content* of the card's configuration. There are any number of things which are represented by the slit in the card. What explains the ability of the Autobot to exploit its representations nonetheless is that the target is *among* them.[39]

Cummins' strategy for overcoming problems of non-uniqueness is not, therefore, to show how to eliminate all but one of those mappings which provide a representation with the content that it has. Rather, Cummins' strategy is to show that non-uniqueness (of the variety described in **Figure 1-B** anyway), does nothing to inhibit the explanatory power of interpretational semantics, which turns out to be more concerned with how well a system's representations match up with their targets than with whether there might be other things that the system's representations could have represented equally well.

> If your map is not properly oriented, you won't get to your destination, but *that is not the map's fault*. The correct information is there; you just do not or cannot exploit it.... Grounding representation in isomorphism entails, as we have seen, that representational content is never unique. This non-uniqueness, however, does not undermine the explanatory value of representation in any way, for the explanatory work is done by the match (or mismatch) between content and target. The fact that a representation correctly represents many things other than the target has no tendency to devalue the important fact that it does (or does not)

[39] Cummins (1996), p. 99.

> represent the current target, for it is the match or mismatch (or rather the degree of match) between target and content that bears an interesting explanatory relation to performance. [40]

But how the distinction between target and content comes into play here may reveal a difficulty in Cummins' account. The Autobot illustrates a situation in which a single representational structure and a single content structure are related by more than one isomorphic mapping relation. According to interpretational semantics, nothing more than the *existence* of an isomorphism relating *R* to *C* is necessary for saying that *C* is the content of *R*. This would make it difficult to see how a mismatch between the content of *R* and its target could ever arise. In this scenario, there is only one content structure that can serve as the target. And there is nothing in the interpretational semanticist's account of content fixation which would imply that the content of *R* ever changes either. It seems to me that, the only way to conceptualize mismatches between target and content in the interpretational semanticist's view is to allow that either (a) there are several content structures available to which a single representation can be mapped, or (b) that there are several representational structures which are related to a single content structure. Given

---

[40] Cummins (1996), 99 & 102. It may seem that Cummins is alluding to something more like the standard version of the uniqueness problem (**Figure 1-C**) when he says: "The fact that a representation correctly represents ***many*** things other than the target has no tendency to devalue the important fact that it does (or does not) represent the current target, for it is the match or mismatch (or rather the degree of match) between target and content that bears and interesting explanatory relation to performance" [emphasis mine]. However, I think this is not what Cummins had in mind here. First, one cannot read Cummins' notion of "correctly represents" as equivalent to "has as its representational content" (which is how the comment above must be read if it is to be interpreted as alluding to the standard problem of uniqueness as I've characterized it) since the content of a representation is not technically correct or incorrect on Cummins' view—only the relationship between it and the target of the representation can give rise to error (and hence to correctness). Second, Cummins' notion of "correct representation" is one which only partly appeals to the notion of representational content. Representational correctness, for Cummins, occurs ***when content and target*** are consistent. But reading the passage above with this in mind makes the claim that "a representation correctly represents many things other than the target" nonsensical. Therefore, I think it best to assume that Cummins meant to describe the same manifestation of the problem he has been addressing all along.

that this is so, it is curious that Cummins does not spend more time on the sorts of uniqueness problems illustrated in **Figures 1-C** and **1-D** respectively.

Cummins' version of the uniqueness problem cannot be "solved" by interpretational semantics if the criterion for a solution is to be able to specify some way of picking one mapping, responsible for giving *R* content, over all of the others. The reason that no such criterion can be specified by interpretational semantics is that it treats isomorphism as a *sufficient condition* for representational content. What this implies is that the interpretational semanticist cannot even *represent* distinct mappings between *C* and *R* let alone choose among them. The uniqueness problem, as characterized by Cummins, is probably intractable for interpretational semantics.[41]

But there are reasons to take heart despite this. I will argue that if a meta-system Σ cannot represent the difference between two mappings of *R* onto *C*, then, relative to Σ, there *is* no difference between them.[42] And as Σ acquires the capacity to represent differences between *C* and *R*, the picture starts to look more like the sort of uniqueness described in **Figure 1-C**, for example, than like Cummins' brand. In particular, as Σ acquires more information about the external world, *C* can been seen as part of a larger, more complex structure than the one of which Σ was previously cognizant and which *R* accurately represented.

Cummins worries less about the effect of the varieties of non-uniqueness on content

---

[41] Once again, this is so on the condition that your criterion for a solution is that some method be specified for choosing between the competing mappings. Perhaps it seems obvious that this is the only satisfactory way of responding to the uniqueness problem, but ultimately I will argue that it isn't.

[42] More on this in Chapter 5.

fixation than he should, but perhaps this is understandable given the de-emphasized role of representational content in explaining system behavior according to his theory. More important than fixing content for Cummins' view is the ability to fix targets, and ultimately, to fix them for the purpose of comparing them with the contents of a system's representations. But note that we will be unable to identify target-content mismatches if we are unable to specify the *content* of representations with equal acuity. In other words, if we cannot determine which external structure provides *R* with the content that it has in the first place, then we will not be in a position to compare the content of *R* with an appropriately chosen target. Even if Cummins is correct, therefore, in putting the explanatory burden on the relationship between targets and contents, rather than on representational content alone, this does not exempt one from the need for solving the uniqueness problem for representational content.

There is, I believe, a solution to the varieties of uniqueness problem for representational *content* and Cummins has most of the tools he needs to construct this solution. Most of the remainder of this dissertation is devoted to explicating and defending the solution I'll propose. Before concluding my summary of Cummins' own views however, I'd like to say a final word about targets and the possible disadvantages of relying on them so heavily in an account of error and representation.

1.4 Two Problems About Targets

As I see it, there are at least two problems with Cummins' use of targets (in addition to those we've already considered in detail above):

1. Degrees of Accuracy in Representation: Most people think that representation can occur in degrees. In fact, Cummins himself has insisted that a desideratum of a theory of representation is that representation admit of degrees[43] and mentions similar requirements on representation in this work. Nonetheless, it's hard to see how Cummins' use of targets will generate a theory of error which, among other things, accounts for representation by degrees. This is true in part because targets don't figure in to a legitimate theory of *representational* error but rather, account for performance error understood as what occurs when *correct* representations are misapplied. And as for representational *content*, since representations have content in virtue of being isomorphic with what they represent, there is a sense in which all representations are perfectly accurate, or better: a sense in which the notions of accuracy and inaccuracy simply don't apply.

2. Degrees of Completeness in Representation: Since representational correctness is achieved "... when the content of a representation is the *same as* the target of its application" targets will have to be structures too, for Cummins [emphasis mine].[44] If not, then there is no sense of "same as" that will account for the match up between target and content. However, if the representation relation consists in a strict isomorphism, then, to have a perfect match, for example, between the city, which is the target of my representation, and the map of the city, there must be an isomorphism between the city and the map. But this is surely requiring the unreasonable. Cummins is aware of this

---

[43] Cummins (1989) and Cummins (1996), pp. 107 & 108.

[44] Cummins (1996), p. 109.

difficulty when he writes:

> ...we speak *loosely* when we say, for example, that a map represents Tucson. Common sense recognizes this looseness by acknowledging such facts as that a topographical map and a street map of Tucson, while both maps of Tucson, nevertheless map different things. A street map abstracts away from many features of the city, including contours and altitudes, while a topographical map abstracts away from many features, including the layout of the streets. All forms of representation, except for particle-for-particle duplication, are abstract in this sense: they capture certain structures and are silent about others [emphasis mine].[45]

Let me begin with the second problem. Although I am comfortable with the notion that all representation is abstract, I believe it an undesirable consequence of what is contained above that all representation short of "particle for particle duplication" is representation only in a "loose" sense. I also believe this to be an entirely unnecessary consequence of a picture theory of representation. As we consider the uniqueness problem in its various manifestations more closely, I will present evidence that the representation relation should be thought of as *homomorphic* rather than as isomorphic. That is, instead of requiring that every element and relation in the content structure be mapped into by the function which connects it with the representing structure (instead of requiring that the representation relation be an *onto* mapping), I will require only that some subset of the elements and relations in the content structure be mapped into by the representation relation. Cummins does not explicitly rule out the possibility of having mappings which are not onto in his intuitive definition of isomorphism expressed earlier in this dissertation. But insofar as "particle for particle duplication" suggests an onto relation between structures, Cummins does seem to imply that anything short of "loose" representation requires the

---

[45] Cummins (1996), p. 109.

stronger, onto mapping. Conceiving of the representation relation as homomorphic seems particularly well-motivated now. If the representation relation is weakened in this way, then, speaking intuitively, representation can remain abstract without applying "only loosely" to anything short of duplication of the object represented. In my view, only some features of the city need be mapped into by features of the map in order to achieve the relationship of representation. This seems to be much more natural than relegating all but particle-for-particle duplication to "loose" representation.

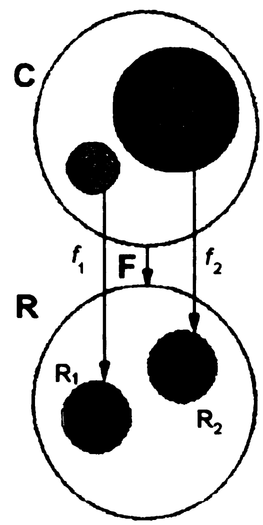

Figure 1-G: Isomorphism 'Modulo' A Substructure

Making this move may help to account for degrees of representation (the first problem mentioned above), and can do so without the use of targets. To see how, suppose that R_1 and R_2 are both representational structures in a cognitive system/structure R and that f_1 and f_2 are homomorphisms of R_1 and R_2 into C. Suppose further that c_1 and c_2 are substructures of C and that R_1 and R_2 are isomorphic to c_1 and c_2 respectively. We may say that R_1 is an inferior representation of C compared with R_2 just in case the domain of c_1 is less than than the domain of c_2. The intuitive idea is, of course, that if R_2 is more complete, i.e., if fewer elements need to be excluded from C in order to achieve an isomorphic mapping with some R_2 when compared with R_1, then R_2 is a better representation of C that R_1 is (see **Figure 1-G** above).[46]

[46] This approach may also have the advantage of providing an account of what makes representations be representations of the same or different types; or equivalently, of explaining the ability of cognitive systems to select from among several different ways of typing objects. When a system classifies objects as belonging to the *same type* we "mod out" the features or substructures of the representational structures which are the objects of our typification. Alternatively, when the system classifies these objects as belonging to different types, the relationship between them remains a homomorphism.

CHAPTER 2: *The Picture Theory of Meaning*

Before moving on to consider the uniqueness problem and its possible solutions in more detail, I'd like to pause to provide the reader with a more detailed introduction to the view of mental representation known as the picture theory in order to compare this view with interpretational semantics and to determine the capacity of interpretational semantics for responding to traditional difficulties with the view. Accordingly, this chapter is organized into three primary parts: First, I discuss the basic tenets of a typical version of the picture theory of mental representation; next, I discuss how interpretational semantics may be seen as a refinement of this view; and lastly, I consider in detail some of the more troubling objections raised against a picture theory of representation and investigate whether or not interpretational semantics can surmount them.

2.1 An Introduction to the Picture Theory of Representation

Perhaps the best known proponent of the picture theory of meaning was Ludwig Wittgenstein, despite the fact that in his later work, Wittgenstein largely rejected the view. Other philosophers, such as Aristotle, and some of the Empiricists also put forth versions of the view.[47] It is arguable that Leibniz held a version of the picture theory where his theory of perception is concerned, insofar as he held that one thing represents or expresses another when "there is a constant and fixed relation between what can be said of one and

[47] Compare for example, the empiricist John Locke's views on primary qualities and how they *resemble* their objects.

what can be said of another."[48] In general, philosophers who held a picture theory of representation claimed that representations have the content that they have in virtue of a *resemblance* between the representation and the object represented. Aristotle took the notion of resemblance quite literally—for him, representations contained the *form* of the things that they represented. In addition, ancient philosophers such as Epicurus and Empedocles, believed that *eidola* or "images" of external objects, impinged on the senses to create our perceptions of them. Epicurus allowed that perception could be further influenced by judgment, occasionally leading to perceptions of things such as unicorns and centaurs, but maintained that all perception can be traced to the reception of *eidola*.[49]

Later advocates of the picture theory, such as Wittgenstein, did not require that representations resemble objects in the sense of "looking like" those objects. Rather, representations, for Wittgenstein, could depict objects and/or states of affairs "... provided there were as many distinguishable elements within the [representation] as within the situation it represents, so that the [representation] possessed the appropriate pictorial form to be isomorphic to [the object or] state of affairs."[50] With Wittgenstein, therefore, the picture theory of meaning took a formal turn—features shared between the representation

---

[48] from "letter to Arnauld, G. 2:112/L 339" referenced in Simmons, Alison. "Changing the Cartesian Mind." in The Philosophical Review. January 2001. Simmons writes: "In other words, [for Leibniz] representation involves an isomorphism between *res repraesentans* and *res repraesentata*. Resemblance is the paradigm case, but other forms of isomorphism will do; planar projective drawings represent solids, maps represent cities, musical notation represents a musical composition and so on." (Simmons (2001).

[49] Copleston, Frederick, S.J. A History of Philosophy Vol. I, Greece and Rome. New York: Doubleday, 1993. pp. 402 & 403.

[50] Logue, James. printed in The Oxford Companion to Philosophy. ed. Ted Honderich. New York: Oxford University Press, 1995. p. 681.

and the thing represented are described more formally (i.e., in terms of isomorphism) than is possible when one invokes the notion of a visual image or picture to account for pictorial similarity. In addition, describing pictorial similarity in terms of isomorphism made it possible to give an account for the representational capacities of things such as propositions, which are presumably not pictorial images of the things they represent at all. Should a proposition be analyzable into elements which correspond to the state of affairs it describes, that proposition would thereby be said to *represent* that state of affairs, in virtue of the formal resemblance which the proposition bears to it.

2.2 Interpretational Semantics and the Picture Theory of Representation

Cummins' interpretational semantics provides a fully formal interpretation of the notion of "pictorial similarity" and hence, is a formalized version of a picture theory of representation. Like Wittgenstein, Cummins also depicts the representation relation as an *isomorphic* relation between representation and thing represented. In his view, representations are mathematical structures and have all of the associated characteristics of such structures. In Chapter 1 we saw that whenever a representational structure R represents a "content" structure C:

> 1. An object in R can represent an object in C.
> 2. A relation in R can represent a relation in C.
> 3. A state of affairs in R—a relation holding of an n-tuple of objects—can represent a state of affairs in C.[51]

---

[51] Cummins (1996), p. 96

Specifically, *R* and *C* are related isomorphically in virtue of the existence of a special kind of function called an *isomorphism* from *R* onto *C*. Although we will want to say more about the notion of isomorphism later, for now, let us define an isomorphic function as follows: A function *h* is an isomorphism mapping a structure $\mathfrak{R}$ onto a second structure $\mathfrak{C}$ just in case:

(a) for each *n*-place predicate symbol *P* and each *n*-tuple $<a_1, ..., a_n>$ of elements of $\mathfrak{R}$, $<a_1, ..., a_n> \in P^{\mathfrak{R}}$ iff $<h(a_1), ..., h(a_n)> \in P^{\mathfrak{C}}$ (where P^i should be read as "the predicate *P* in structure *i*").

(b) for each *n*-place function symbol *f* and each such *n*-tuple, $h(f^{\mathfrak{R}}(a_1, ..., a_n)) = f^{\mathfrak{C}}(h(a_1), ..., h(a_n))$ (where f^i should be read as "the function *f* in the structure *i*).

(c) *h* is one-to-one and onto.[52]

The proponent of interpretational semantics claims that whenever such a function exists between two structures, the two structures are *similar* or *resemble* one another, where, like Wittgenstein, we no longer mean to imply by this that the two *look* similar.

2.3 Common Criticisms of Picture Theory and Responses from the Perspective of Interpretational Semantics

Although which criticisms are most effective against a picture theory of meaning generally depends, in part, on whether the theory has undergone the formalization characteristic of Wittgenstein's view and interpretational semantics, some criticisms are classical objections to picture theory in any of its manifestations. I summarize versions of some of these criticisms here and suggest strategies for responding to them from the

[52] Enderton (1972), pp. 89 & 90. Enderton does not require that isomorphisms are onto mappings. I append this condition to bring the definition of isomorphism into accord with the way it is understood in contexts such as group theory and abstract algebra.

perspective of an interpretational semanticist. Finally, I consider a criticism of picture theory (raised indirectly by both Dennett and Fodor in their discussions of the picture theory of representation) which I believe to be related to what I have been calling the "standard" uniqueness problem for interpretational semantics.

2.3.1 The Problem of Abstraction Via Daniel Dennett

Daniel Dennett makes his case against the picture theory of representation in an article entitled, "The Nature of Images and the Introspective Trap."[53] His use of the term 'image' may be understood as highly analogous to our use of the term 'picture' thus far, insofar as images for Dennett have representational capacities in virtue of the relationship they bear to the thing of which they are an image. However, as we shall see, Dennett's characterization of picture theory is more restrictive than the view embraced by a Wittgensteinian-type of picture theory. Specifically, Dennett portrays picture theory as a view which requires that pictures share *monadic* properties with what they represent—e.g., the picture must have some color, shape, smell, taste, etc. in common with the represented object.

> An image is a *representation* of something, but what sets it aside from other representations is that an image represents something else always in virtue of having at least one quality or characteristic of shape, form, or colour in common with what it represents. Images can be in two or three dimensions , can be manufactured or natural, permanent or fleeting, but they must *resemble* what they represent and not merely represent it by playing a role—symbolic, conventional, or functional—in some system.[54]

Dennett's chief concern is whether there are mental representations that represent in virtue

---

[53] Dennett, Daniel. "The Nature of Images and the Introspective Trap." in Readings in Philosophy of Psychology. Vol. II. ed. Ned Block. Cambridge: Harvard University Press, 1981.

[54] Dennett (1981), p. 129.

of being pictures (i.e., whether there are mental representations which are pictorial). His strategy is to show that similarity in general (whether structural or monadic) cannot ground representation at all (and therefore, that pictures cannot be representations). Since interpretational semantics *just is* the attempt to show that the representation relation is grounded in a kind of similarity, namely structural similarity, Dennett's criticisms are applicable here.

Dennett thinks that mental representations have at least one characteristic which is essentially unlike the characteristics of pictures,[55] arguing that if mental representations and pictures do not share all of the same characteristics, then mental representations cannot be pictorial. Specifically, Dennett thinks that mental representations can be *abstract* in a way that pictures cannot. This is true, he claims, because unlike pictorial representations of a thing, the mind, and in particular, the imagination, can represent while simply *leaving out* some of a thing's characteristics from consideration. For example, a *picture* of a man with a wooden leg could not refrain from mentioning the detail that the man either has or fails to have a hat. But on the other hand, the *imagination* can conjure up a representation of the man with the wooden leg without taking information regarding his hat into consideration—*that* piece of information is simply not part of the mental representation. In conclusion, argues Dennett, since mental representations, formed in the imagination, can have the quality of being detail-poor, but pictures cannot have this quality, mental representations cannot be pictures.

[55] Note that I (and most authors) concentrate on pictures whose properties are detectable through *vision*. However, presumably there are auditory and tactile "pictures" as well.

> We can, and usually do, imagine things without going into great detail. If I imagine a tall man with a wooden leg, I need not also have imagined him as having hair of a certain colour, dressed in any particular clothes, having or not having a hat. If, on the other hand, I were to draw a picture of this man, I would have to go into details. I can make the picture fuzzy, or in silhouette, but unless something positive is drawn in where the hat should be, obscuring that area, the man in the picture must either have a hat on or not. As [J.M.] Shorter points out, my not going into details about hair colour in my imagining does not mean that his hair is colored 'vague' in my imagining; his hair is simply not 'mentioned' in my imagining at all.[56]

In addition to persuading Dennett that mental representations cannot be the same as pictorial images, this argument convinces him that a mental representation is more like a description than like a picture. For like mental representations descriptions can either be detail-poor or detail-rich. I can describe Sephra with or without mentioning that she is a cat; I can describe the man with the wooden leg with or without mentioning that he has a hat. *Pictures* of Sephra however, would address the information about the type of animal she is necessarily (either by revealing that she is a cat or representing her in such a way that she is not), and so insofar as mental representations of her need not do so, Dennett believes that they are distinct from pictures and closer to descriptions.

It is arguable that pictures are not always as (necessarily) revealing of the details as Dennett suggests here. For example, suppose that I want to represent Sephra not only as a cat, but as a *Siamese* cat. Since Sephra's Siameseness is something primarily detectable by her coloring (at least, if your only source of information is a picture), then a monochrome picture of Sephra might not convey the information in question, nor would a silhouetted picture of her. Even her shape in silhouette would not reveal the breed of cat she is, since

[56] Dennett (1981), p. 130.

Russian Blues, for example, look strikingly similar to Siamese in silhouette.[57] Dennett's example really only demonstrates that pictures convey all of the details about an object that can be conveyed in virtue of things like its shape (and maybe its size, although this is doubtful). But this doesn't show that pictures are incapable of abstraction *per se*—only that they do not abstract away from certain kinds of features. And this is certainly true of mental representations (if not of descriptions) as well,[58] for it is doubtful that I can have a representation of the man with the wooden leg unless I have a representation of certain of his (perhaps more fundamental) characteristics.

There is an additional argument Dennett raises for believing that abstraction is a persistent problem for the view that mental representations are pictorial (and for believing that they are more appropriately thought of as descriptive). The following is his famous "striped tiger" example:

> Consider the Tiger and his stripes. I can dream, imagine or see a striped tiger, but must the tiger I experience have a particular number of stripes? If seeing or imagining is having a mental image, then the image of the tiger *must*—obeying the rules of images in general—reveal a definite number of stripes showing, and one should be able to pin this down

---

[57] This is particularly true of "stick figures" and other line drawings. This example (**Figure 2-A**) probably conveys that the represented thing is some sort of cat, but it certainly doesn't convey all of the available details.

Figure 2-A:
'Sephra Approaching the Food'

[58] **Figure 2-A** also illustrates how pictures can be descriptive without filling in all of the details. This figure might serve the same purpose as its more descriptive caption.

with such questions as 'more than ten?', 'less than twenty?'. If however, seeing or imagining has a descriptional character, the questions need have no definite answer. Unlike a snapshot of a tiger, a description of a tiger need not go into the number of stripes at all; 'numerous stripes' may be all the description says.[59]

I believe that this argument is similar to the first and makes the same mistaken assumptions about the capability of pictures to abstract. Dennett's criticisms here seem to be based on the assumption that pictures must have features isomorphic to the features of that which they are images of. It is not difficult to see why Dennett would make this assumption—indeed, I have characterized both Wittgenstein's version of picture theory and interpretational semantics in a way that would suggest it. It is the assumption which is behind his contention that images cannot abstract away from even some of the features of their object—all of the stripes of the tiger must be represented in the image of the tiger, for example. This is the grounds for rejecting images as good candidates for mental representation. But on closer examination, this assumption is curious, to say the least. Not only is it obvious that images *do not* depict, as a rule, every feature of the thing of which they are an image, but assuming so would be inconsistent with Dennett's own initial definition of "image" as something which must have "... *at least one* quality or characteristic of shape, form, or colour in common with what it represents."[60] In any case, it is Dennett's initial definition of "image" that is more analogous to the notion of pictorial similarity I will ultimately defend. I will argue that, contrary to Cummins, the representation relation is better characterized as a *homomorphic* relation, rather than as an *isomorphic* one. If resemblance does not require an onto relation between representation

[59] Dennett (1981), p. 130.

[60] Dennett (1981), p. 129.

and represented object, then, contrary to Dennett's view, the picture theorist *need not* portray the representing "image" as one which shares every feature with the object it represents. If some features can be left out, then perhaps the door is opened to the possibility of accommodating abstraction in representation after all.

Conceptualizing the representation relation as one which does not require an onto mapping between structures, namely, as a homomorphism, may prove to have the advantage of making abstract representations easier to conceptualize within the context of a picture theory of representation. But Dennett's arguments raise another difficulty for picture theory which we have not yet acknowledged or responded to. Although it now seems clear that pictures *can* represent only some of the properties of their objects, Dennett's examples also raise the possibility that pictures can, and possibly must, have properties that are not meant to correspond to *anything* in the represented object. For example, in **Figure 2-A** above, the color of the lines used to draw the stick-image is blue, but we would not ordinarily imagine that the color of the lines is a representation of any property of Sephra. For interpretational semantics, wherein the representation relation is an isomorphism, this presents a special challenge. On this view, *R* is a representation of *C* *if and only if* *R* and *C* are isomorphic structures (and are therefore, among other things, one-to-one); hence, it would seem that there can be no element in *C* which is not mapped into by some corresponding element in *R* and therefore no element in *R* which does not *represent* some aspect of *C*.

Strictly speaking, *something* in *C* must get mapped into by each element of *R*, else we cannot characterize the representation relation as a proper function. But we can imagine that in cases such as the one given above, elements of R map into $\varnothing$, providing the intuitive grounds for saying that some elements of our representations do not correspond to "anything" in the external world. But such a mapping would not be an isomorphism, since it would not be a one-to-one mapping (presumably there would be more than one thing mapped into $\varnothing$). However, this *would* qualify as a *homomorphic* mapping of R into C.

So the move to characterize representation as homomorphism may turn out to be useful in formulating a response to this criticism as well since it drops the requirement that the mapping is one-to-one. But before further examining how this is so, let us take a look at a similar criticism of picture theory described by Jerry Fodor.

2.3.2 The Problem of Abstraction Via Jerry Fodor

Jerry Fodor raises similar problems concerning the ability of pictures to represent abstractly and concerning the non-representational features of pictures in an article entitled "Imagistic Representation."[61] Fodor's notion of a "picture is similar to Dennett's although he is less precise about the conditions under which pictures resemble their objects. For him, pictures are simply representations that refer by resembling.[62]

Like Dennett, Fodor sees the problem of accounting for mental representation as a problem of choosing between two possible alternatives: we will either get a theory of representation which is pictorial, or we will get one which is discursive (what Dennett has

---

[61] Fodor, Jerry. "Imagistic Representation." in Readings in Philosophy of Psychology. Vol. II. ed. Ned Block. Cambridge: Harvard University Press, 1981. pp. 135-148.

[62] Fodor (1981), p. 135.

called "descriptive"). Fodor reconceptualizes this dilemma as one between "iconical" and "symbolic" theories of representation and will ultimately favor the latter over the former.[63] Fodor associates iconical theories of representation (theories wherein the reference of icons is mediated by resemblance) with Jerome Bruner.[64] His own view is that the reference of symbols is mediated by "convention or something"[65] but that regardless, representations are fundamentally symbols rather than pictures.

Also like Dennett, Fodor's strategy is to cast doubt on iconical theories of representation by showing that icons are not the kinds of things that can genuinely represent. A good theory of representation, Fodor claims, will make representations capable of denoting *states of affairs*, by which he means facts of the matter regarding relationships between objects, persons, and events. Icons, he claims, can never do this.

> ...it makes a sort of sense to imagine a representational system in which the counterparts of words resemble what they refer to, [but] it makes no sense at all to imagine a representational system in which the counterparts of sentences do.[66]

---

[63] Both Dennett and Fodor seem to assume these are the only two alternatives available to theorists of mental representation. Both use this assumption to support a theory of discursive representation by rejecting the plausibility of the picture theory. However, if a theory like structural or interpretational semantics is right, then there may be a third alternative. In particular, structural semantics will be revealed to have many of the properties of both symbolic and discursive theories of representation, leading one to speculate on whether the alternatives considered by Fodor and Dennett may form a false dilemma.

[64] See Bruner, J.S. "On Perceptual Readiness," in Psychological Review. Vol. 64, 1957. pp. 123-152.

Bruner, J.S. , Goodnow, J.J. and Austin, G.A.. "A Study of Thinking." New York: Paperback Wiley Science Editions, 1962.

Bruner, J.S., Olver, R.R. and Greenfield, P.M. "Studies in Cognitive Growth," Wiley, New York: 1966.

[65] Fodor (1981), 136.

[66] Fodor (1981), 136-37.

Since we know that words and sentences represent, the idea is that their representational capacities must be explained in terms of pictorial similarity (if an iconical theory of representation is correct). Iconical representation makes words into icons (i.e., things with representational capacity), but it cannot, thinks Fodor, do the same for sentences. But if icons cannot represent what sentences represent, then, representation (and specifically, the representational capacity of sentences) cannot be iconical.[67]

> In Iconic English, *words* resemble what they refer to *but sentences don't resemble what makes them true* [e.g., states of affairs]. Thus, suppose that, in Iconic English, the word 'John' is replaced by a picture of John and the word 'green' is replaced by a green patch. Then the sentence 'John is green' comes out as (say) a picture of John followed by a green picture. But *that* doesn't look like being green; it doesn't look much like anything. Iconic English provides a construal of the notion of a representational system in which (what corresponds to) *words* are

---

[67] Fodor is primarily concerned about the capacity of sentences to represent, and though I do not believe that he provides us with reasons to doubt that icons can represent abstractly here, there may be something to his more general point: that sentences are harder to think of as icons than are individual words. The problem with words (vs. sentences) and their meanings arises in the context of questions about how *language* acquires meaning whereas interpretational semantics, and ultimately structural semantics, are views about the representational capacities of *mental states*. I don't spend a lot of time talking about how language acquires meaning, although I think that the account would have to involve something like the notion that linguistic terms acquire meaning conventionally. Specifically, we might say that people associate the contents of their mental states with words by convention, and that sentences have meaning in virtue of expressing a relation between mental states with representational content. Since both interpretational and structural semantics have it that the contents of mental states and networks of mental states is given by the holistic role which they occupy within a larger representational network, the upshot of this is that linguistic meaning is likewise given holistically. That is, insofar as individual mental states have content in virtue of their role in the representational network, individual words, whose meanings derive from the content of individual mental states, have meaning in virtue of their role in sentences. This is a decidedly different view about language from Fodor's in that Fodor views language as *compositional*: The meaning of a sentence is a function of (is "composed" from) the meanings of its words. Alternatively, Randall Dipert has expressed views on the holistic nature of language which would accord well with the position of structural semantics. In an article entitled, "The Mathematical Structure of the World: The World As Graph," (The Journal of Philosophy Vol. XCIV, No. 7, July 1997), Dipert writes, "It is a long-standing custom to speak of a word or phrase as *having* meaning; this everyday approach, of treating words as having semantic and syntactic (monadic) properties, has been aided and abetted by theories of language since the Middle Ages. As is often briefly noted but rarely developed, however, this is quite misleading: no series of marks, or of sounds, really—intrinsically—has a meaning. A word *has* a meaning only in the bosom of a language."

icons, but it provides no construal of the notion of a representational system in which (what corresponds to) *sentences* are.[68]

Since thoughts are sentence-like (according to Fodor) and since sentences cannot be iconical, thoughts cannot be iconical either.[69] But this is just to say that mental representation is not iconical.

Of course, Fodor's guess about how the statement 'John is green' would be conveyed iconically is probably not the best guess one could make. Intuitively, it seems a much clearer conveyance of the information could be accomplished with a picture of John colored all-green. But this does not avoid the central point of Fodor's objection. Fodor admits that if an iconical theory was to seriously approach the representation of states of affairs, it would probably proceed more in accordance with my guess: He acknowledges, for example, that a picture of the sentence 'John is fat' would probably picture John with a bulging tummy.[70] But there is still a problem: Even if my intention is to represent John's fatness with the image, the image certainly represents many other facts about John having nothing to do with this. How are we to know which of these ways of depicting John the picture represents? For example, if the picture of John with a bulging tummy is also a picture of John standing, then does the picture in question represent the thought: "John is

[68] Fodor (1981), 136.

[69] Fodor (1981), 136.

[70] Fodor (1981), 136.

fat," or does it represent the thought "John is standing?" Fodor contends that theories of iconical representation have no way of determining this.[71]

> ...symbols really *are* abstract. A picture of fat John is also a picture of tall John. But the sentence 'John is fat' abstracts from all of John's properties but one: It is true if he's fat and only if he is. Similarly, a picture of a fat man corresponds in the same way (i.e., by resemblance) to a world where men are fat and a world where men are pregnant. But 'John is fat' abstracts from the fact that fat men *do* look the way that pregnant men *would* look; it is true in a world where John is fat and false in any other world.[72]

Fodor puts the problem like this: "The trouble is *precisely* that icons are insufficiently abstract to be the vehicles of truth."[73] So, like Dennett, Fodor believes that pictures lack at least one of the characteristics necessary for grounding representational content, namely, the ability to represent abstractly. But Fodor's emphasis in this example may be slightly different from Dennett's. Dennett's claim was that pictures cannot fail to represent all of the properties of the objects of which they are pictures, and therefore cannot be capable of abstraction. I argued that this position was inaccurate by showing examples of pictures which do convey only some of an object's properties. Fodor acknowledges the ability of pictures to convey only some of an object's properties, but seems to think that icons still fall short of being "sufficiently abstract." To see why, suppose that we have a picture of John and that this picture in fact conveys only some of his characteristics. In particular, the picture is one which portrays John standing up, colored green, and with a

---

[71] Of course, one could always ask me what I *intended* to represent with the image, but as soon as one remembers that we are ultimately interested in claiming that *mental* representations are images, invoking intentions threatens the project of providing a fully naturalistic account of mental representations.

[72] Fodor (1981), 138.

[73] Fodor (1981), 137.

sightly bulging tummy. Fodor's claim is that when we are presented with a picture of a green, standing John with a bulging tummy, which of these features is currently being represented by us is not clear, and the picture can do nothing to make it any clearer. On the other hand, if someone were to utter the sentence, 'John is fat' it would be clear from this that this person represents those characteristics of John's which contribute to his fatness, and not necessarily those which may indicate his color, his stance, etc. Since sentences can "abstract away" from John's other properties in this way, while pictures cannot, Fodor maintains that representations must be more like sentences than like pictures.

In addition, like Dennett, Fodor's arguments show that there will probably be properties of pictures which are not ordinarily understood to be representational at all, and therefore do not necessarily "resemble" anything about the objects they depict. For example, one may find that there are variations in the darkness of the lines used to create the picture of John, the result of unevenly applied pressure on the artist's pencil. We would not ordinarily think of such variations as representations of any particular property of John's though they are most certainly properties of the representation itself.

There are at least two responses which can be made to Fodor's criticisms concerning the ability of pictures to represent abstractly. First, it is arguable that although pictures do not always pick out only one of the available properties of a subject and frequently (perhaps always) contain properties which are non-representational, the same can be said of sentences. If this is so, then Fodor has not provided us with a convincing reason to prefer descriptive theories of representation to iconical theories of representation. Second, there may be clear criteria for judging that a picture represents but one of a wide variety of possible targets, despite Fodor's worries that there are no such criteria. To see why, it is necessary to be more specific about how "pictures" ought to be conceptualized in the context of a picture theory of representation. I'll consider both of these points in turn.

2.3.2.1 Non-Representational Properties of Pictures and Sentences

Fodor's case requires that both of the following are true: (1) sentences do a better job than icons do of abstracting away from all of a subject's properties except for those the sentence represents, and (2) sentences, unlike pictures, do not have properties which are non-representational. I believe that both of these statements are false. Let's consider (2) first.

It is important to point out that it is easy to get into trouble if one takes it for granted that there are properties of pictures which are not supposed to be representational. Any claim that a property of some icon is not one which represents is a claim which must be couched in an antecedently accepted theory of representational content. It wouldn't be surprising therefore if advocates of a non-iconical theory of representation were able to identify properties of icons that didn't seem to represent anything. But for the sake of argument, let us suppose that icons *can* have representational content, but that they may nonetheless have specific properties which do not represent anything in the object they depict. Since interpretational semantics and Wittgensteinian picture theory both require that representations be isomorphic to the structures which give them content, they are hard-pressed to explain such a scenario. Isomorphism requires that the representing structure and represented structure are one-to-one, and this prohibits the possibility of having properties in the representation which do not correspond to anything in the object represented.

Understanding the representation relation as a homomorphism may provide a way out of this quandary. Since homomorphism drops the one-to-one requirement, there is nothing preventing a situation in which a number of elements in the representational structure get mapped to $\varnothing$ in the represented structure.[74] This is one way to explain what we intuitively describe as a case where some features of a representation don't correspond to any features of a structure in the external world. We say nonetheless that we have a representation of the external structure in such cases. On the other hand, as we shall see, dropping the one-to-one and onto requirements on the representation relation may raise other difficulties for this account of representational content. Among the most significant of these is an exacerbation of the uniqueness problem, which we will consider in detail presently.

Interestingly, it may not be necessary to weaken the representation relation in order to respond effectively to Fodor's criticism. Instead, advocates of interpretational semantics could argue that non-representational properties of pictures are always *monadic*, and that though such properties always have to be mapped to monadic properties of the represented object, these monadic properties don't have to be *the same*. For example, if the picture of John is drawn in green, the isomorphism does not have to map this property

[74] $\varnothing$ will always be an element of the represented structure, since $\varnothing$ is a member of all sets.

to an actual instance of the color green, or to any color at all even if it is necessary that this feature is mapped to *some* feature of John.

So in summary, it looks as if pictures may be able to contain non-representational properties without compromising the possibility that iconical theories of representation are true. But regardless, the existence of non-representational properties of icons does not distinguish them from sentences, since sentences can also have properties which have nothing to do with their representational content. For example, the proposition, 'John is green' can be tokened in pencil on a piece of paper, can be spoken, can be stored as magnetic media on a hard disk, or can be emblazoned on the back of a vintage jacket. Each of these manifestations of the proposition has properties which in no way represent aspects of the proposition itself. As a result, it is hard to see why the possible presence of non-representational properties in pictures would be cause to compare them disfavorably with sentences.

2.3.2.2 Can Discursive/Symbolic Devices Uniquely Specify A Subject's Properties?

Next we need to consider whether sentences do a better job than icons do of abstracting away from all of a subject's properties except for those the sentence represents (Fodor's assumption (1) from above).

Fodor argues that pictures usually portray a number of the properties possessed by their subjects, even if not all of them, and there is no way to know which of these is the one of which the picture is a representation.[75] Since sentences presumably do better at uniquely specifying the content of a representation, representations are probably more like sentences than pictures. But do sentences really pick out the things they represent as uniquely as Fodor suggests? Probably not. Consider the following sentences:

(a) 'Bob is larger than Tom.'
(b) 'Mary is swift.'
(c) 'Cleveland is not close to Sydney.'

In the first case, if we suppose that Bob is both heftier and taller than Tom, then (a) does not pick out one or the other of these properties uniquely. The sentence could function to specify that Bob is heftier, taller, or even both heftier and taller, but nothing about its structure alone will serve to determine which. Sentence (b) is ambiguous due to the ambiguity of the adjective it contains. Assume that Mary is in fact both intelligent and fleet-of-foot. Perhaps (b) is supposed to represent the fact that Mary is smart, but we have no way of knowing whether this was the intention, or whether the fact that Mary runs fast was the target of the expression. Finally, (c) contains two nouns, 'Cleveland' and 'Sydney' whose references might change depending on how one takes the ambiguous phrase 'is not close to.' For example, if we take the phrase to represent a judgment about distance, then both nouns would likely name locations, specifically cities. But if we interpret the phrase as one which describes the status of a personal relationship, then 'Cleveland' and 'Sydney' might well name people. Therefore, not only does (c) fail to specify uniquely which

[75] That is, short of asking someone what they intended to represent with the picture, which is not an option here, since we want an explanation of the representational capacities of pictures which does not appeal to the intentions of a "designer" or "interpreter."

properties of an object are represented, it does not even uniquely pick out the objects themselves.

Of course, the sentences I have used here all have one thing in common: They are ambiguous in one way or another and could be disambiguated if more information was obtained from the speaker. But note that we cannot ask the speaker about his or her intentions in uttering these statements for the same reason that we could not inquire about the intended target of a pictorial representation. If pictures or sentences are going to be the kinds of structures which possess representational content primitively, then they cannot possess it in virtue of having been interpreted or designed to do so. Whatever the mechanism by which our mental structures have representational content, it cannot be one which relies on the more primitive capacities of an outside designer if we are to achieve a fully naturalistic account of mental representation.

Fodor might agree that these sentences are not, by themselves, capable of revealing, say, whether it is Bob's relative tallness or his heftiness that we represent. However, he might argue that statements are, *in general*, capable of such abstraction, whereas pictures, *in general*, are not. In other words, while (a) does not achieve the kind of abstraction Fodor wants representations to be capable of achieving, perhaps (a') does:

(a') Bob is larger than Tom because he is taller than Tom.

It is probably true that most sentences, even if perhaps not sufficiently descriptive to represent a unique property of their subjects, can be made to be descriptive enough for this purpose. But I want to argue that the same can be said for pictures. In order to understand why this is so, let's take a closer look at how both Fodor and Dennett understand the notion of "picture" for picture theory semantics and how this understanding differs from Cummins' and my own notion of "picture" for interpretational and structural semantics.

2.3.2.3 Can Iconical Devices Uniquely Specify A Subject's Properties?

It seems as if both Dennett and Fodor want to require that in order for pictures to be used as representations, they must share at least one *monadic* property (e.g., a color, a shape, a taste, a smell: in general, a *look*) with what they represent. Neither Dennett nor Fodor explicitly state this to be a criterion of picture theory as far as I know, but the examples they use, and specifically, the arguments which they launch against the ability of pictures to represent abstractly, suggest that this criterion is implicit in their views. In particular, by arguing that pictures cannot specify which of an object's properties they are representations of, Fodor seems to ignore some of the non-monadic properties of pictures which might be useful for exactly this purpose. Take the following case, illustrated in **Figure 2-B**.

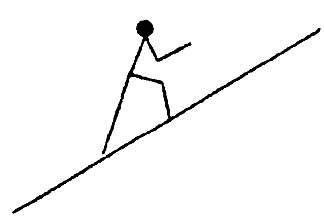

Figure 2-B: Man Climbing/Descending A Mountain

It cannot be clear from the depiction in **Figure 2-B** whether the figure pictured is climbing the incline or descending (backwards) along it. Presumably the picture represents one or the other and, by hypothesis, it shares certain monadic properties with the scenario which it represents. For in either case, there is an incline at (roughly) the angle described by the picture, and a figure who (very roughly) resembles in shape the figure pictured here. Despite these shared properties however, there is no way to determine which of two scenarios (climbing or descending) the picture represents.

Interestingly, interpretational and structural semantics do not require that a picture share monadic properties with the thing which it represents. Many representations *will* share such properties, but pictures and the things they represent are not *required* to share them on these views.[76] Instead, for structural and interpretational semantics, pictures are structures characterized in terms of their *relational* properties. Whether or not a given structure represents a given object (read: is a picture of a given object) has to do with both the formal relationship which it bears to that object and with the relationship which it bears to a larger network of representations. This ends up allowing pictures a kind of flexibility not apparent in either Dennett's or Fodor's description and criticism of picture theory semantics. To see how this works in the case of the example illustrated by **Figure 2-B** let us imagine a slightly altered version of the figure, such as that shown in **Figure 2-C**.

---

[76] There may be one exception to this: homomorphism implies preservation of reflexivity, and reflexivity is arguably monadic. But even if this is true, it in no way compromises the possibility that monadic properties of pictures are mapped to different monadic properties of the represented object, implying that relational properties are what matter for object individuation.

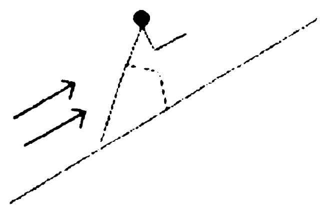

Figure 2-C: Man Climbing A Mountain

While the picture given in **Figure 2-C** may have characteristics which fail to "look like" those present in the actual state of affairs depicted, it is nonetheless a representation of that state of affairs, just in case the two have a similar structure. In Cummins' case, establishing structural similarity amounts to showing that the representation is isomorphically related to the state of affairs it represents. Although I do not suggest that the climber represented here has a set of arrows at his back in reality, what I do suggest is that the arrows restrict the available interpretations of the picture's meaning in virtue of their placement on the diagram in relation to the climber and the incline, and that the *relationship* between the climber and incline that they thereby express is one that is in fact mirrored in reality.[77] Thankfully, this is all that picture theory semantics needs to require

[77] Note that the arrows themselves are symbols which have meaning. Probably most of us take them to indicate the direction of the arrow's pointed side. How these symbols come to have the meaning that they do undoubtedly has a similar explanation to that given for the meanings of symbols in a language. On the other hand, if you see the arrows as more akin to pictures than to descriptors, then it is arguable that **Figure 2-C** does no better at specifying a unique interpretation of the man's activity than **Figure 2-B** does. This is because the arrows, *qua* pictures, do not necessarily represent a direction of movement uniquely. Although this is a legitimate criticism of the example used here, it does very little beyond reestablishing the relevance of the primary problem which structural semantics must address, namely, the uniqueness problem.

of representations, and thus it seems that pictures may well be capable of representing abstractly after all.

Of course, showing that *some* pictures are capable of representing abstractly does nothing to avoid the fact that there are pictures with representational content (such as that given in **Figure 2-B**) which do not clearly represent only one or the other of two very different objects or states of affairs. Fodor is wrong to think that this problem is restricted only to pictorial representations, but is correct to point out the problematic nature of pictures which are not related to a unique object. If all that is required for representation is that the representation is homomorphically mapped to the thing which it represents, as I shall ultimately argue, then **Figure 2-B** represents a climbing figure as well as it represents a descending one. Taken by itself, there is really no criterion for choosing which of these two distinct states of affairs the picture ought to denote. But this problem is, by now, a familiar problem—it is the problem of uniqueness.

2.3.3 The Problem of Uniqueness for Picture Theory

Having attempted to address the allegation that picture theory cannot support representations which are sufficiently abstract, let us examine Fodor's criticism more carefully as it applies to the uniqueness problem for interpretational semantics and picture theory in general. Very generally, this problem arises as the result of the fact that pictures can be interpreted in a variety of different ways. As we have seen, the state of affairs described by the sentence, 'John is fat' could be represented pictorially, but the representing picture would simultaneously represent many other aspects of John, making its content indeterminate. This problem is easy to conceptualize as a kind of uniqueness problem and in particular, as what I have called the "standard" sort of uniqueness problem. In **Figure 2-D**, each C_i is a possible interpretation of the pictorial image doing the representing (R). Each f_i is a homomorphism from R into C_i.

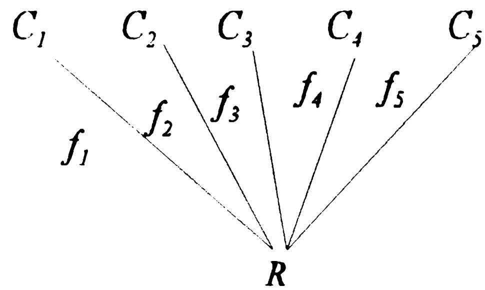

Figure 2-D: Standard Non-Uniqueness Via Fodor's Critique

Although Fodor's criticism raises the standard uniqueness problem for picture theory, it is consistent with Cummins' version of the problem as well. To illustrate how both problems can arise for pictorial representations, consider Dennett's rendition of the dilemma associated with Joseph Jastrow's infamous duck-rabbit (see **Figure 2-E**):[78]

Figure 2-E: The Duck-Rabbit

[78] Joseph Jastrow 1900 "Duck-Rabbit Drawing." in Henry Gleitman, Psychology New York: W. W. Norton and Co., 1981.

What can possibly be the difference between seeing it [the duck-rabbit] first one way and then the other? The image (on the paper or the retina) does not change but there can be more than one description of that image. To be aware of it first as a rabbit and then as a duck can be just a matter of the content of the signals crossing the awareness line, and this in turn could depend on some weighting effect occurring in the course of afferent analysis. One says at the personal level 'First I was aware of it *as* a rabbit, and then *as* a duck,' but if the question is asked 'What is the difference between the two experiences?', one can only answer at this level by repeating one's original remark. To get other more enlightening answers to the question one must resort to the sub-personal level, and here the answer will invoke no images beyond the unchanging image on the retina.[79]

The idea is this: Let the duck-rabbit picture be R (see **Figure 2-F**). Then there are two content structures, C_1 and C_2 such that, under one homomorphic mapping, R has content C_1 (the duck) and under the other, R has content C_2 (the rabbit).

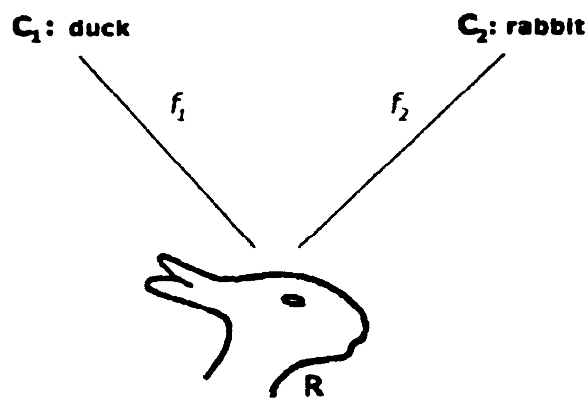

Figure 2-F: Standard Non-Uniqueness Via The Duck-Rabbit

This is the standard sort of uniqueness problem wherein both content structures seem to serve as equally good candidates for specifying the content of R. On the other hand, if we think of the mental image which we have when viewing **Figure 2-F** as R and the picture

[79] Dennett (1981), p. 131.

itself as the object of our representation, *C*, then we have something which looks more like Cummins' version of the uniqueness problem (see **Figure 2-G**).

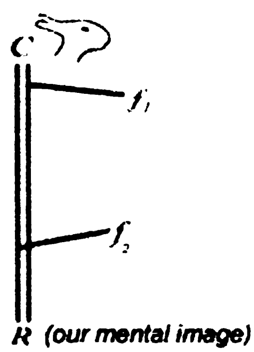

Figure 2-G: Cummins' Non-Uniqueness Via The Duck-Rabbit

Since Fodor's articulation of the uniqueness problem for picture theory can be construed as similar to (indeed, perhaps even identical to) the corresponding problems facing interpretational semantics, a more in-depth look at these problems and the responses available to advocates of interpretational semantics is in order. In the next chapter, I characterize non-uniqueness more formally by exploring how it arises in the context of measurement theory. This more formal characterization of non-uniqueness will put us in a better position to formulate solutions to the problem, and will also reveal at least one additional context in which problems with uniqueness can arise. I'll also consider the proposal to characterize the representation relation as a homomorphism rather than an isomorphism more carefully. This may reveal additional options for responding to the uniqueness problem as it has been characterized so far.

CHAPTER 3: *Representation As Homomorphism And The Problem of Uniqueness*

Until now, I have only briefly alluded to the need for a revision in the way that Cummins characterizes the representation relation: the isomorphic relation he claims holds between representations and what they represent. However, from the preceding discussion and in what follows, it should be clear that the uniqueness problem is not unique to interpretational semantics, and that where it arises in other contexts, part of the approach to dealing with the problem involves, ironically, a *weakening* of the relation between the structures that are not uniquely mapped. I'll begin this chapter with a more detailed look at the nature of the representation relation as Cummins characterizes it. This I will follow with a detailed proposal for a weakening of the representation relation and will argue in particular that the relation is a *homomorphism* rather than an *isomorphism*. I'll discuss some of the advantages that this revision would have over Cummins' original view and will attempt to say how the move to a homomorphic representation relation influences our understanding of how mental representation works, and most importantly, how it can contribute to a solution to various manifestations of the uniqueness problem. Ultimately, the theory that emerges to account for mental representation is a variety of picture theory which is, nonetheless, different enough from Cummins' own view, that I refer to it as "*structural semantics*."

3.1 Cummins' Isomorphic Mapping Relation

Cummins characterizes mental representation as consisting of a mapping between structured features of a cognitive system (e.g., a brain or computer) and things in the external world. Specifically, a structure of the cognitive system is a *representation* just in case a mapping, and in particular, an *isomorphic mapping* exists between it and some structure in the external world. Since the account of representation proposed by Cummins requires only that isomorphisms of this sort *exist* and not that we or any other agent are aware of them, this theory of the nature of representation is a non-question begging one (that is to say, we do not require that an agent who already possesses representations identify or construct the isomorphism which is in turn to account for the presence of representation in cognitive systems originally).

A closer look at the precise nature of the relation Cummins uses to account for the representational capacities of cognitive systems will reveal why a friendly amendment to his account may be in order.

Cummins does not, to my knowledge, describe the notion of isomorphism with much precision in any of his works on interpretational semantics, but does say that it should be understood the same way in the context of mental representation as it is in mathematics. Following this lead, I will endeavor to describe a view of isomorphic relations for the reader which is more precise than that offered in the previous chapters so that their role in the explication of mental representation for Cummins, and for me, can be understood more clearly. I will characterize isomorphic relations, at least initially, as relations that hold between relational systems, where a *relational system*[80] is "a finite sequence of the form $\mathfrak{S} = \langle A, R_1, ..., R_n \rangle$, such that A is a nonempty set of elements called the *domain* of the system $\mathfrak{S}$, and $R_1, ..., R_n$ are relations on A."[81] Relational systems may be of a certain *type*. The following are examples of relational systems which differ in type:

$\mathfrak{S}_1 = \langle A_1, R_1 \rangle$	where A_1 is the set of all cats and R_1 is the binary relation of being cuter than, such that for any $a, b \in A_1$, aR_1b iff a is cuter than b.
$\mathfrak{S}_2 = \langle A_2, R_2 \rangle$	where A_2 is the set of all people and R_2 is the quaternary relation of one pair of siblings being more alike than another. That is, for any a, b, c, and $d \in A_2$, abR_2cd iff a and b are siblings, c and d are siblings and a and b are more alike than c and d.[82]

In general, we may define the *type* of a relational system as follows: "If $s = \langle m_1, ..., m_n \rangle$ is an n-termed sequence of positive integers, then a relational system $\mathfrak{S} = \langle A, R_1, ..., R_n \rangle$ is of type s if for each $i = 1, ..., n$ the relation R_i is an m_i-ary relation."[83] Note that this implies that $\mathfrak{S}_1$ in the example above is of type $\langle 2 \rangle$ while $\mathfrak{S}_2$ is of type $\langle 4 \rangle$. A relational system $\mathfrak{S}_3$

---

[80] Tarski, A. Contributions to the Theory of Models, I, II. Indagationes Mathematicae. 1954, pp. 572-588. and P. Suppes and J. L. Zinnes, "Basic Measurement Theory", in R.D.Luce, R.R. Bush, and E. Galanter (eds), Handbook of Mathematical Psychology, Vol. I, New York: Wiley, 1963. pp. 1-76.

[81] The concept of a relational system is analogous to the concept of a structure (Enderton (1972)) for a formal language.

[82] Of course, the relations that form a relational system are really sets of ordered n-tuples. In the first case, for example, R_1 is a set of ordered pairs (of cats). The phrase "is cuter than" is the intuitive English rendition of this relation. I use the intuitive formulation here because I will ultimately want to consider the role which relational systems play in representation, and the more intuitive rendition of the relation will help make that role easier to understand.

[83] Suppes, Patrick and J. L. Zinnes, "Basic Measurement Theory." in R.D.Luce, R.R. Bush, and E. Galanter, eds. Handbook of Mathematical Psychology. Vol. I, New York: Wiley, 1963. p. 5.

$= \langle A_3, R_1, R_2 \rangle$ where R_1 and R_2 are both binary relations would be of type $\langle 2, 2 \rangle$.[84] Two relational systems are *similar* if they are of the same type.[85]

We are now in a position to state a definition of isomorphism for relational systems. Let us start with the particular case of relational systems of type $\langle 2 \rangle$. Suppose that $\mathfrak{S}$ and $\mathfrak{R}$ are type $\langle 2 \rangle$ relational systems such that $\mathfrak{S} = \langle A, R \rangle$ and $\mathfrak{R} = \langle B, S \rangle$. Now, $\mathfrak{S}$ and $\mathfrak{R}$ are *isomorphic* if there is a one-to-one function f from A onto B such that for every a and b in A:

$$aRb \text{ iff } f(a) = S(f(b)).$$

In general, let $\mathfrak{S}$ and $\mathfrak{R}$ be similar relational systems, such that $\mathfrak{S} = \langle A, R_1, ..., R_n \rangle$ and $\mathfrak{R} = \langle B, S_1, ..., S_n \rangle$. Then, $\mathfrak{S}$ is isomorphic to $\mathfrak{R}$ if there is a one-one function f from A onto B such that, for each $i = 1, ..., n$ and for each sequence $\langle a_1, ..., a_{mi} \rangle$[86] of elements of A,

$$R_i(a_1, ..., a_{mi}) \text{ iff } S_i(f(a_1), ..., f(a_{mi}))$$[87]

The objective, of course, is to use the notion of isomorphism between relational systems to account for a cognitive system's ability to represent the external world. As we

---

[84] Suppes et al. (1963), p. 5.

[85] Note, knowing when two relational systems are of the same type is important because it formalizes part of the notion of "same structure" that is employed in the claim that a cognitive relational system is a representation of a structure in the external world in virtue of the fact that the two systems have the *same structure*.

[86] Please read a_{mi} as "*a* sub *m* sub *i*".

[87] Suppes' and Zinnes' definition of isomorphism also assumes the restriction that mappings be both one-to-one and onto.

have seen, Cummins wants to say that a given representation has the content that it does (and indeed, *is* a representation) in virtue of the existence of an isomorphic mapping between it and the thing it presumably represents. Suppose, to take a concrete example, that we want to explain how some cognitive system rudimentarily represents its neighborhood. There is a relational system $\mathfrak{I}_1 = \langle A, R_1, ..., R_n \rangle$ such that A, the domain of $\mathfrak{I}_1$, contains streets[88] and $\langle R_1, ..., R_n \rangle$ is a set of relations between streets (for example, streets may be related in the following common ways: "is longer than," "intersects," "is to the east of," "dead-ends into," etc.). Together the domain of $\mathfrak{I}_1$ and the relations on that domain constitute a neighborhood. Now, let $\mathfrak{R}_1 = \langle B, S_1, ..., S_n \rangle$ be a relational system instantiated by some cognitive system P. P instantiates $\mathfrak{R}_1$ in the sense that the domain, B of $\mathfrak{R}_1$ consists of a set of data structures (implemented in P) which are ordered by the set of relations $\langle S_1, ..., S_n \rangle$. If it is true that there is an isomorphism between $\mathfrak{I}_1$ and $\mathfrak{R}_1$ in the sense described above, then, on the present view, $\mathfrak{R}_1$ is a representation of the neighborhood of P (since P's actual physical neighborhood *just is* the relational system $\mathfrak{I}_1$).

Of note is that the elements of the domain in this example may be complex entities that, viewed from a lower level, are themseleves relational systems—systems which could, from that lower level, be representations for the cognitive system. For example, in the present case, streets may be—and probably are–relational systems with

[88] I consider the representation of the system's neighborhood to be a relatively un-complex case (in that it does not involve *a lot* of subsystems (like streets) of the main relational systems under discussion which might themselves be thought of as representable), but I do not intend to suggest that it is the simplest case of representation. What the "simplest" case of representation is might be an interesting question to pursue in this context.

their own elements and relations, capable of being independently represented by the system. In general, this example demonstrates that there is an ambiguity, thus far unexplored, in what constitutes the "structures" being mapped in the representation relation. In some cases, the representation relation maps structures that consist of what are intuitively simple elements and relations on those elements. But other times what are mapped are structures whose elements might normally be seen as complex entities themselves. Indeed, the elements could even be properties and relations. One way of understanding the ambiguity described here is to think of there being different levels of representation; sometimes we are representing things at a higher level of abstraction, sometimes at a lower level. One place one often gets such changes in level is in moving from one scientific field to another (e.g., ecology often treats species as individual elements where a lower level in biology would treat them as complex interacting systems.

Therefore, it is important to recognize that we must be able to specify what elements belong to the domain of a relational system, and what relations are defined over these elements (i.e., what ordered n-tuples belong to the relational systems in question) before it is possible to determine whether two (or more) systems meet the conditions required for them to be isomorphic to one another, or alternatively, what two (or more) "systems" we are even dealing with in the first place.

As a concrete example, consider the construction of a database table for storing user information. Suppose that I have a database housed on a mainframe computer which can only store information using scalar data types. A table containing the user's unique id, his first name, last name, and phone number has been created in the system which contains several million records. Though the company I work for has historically been interested only in the home phone number of each user, it would now like to begin collecting and discriminating between the user's home, work and cell phone numbers. Though I would like to be able to store all three numbers in the "phone" field of the users table (see **Figure 3-B**), the limitation on the type of data storage the mainframe supports will not allow me to treat it as an array of values. As a result, I decide to create a new table called "phone_numbers" which contains the three types of number, plus an integer value corresponding to the value of the "phone" field in the original "users" table. Where before this field was a discrete representation of the home phone number of individual users, now it serves as a pointer to an array of numbers, each related in virtue of belonging to the same individual.

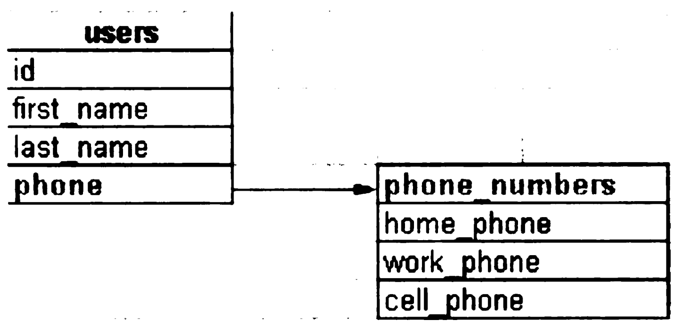

Figure 3-A: Mainframe Database Tables

Examples like these, illustrating the value of locating representations at different "levels" of analysis, suggest that Cummins' requirement that features of the cognitive system be isomorphic to features of the external world is a significant restriction on the representation relation. This is true because of the requirement that, to be isomorphic, two relational systems must be related by a function which is one-to-one. Since one-to-one, onto functions must map distinct elements of the domain into distinct elements of the range, it is difficult to see how isomorphisms could support mappings involving "higher-order" representations of the sort described here. Specifically, if we take the "phone" field of a particular users table entry to represent the entire table of phone numbers for that record, then it seems that we have a mapping which resembles something more like the following:

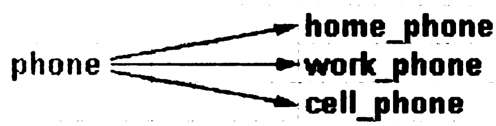

Figure 3-B: Failure of One-to-One

Although it is possible to think of the set {home_phone, work_phone, cell_phone} as a single object: phone_numbers, thereby preserving a one-to-one, onto mapping between phone and phone_numbers, this understanding of the mapping relations obscures the complex structure of the phone_numbers construct, and is not, in my view, the most natural way of viewing the relation between phone and the phone_numbers table. If I am correct, then we need a construal of the representation relation which does not require that the relation is a one-to-one, onto mapping between structures. This condition is met by the notion of homomorphism.

In addition to allowing only one-to-one mappings between relational structures, isomorphism, in virtue of the onto requirement, guarantees that the sets of elements composing each relational system are equivalent in cardinality. Despite these requirements, it seems obvious that the relational system which constitutes the cognitive system's representation and the relational system which exists in the external world may not always have the same number of elements in their domains. Speaking intuitively, if the cognitive system's representation is one that does not pick out each and every one of the features of the thing which it represents, then no isomorphic mapping between the two structures is possible.[89] Likewise, if the cognitive system's mental structure has features not mirrored in the object it represents or possesses more than one mental structure that can be mapped to this object, then no isomorphism obtains between them in this case either. It seems likely that most relationships between mental and external relational systems will fall under one of these two cases, leading one to conclude that a better description of the representation relation is that it is a *homomorphic* relation between relational systems. We understand homomorphic relations in exactly the same way that we understand the relation of isomorphism, with the exception that for homomorphic relations we drop the requirements that f be one-to-one and onto. Notice that even cases wherein the mental and external relational systems *do* possess domains of equal cardinality will meet the conditions necessary for representation, since, under this formulation, it is implied that every isomorphism is a homomorphism, but not the converse.

It makes sense to ask therefore, whether we might be able to be satisfied with the weaker relation of homomorphism in accounting for the nature of mental representation given that it does not require that the function mapping the two relational systems be one-to-one and onto. In the next section, I'll endeavor to show how the account of

[89] I use the term "features" here to denote elements of both the representational relational system and the external relational system.

representation is affected by the move to a homomorphic relation, what advantages the move may present, and what difficulties, if any, it may raise.

3.2 Understanding the Representation Relation as a Homomorphism: Structural Semantics

3.2.1 How Homomorphism Accounts for Paradigm Cases of Representation

The way in which the notion of homomorphism replaces Cummins' use of isomorphism is really quite simple. Instead of saying that the representation relation consists in an isomorphic mapping *f* between the relational system of the cognizer and the relational system which constitutes representable features of the external world, we now state that only a homomorphic mapping is required. This amounts, in part, to allowing that mappings occur between individuals in one relational system and the *subsystems* of other relational systems. We define a *subsystem* of a relational system $\mathfrak{S}$ as "a relational system obtained from $\mathfrak{S}$ by taking a domain that is a subset of the domain of $\mathfrak{S}$ and restricting all relations of $\mathfrak{S}$ to this subset."[90]

This may be a more significant alteration of the relation *f* than it initially seems to be. Part of the advantage of requiring an isomorphic relation between the network of representations in the cognitive system and representable things in the external world may have been that no one needed to decide which features of the cognitive system (the range) should belong to the relationship. It is as simple as this: What features *exist* in the cognitive system are those that belong to the range of the function mapping the two

[90] Suppes et al. (1963), p. 10.

systems together.[91] Setting it up this way seems to help one to avoid the problem of requiring an agent (who would have to already possess representational capacities) to identify which features are most important to the representation relation (and hence, which features belong in the range)—a problem because it would make the account of representational content dependent on a more primitive representational capacity. Since an isomorphic relation has it that *all* elements belonging to each of the relational structures get mapped to one another, no choices must be made regarding which elements in the range ought to be mapped into and therefore, no agent capable of such choice is required.[92] A homomorphic mapping, on the other hand, will be constructed in such a way that only *some* elements of the cognitive relational system get mapped into by the system realized in the external world. Hence, it may seem that a decision about which elements these are will be required.

This is not an unfamiliar difficulty: A version of it was raised in connection with Fodor's criticism of picture theory semantics, when he argued that pictures, unlike descriptions, do not abstract away from any features of the things they represent. Typically, a picture immediately depicts more about the represented object than our thought about it does, and may say nothing about which of the features it depicts are the ones we care about. Since mental representations are frequently more targeted than this,

---

[91] Or alternatively, what features of a represented *object* exist are those that belong to the range of the relational system which is mapped into by the cognitive system. Above I am speaking at the level of the entire network of represented objects rather than at the level of individual represented things.

[92] It may be necessary, however, to determine which elements there *are*. This problem is closely related to the problem of specifying the "level" of analysis for a particular relational system.

that is, since they do pick out the one thing we are focused on in any given situation, Fodor thinks that pictures are not the kinds of things that can be mental representations.

It may seem that Cummins' isomorphism fares better in this regard, but this assumption is mistaken. While it is true that characterization of the representation relation as homomorphic seems to leave one without a clear criterion for identifying which of the features of represented objects ought to belong in the range of the mapping in any given case, characterization of it as isomorphic eliminates the need for this decision at a fairly heavy price.[93] For if all properties belonging to an object must be mirrored in the representing structure, then the we reopen ourselves to Dennett's arguments against picture theory. Recall that Dennett, too, was concerned with the ability of pictures to support abstract representation. His view was that they would inevitably portray all of a subject's features, making abstraction from one or more of those features impossible. Dennett's view makes sense if the representing structure must in fact be isomorphic to the structure it represents. But without this requirement, I have argued that it is difficult to see why pictures cannot do as good or even better a job at portraying only some of a subject's characteristics as descriptions can.

It seems, therefore, that at least one advantage of viewing the representation relation as homomorphic is that it enables pictures, understood now as relational systems, to pick out only some of the properties of external objects. What remains to be explained is how one ought to determine which elements of each relational system are the ones belonging to

---

[93] Actually, conceiving of the relation as isomorphic doesn't eliminate this need at all, although it may reduce it somewhat. See below for a discussion of why this is so.

the homomorphic relation that constitutes our representation. But this is, as the reader has probably already noticed, simply another way of articulating the uniqueness problem for picture theory semantics. Which elements of the cognitive structure are mapped into elements in the external structure depends on how the elements of each relational system are defined and which relations on the elements in these systems are preserved. In other words, figuring out which elements are the ones relevant to a given representation amounts to deciding which *homomorphic function*, out of many homomorphic functions between the two relational systems, constitutes the representation.

Here, in dealing with Fodor's critique, is where it may appear that Cummins' isomorphic relation performs better, but this too is a mistaken view. Even though conceiving of the representation relation as isomorphic does eliminate mappings between relational systems which are not of equal cardinality, and hence the need for choosing among functions which map the domain of the representational relational system into different elements of the cognitive relational system, it does not prevent the possibility of obtaining several isomorphic mappings between two relational systems of equal cardinality. In other words, it does not prohibit versions of non-uniqueness such as those seen in connection with Cummins' Autobot, wherein it was possible to specify one function mapping each of the turns in the card to each of the turns in the maze, and another mapping each and every one of the turns in the card to the inverse of each and every turn in the maze.

Since both homomorphism and isomorphism give rise to versions of the uniqueness problem, but only homomorphism is consistent with the idea that representations do not consistently pick out all of the elements of those objects which they represent, it seems that, in this respect at least, understanding the representation relation as homomorphic is the preferable choice.

3.2.2 How Homomorphism Accounts for Representation Under Degraded Conditions

As we have seen, there is an ambiguity in this account of the representation relation about whether the relational system instantiated by a cognizer and mapped to the external world is an *individual* representation or an *entire network* of representations. This ambiguity arises in both Cummins' account and in mine. Although I believe that for a picture theory of representation to overcome the uniqueness problem, the focus must be on a mapping between a wider cognitive network and a wider network of things it represents, part of this picture definitely involves the presence of individual mappings between nodes of the cognitive network and individual objects in the external world. I propose that these individual mappings also be thought of as homomorphic. Doing so will allow us to account for the ability of a cognitive system to represent objects successfully with only partial information about their features (something we believe cognitive systems have the ability to do).

One can graphically represent a mapping between a cognitive system and an external world structure as shown in **Figure 3-C**.

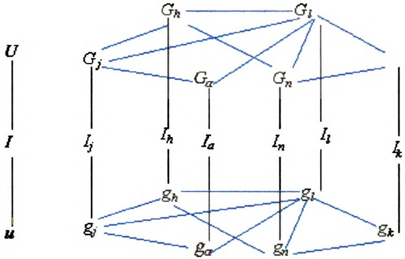

Figure 3-C: A Representational Network

Let each g_i be an element of the cognitive system that, once mapped to some element G_i of the external world, constitutes an individual representation for that system. Let each I mapping g_i into G_i be a homomorphism. ***I*** is also a homomorphism, but one which maps the entire network of *g*s (***u***) to the entire network of *G*s (***U***). In this model, ***u*** and ***U*** are also relational systems.

In general, each g_i may be thought of as a relational system itself. Suppose in particular, that g_n is a representation I have of my cat Sephra in virtue of a homomorphic mapping between it and Sephra herself (G_n). By hypothesis, the relational system g_n maps *into* the relational system G_n. In other words, not *every* feature Sephra has (i.e., every element in the range of I_n) will be captured by my representation of her. For example, if Sephra were to have an above average size liver organ, my representation of her would not likely map into that feature. Our intuition is that, despite such shortcomings in the completeness of our representational picture with respect to many of the things which occupy the external world, we can nonetheless aptly describe ourselves as possessing a legitimate representation of such things. Moreover, by characterizing the representation relation as homomorphic, we allow for the possibility that we can acquire additional information about external objects which was not previously available to us, with the result that representations which were "incomplete" depictions of some particular object at one time, can become more "complete" at a later time. If an ultrasound of Sephra's liver was shown to me, for example, then additional elements of g_n would become part of the range of the function mapping g_n into G_n.[94] In this way, we can see how conceiving of the representation relation homomorphically allows for the involvement of mental representations in learning and knowledge acquisition in general.

Of course, it will always be possible to specify an isomorphic mapping between a subset of the elements in the domain of the cognitive relational system to a proper subset of the elements composing the domain of the real-world external relational system. Given that this is the case, it may seem that there is not a lot of motivation to conceive of the representation relation as homomorphic rather than isomorphic. After all, in any given case, what we do represent "completely," we represent isomorphically. But requiring the representation relation to be homomorphic makes sense if we think of what is represented as being the real world "out there" in all of its complexity, assuming that this complexity is usually not represented by cognizers to the fullest extent (and perhaps never is). Relative

[94] Or, equivalently, we would simply specify a new function whose domain and range now included the relevant elements.

to this assumption, it makes sense to imagine that our representations always fall short, in the direction of the weaker relation of homomorphism as opposed to full isomorphism. Indeed, it may make sense to assume that even our representations of complex substructures of the external world typically fall short in this way.

3.2.3 How Homomorphism Accounts for Substitutivity of Identity Failure

Recall that in my introduction to the problem of mental representation, I stated that any good theory of representation ought to provide a plausible explanation for a peculiar characteristic of mental representations, namely, that when representations refer to the same external object they cannot necessarily be substituted for one another without altering the content of the thought in which they play a part. There, we used the following example:

(*s*) I want to be a Jedi like my father.
(*s*') I want to be a Jedi like Darth Vader.

where *s* is a thought possessed by the character of Luke Skywalker in the film *Star Wars* and *s'* is the result of substituting 'Darth Vader' for 'my father' where it appears in *s*. In the context of the film, both 'my father' and 'Darth Vader' are descriptions of two representations which can be mapped to the same external relational structure, namely, the Jedi knight, Anakin Skywalker.

Since interpretational semantics has it that a cognitive relational structure represents an external relational structure just in case there is an isomorphic relationship between the two structures, it is certainly arguable that so long as Luke Skywalker has a representation of his father, he also has a representation of Darth Vader, and visa versa. That is, since Luke's father and Darth Vader are the same person (read: are one and the same external relational structure), then so long as Luke's possesses a representation of his father, he must also possess a representation of Darth Vader. Indeed, given that representations must be isomorphic to the external relational structures which provide them with the content they have, we may say that Luke representation of his father and his representation of Darth Vader are one and the same.[95]

This is a problematic result for interpretational and structural semantics because it seems to leave no way of distinguishing between thoughts such as those expressed in *s* and *s'* above. But clearly there is a difference of content between these thoughts, for one (*s*) accurately expresses Luke's desire, while the other (*s'*) does not.

If the representation relation is homomorphic, we can argue that failure of substitutivity of identity arises when a subject's cognitive relational system $\mathfrak{R}$ contains two different representational substructures R_1 and R_2 which are both homomorphic to a single external substructure C. R_1 and R_2 are distinct substructures bearing distinct relationships to the other substructures in $\mathfrak{R}$ despite the fact that both are homomorphically related to the same external relational structure. For example, in **Figure 3-D** below, though both R_1 and R_2 are semantically related to Luke's representation of a

---

[95] I base this claim on the assumption that there would be nothing structurally distinctive about the individual representations, and that differences in their monadic properties are not relevant to distinguishing them. However, it may be that if 'my father' and 'Darth Vader' are treated as substructures of a larger representational scheme, there would be ways of distinguishing between them. Specifically, they might be distinguished in virtue of the different relationships they bear to other substructures of the cognitive relational system. I'll address this possibility presently.

Jedi knight, only one is related to his representation of Obi Wan Kenobi's disciple. In the external world, Anakin Skywalker is defined by all of these relations.[96]

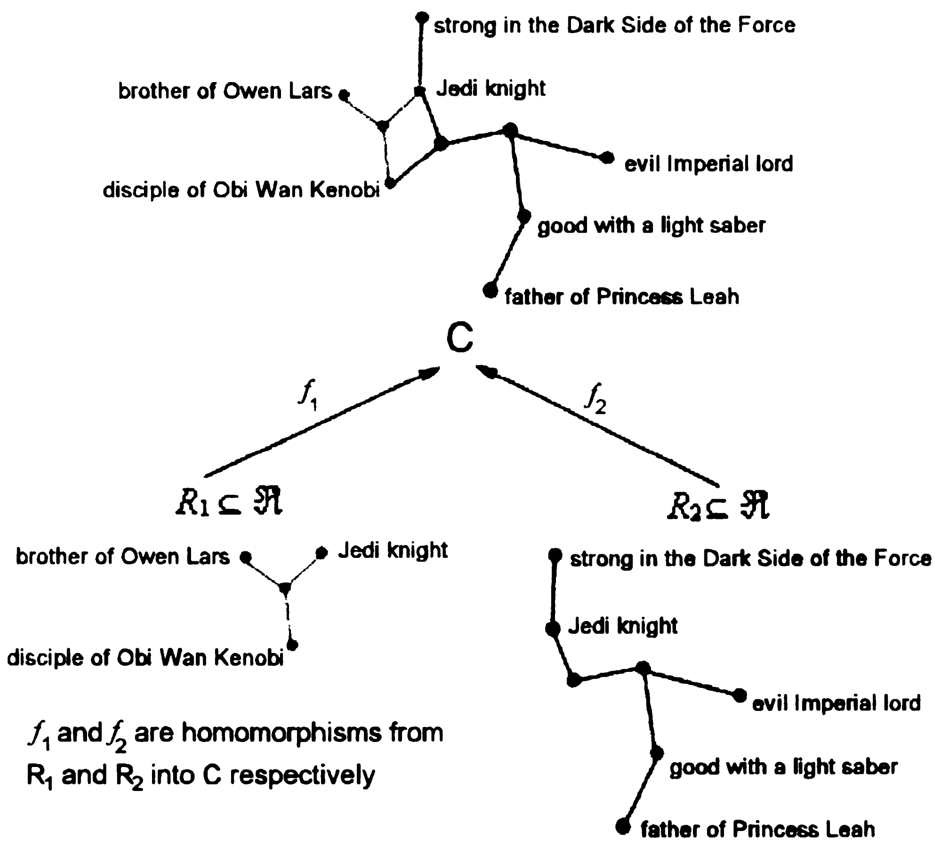

Figure 3-D: Substitutivity of Identity Failure

---

[96] One response available to both interpretational and structural semantics is to argue that the contents of representations is not best specified by considering how *individual* representations are mapped to the external world, but rather, by understanding the contents of representations *holistically*. In particular, one could argue that the failure of substitutivity of identity arises when a subject's cognitive relational system $\mathfrak{R}$ contains two different representational substructures R_1 and R_2 which are both isomorphic (and therefore homomorphic) to a single external substructure C. Ordinarily, we would not be able to say that R_1 and R_2 are "different" substructures under these circumstances, and relative to the external structure to which they are mapped, they are not. But relative to the entire cognitive relational system $\mathfrak{R}$ of which they are a part, they are distinct substructures. They are different in virtue of being related to other substructures of $\mathfrak{R}$ in distinct ways.

The reason, therefore, that "Darth Vader" (R_2) and "my father" (R_1) can't be treated as identicals in the context of Luke Skywalker's representational scheme, is that they are *not* identicals. They nonetheless have something important in common, which is that both are homomorphically related to the same external "real-world" structure.[97]

Characterizing the representation relation as a homomorphism brings with it a number of advantages, such as providing part of the explanation for how representations can be abstract, accounting for the failure of substitutivity of identity in mental contexts, and allowing for the possibility that cognitive systems can represent objects more or less completely, opening the door for a kind of representation-driven explanation of learning about the external world. There are other advantages to "structural semantics," my name for the view that representation is a homomorphic relation between relational systems. In particular, we shall ultimately see how structural semantics can address issues of non-uniqueness in representation. Before considering possible ways of responding to the problem of uniqueness however, there is still one more version of the problem to consider. This manifestation of the uniqueness problem clearly arises in the context of measurement theory and is described by Suppes and Zinnes in their work entitled "Basic Measurement Theory."[98]

---

[97] Notice that this explanation accounts for both the "*de-re*" and "*de-dicto*" readings of *s'*. There is a sense in which Luke *does* want to be a Jedi knight like Darth Vader, insofar as he wants to be a Jedi knight like his father, and Darth Vader *is* his father. This is the "transparent" or *de re* reading of *s'*. The *de re* reading of *s'* is justified by the fact that an actual relationship obtains between the elements of R_1 and those of R_2: namely, both sets of elements are mapped to elements of the *same* external substructure. The *de dicto* or "opaque" reading of *s'*, is justified by the observation that nonetheless, R_1 and R_2 are not the same structure. The fact they stand in different relations to the other substructures composing $\mathfrak{R}$ is enough to establish the non-necessity of treating the two as identicals within the context of $\mathfrak{R}$.

[98] Suppes et al. (1963), pp. 1-76.

3.3 A Final Form of Non-Uniqueness

3.3.1 Suppes' and Zinnes' Project

In "Basic Measurement Theory" Suppes and Zinnes propose to describe a "fundamental" theory of basic measurement. By offering a theory of fundamental measurement, Suppes and Zinnes mean to provide a mathematical interpretation of many of the concepts common to measurement without appealing to alternative concepts already in place in the theory. This is the sense in which they account for "fundamental" (as opposed to "derived") concepts of measurement. Suppes and Zinnes identify two basic problems in the theory of measurement which parallel some of the issues I have discussed above in the context of mental representation. These are "The Representation Problem" and "The Uniqueness Problem." Although there are some divergences between what Suppes and Zinnes attempt to explain in the field of measurement theory and what I have discussed in the context of a theory of mental representation, there are also some very striking parallels which make their view worth examining.

3.3.2 The Problem of Representation in Measurement Theory

Suppes and Zinnes write: "The first problem for a theory of measurement is to show how various features of [an] arithmetic of numbers may be applied in a variety of empirical situations. This is done by showing that certain aspects of the arithmetic of numbers have the same structure as the empirical situation investigated."[99] This statement captures, if only in rudimentary form, the problem referred to as *the problem of representation* for

---

[99] Suppes et al. (1963), p. 4.

measurement theory. More precisely, the problem of representation is the problem of, "characteriz[ing] the formal properties of the empirical operations and relations used in the [measurement] procedure and show[ing] that they are isomorphic to appropriately chosen numerical operations and relations."[100] Suppes and Zinnes point out that this problem is equivalent to proving what it called a "numerical representation theorem."[101]

The authors employ the notion of a relational system in their efforts to provide a numerical representation theorem for certain procedures of measurement. It is their notion of isomorphism (and homomorphism) between relational systems that I employed above.[102] But before we can understand the kind of numerical representation theorem the authors provide, we must understand the specific *kinds* of relational systems they believe are most relevant to a theory of measurement. Our intuitive notion of measurement is that it is a process through which sets of numbers (and relations on those numbers) or some other appropriate formal system (such as a formal language, or a natural language) come to *represent* items in the external world and the relationships between those items. For example, we measure the height of a person by associating a number with the physical distance corresponding to the spatial interval between that person's vertical extremities. We measure the ratio of one temperature difference to another by assigning numbers to

---

[100] Suppes et al. (1963), p. 4.

[101] Suppes et al. (1963), p. 4.

[102] Although the following account of Suppes' and Zinnes' solution to the problem of representation in a theory of measurement is articulated in terms of an *isomorphic* relation between relational systems, they readily acknowledge the usefulness of considering the relation a ***homomorphic*** one when the function mapping one system onto another is not one-to-one (see Suppes et.al. (1963), pp. 7 & 14-16).

each value in the two pairs of values, determining what the relationship is between the two numbers associated with the first pair, and comparing that relationship to the one that holds between the two numbers associated with the second pair. Suppes and Zinnes seek to formalize our intuitive notion of measurement, in part, by conceiving of the items in the external world and the relations between them as a relational system and in particular, as what they call an *empirical relational system*. In short, an empirical relational system is "... a relational system whose domain is a set of identifiable entities, such as weights, persons, attitude statements, or sounds," while the formal system which is used to represent the empirical relational system is a *numerical relational system*.[103] A numerical relational system is defined by Suppes and Zinnes as "... a relational system $\langle A, R_1, ..., R_n \rangle$ whose domain A is a set of real numbers."[104]

Keeping in mind the concepts of numerical and empirical relational systems, the problem of representation can be stated in still another, and final way. Suppes and Zinnes write: "The first fundamental problem of measurement may be cast as the problem of showing that any empirical relational system that purports to measure ... a given property of the elements in the domain of the system is isomorphic (or possibly homomorphic) to

---

103 Suppes et al. (1963), p. 7.

104 Suppes et al. (1963), p. 7. Suppes and Zinnes note that although this definition of "numerical relational system" places no restrictions on the relations which help to compose the numerical system, in practice the relations will be limited to relations which obtain between numbers (Suppes et.al. (1963), p. 7). For example, you would not expect to encounter a relation like "is the sibling of" in a numerical relational system, despite the fact that the definition does not officially disallow it. I believe that the analogue to a numerical relational system in representation theory will not need this practical restriction: In Chapter 5, I'll consider the reasons why in more detail.

an appropriately chosen numerical relational system."[105] The problem of representation for a theory of measurement is, therefore, quite similar to the problem of representation for a theory of cognition. In the former case, it is in virtue of an isomorphism (or possibly a homomorphism) between a numerical relational system and an empirical relational system that the measurement has meaning; that is to say, the members of the numerical system's domain and relations between them *represent* what they do. In the case of mental representations, it is in virtue of the existence of an isomorphism (or perhaps, a homomorphism) *f* between the cognitive relational system and empirical relational systems that the elements in the domain of the cognitive relational system (representations) as well as the relations between them[106] have the content they have.

Demonstrating the existence of the sort of isomorphism (or homomorphism) described above amounts to proving a numerical representation theorem for the measurement procedure. Some comments are in order about the significance of providing a numerical representation theorem however. The authors argue that,

> The representation problem is not solved if the isomorphism is established between a given empirical system and a numerical system employing unnatural or 'pathological' relations. In fact, if the empirical system is finite or denumerable (i.e., has a finite or denumerable domain), some numerical system can always be found that is isomorphic to it. It is of no great consequence therefore merely to exhibit some numerical system that is isomorphic to an empirical system. It is of value, however, to exhibit a numerical system that is not only isomorphic to an empirical system but employs certain simple and familiar relations as well. A complete or precise categorization of the intuitively desirable relations is unfortunately somewhat elusive, so for this reason, the statement of the representation problem refers to an

---

[105] Suppes et al.(1963), p. 7.

[106] These relations could be statements in Mentalese (Fodor) or perhaps simply relations between subsets of representations which compose the domain of the cognitive system—relations which are mirrored in the actual world.

"appropriately chosen" numerical system.[107]

Although much needs to be said regarding the notion of an "appropriately chosen" numerical system, especially if Suppes and Zinnes are to avoid a question-begging account of representation, the main import of this comment is that it raises the issue of whether there are unique isomorphic mappings between numerical and empirical relational systems.[108] As in the context of mental representation, since the existence of an isomorphic (or homomorphic) mapping between numerical system and empirical system is what provides meaning to the process of assigning a measurement to empirical quantities, it is important that some way of dealing with non-unique mappings be identified. In what follows, I will characterize the formal problem which Suppes and Zinnes refer to as the "problem of uniqueness" more carefully.

3.3.3 The Problem of Uniqueness in Measurement Theory

Informally, the uniqueness problem arises in the theory of measurement when it is realized that there can be more than one numerical relational system mapped into by a

---

[107] Suppes et.al.(1963), p. 8.

[108] My guess is that, for Suppes and Zinnes, an example of a numerical system employing "unnatural or pathological relations" would be a system using the relation K, defined as $\{\langle 2, 17\rangle, \langle 26, 5\rangle\}$. We would then have that K(2, 17) is true and that K(26, 5) is true and that K is false for all other pairs of numbers $\langle n_i, n_j\rangle$. This is "pathological" because it has no intuitive or heuristic mathematical value. It's pathological in the same way that the set consisting of {the empty box of apple juice on my table, Times Square, Bill Clinton} is—there is just no clear purpose to its specification. Contrast this with the "healthy" relation corresponding to "less than": $\{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 3, 4\rangle, \langle 4, 5\rangle, \langle 1, 5\rangle \ldots\}$ or the healthy set of square numbers: {1, 4, 9, 16, 25, ...}. Of course, it is not enough to simply give examples of such relations—it seems as though we need a more formal way of guaranteeing that such numerical systems are somehow ruled out as candidates for mappings with empirical relational systems. One consequence of failing to establish the grounds for their elimination (in the context of a theory of mental representation) would be acceptance of the claim that absurd statements (or more generally, absurd cognitive states of affairs) such as these have representational content.

single empirical relational system. This is perhaps the most significant dissimilarity between the problems Suppes and Zinnes consider in the context of measurement theory and the problems I consider in connection with interpretational and structural semantics. The standard uniqueness problem for interpretational semantics is the problem of the existence of more than one empirical relational system for a single cognitive relational system. And in Cummins' version the problem is that more than one mapping relation exists between a single empirical relational system and a single cognitive relational system. Suppes' and Zinnes' uniqueness problem appears to be yet another variant on the basic difficulty of non-uniqueness: In fact, if we take the analog of the cognitive representing system to be the numerical system, it is an inversion of the standard version of the problem (see **Figure 3-E**).

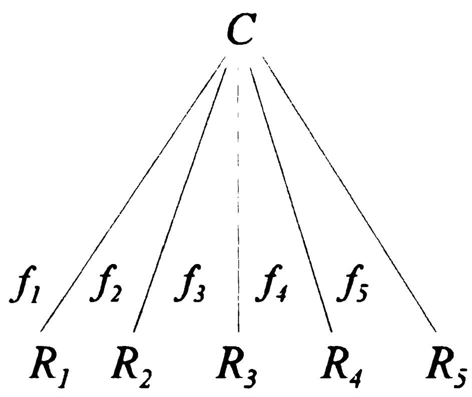

Figure 3-E: Non-Uniqueness in Measurement Theory

In the case of mental representation, non-unique mappings are a problem because the elements of the cognitive relational system get the representational content they have in virtue of a mapping between the cognitive system and an empirical system. If there is more than one empirical system to which the cognitive system can be mapped, the representational content of the elements of the cognitive system is at best ambiguous, and at worst, unspecifiable (because the existence of non-unique mappings might render the whole account of representation untenable). In the case of fundamental measurement theory, elements in the numerical relational system (and statements constructable in that system) will represent the same things (in the empirical relational system to which they are mapped) as elements and statements in competing numerical relational systems. The two situations are different in that in the former case, we find *single* elements in the cognitive system representing *different* things in the external world whereas in the latter case we find *multiple* elements from competing numerical systems representing the *same* things in the external world.

There are a number of additional contexts in which Suppes' and Zinnes' version of non-uniqueness can be identified. In semi-recent years for example, there was a move from analog representation of sound (via the old-style LP and/or cassette tape) to digital sound representation via the modern compact disc recording. Analog systems use a continuum of values to represent any real-world structure. Analog recordings of sound are theoretically quite accurate, since there is no principled limitation on how finely grained the representation of the sound being recorded can be. However, how much of the auditory information can be captured in an analog recording depends heavily on the media used to capture and store the information. In contrast, digital representations of sound can be much better because they assign pre-set values to physical properties in such a way that perturbations in these properties rarely cause error. Both models can be used to represent one and the same set of sounds however. The first record I ever owned was Bruce Springstein's "Born in the U.S.A." These days I listen to Bruce on a CD. But there was but one studio session in which Springstein recorded the sounds I hear on both the analog and the digital medium.

Computer scientists interested in the problem of representing knowledge in artificial systems encounter a Suppes-and-Zinnes form of non-uniqueness as well. For them, a single body of data can be represented using at least four very common methods. The first is via the standard relational database, wherein a set of relationships between elements in a relational system is symbolized by placing these elements in one or more structured tables (see **Table 3-A**).

users_id	First Name	Last Name	Address	State/ Country	Occupation
3456	Jerry	Fodor	56 Cambridge	MA	philosopher
9679	Socrates	Smith	76 Athens Pkwy	Greece	philosopher

Table 3-A: Relational Database Table

An alternative method often used is one which employs the notion of inherited knowledge. This method can capture more about the data it represents and the relationships between the chunks of information from which the data is formed. Specifically, ***inheritance models*** or ***property inheritance*** allows individual elements or classes of elements to inherit attributes and values from more general classes in which they are included.[109] The notion of a semantic network as it is understood in cognitive science is a good example of an inheritance model for knowledge representation.

Two additional forms of knowledge representation, *inferential knowledge* and *procedural knowledge* are also commonly used for the purpose of knowledge representation by computer scientists. Inferential knowledge is very similar to inheritance models but implements the full set of axioms and inference rules for first-order predicate logic in order to describe the relationships between elements in the system it represents. As a result, inferential knowledge systems often capture more of a system's structure than do either of the former kinds of representing systems. Procedural knowledge systems are probably the most complex of all of these methods in that they too capture a system's structure through the use of inference rules, but can, in addition to this, perform certain actions based on the information they possess. Each and every one of these methods can be used to describe one and the same knowledge base and therefore collectively constitute a good example of the sort of non-uniqueness Suppes and Zinnes identify in the context of a theory of measurement.

In choosing which of these models is the best for representing data in computer science, or which method is preferable for audio recording, a number of factors such as efficiency, purpose, ease of implementation, and space play a decisive role. Non-unique ways of representing one and the same external relational structure do not typically give

---

[109] Rich, Elaine & Kevin Knight. Artificial Intelligence. 2nd ed. New York: McGraw Hill, Inc., 1991. p. 110.

rise to serious philosophical concern. However, as illustrated by the examples above, the choice of which representational scheme to use is often informed by practical considerations, and does not necessarily present any deep philosophical problems about meaning. As a result, I will focus primarily on the two sorts of non-uniqueness which I have described as the "standard" and "Cummins" versions of non-uniqueness and in general, will only attempt to defend structural semantics against versions of the problem which *do* present deep problems about meaning and representational content. In the next chapter, I will consider how the notion of translational indeterminacy, as explicated by W.V.O. Quine, presents one such threat to the structural semanticist's account of representational content. This examination of non-uniqueness, manifested as a form of translational indeterminacy, will begin to reveal some ways in which to respond to the various forms of the problem we have seen thus far, and will consider some of the potential difficulties that these responses may inspire.

CHAPTER 4: *Withstanding Quine's Criticisms of Picture Theory*

A proposal for a picture theory of meaning cannot be complete without some consideration of William Van Orman Quine's devastating critique of traditional theories of meaning. In some sense, interpretational and structural semantics are unique and unprecedented approaches to the problem of accounting for meaning and mental representation. However, as has already been demonstrated, both theories bear striking similarities to the picture theory of meaning, and these similarities make interpretational and structural semantics likely objects of criticism—in particular, the relationship of these theories to the picture theory inspires charges not unlike those raised by Quine against what he calls the "museum" view of meaningfulness.

4.1 The "Museum View" of Meaning

In "Ontological Relativity," as in most of his works, Quine is interested primarily in the meanings of terms in a language. Throughout this chapter and for reasons that I will make clearer presently, I will view Quine's comments concerning the meaningfulness of terms as directly relevant to considerations of content for mental representations, and hence Quine's assertions regarding how terms acquire their meanings as applicable to questions about how representations acquire content.

Quine describes the "museum" view:

> [It] is the myth of a museum in which the exhibits are meanings and the words are labels. To switch languages is to change the labels.[110]

According to the museum view, there are three kinds of things that are relevant to an explanation of meanings and our knowledge of meanings.[111] These are (1) the things in the world to which terms with meaning refer ("referents"), (2) the meanings themselves, and (3) the behavior through which we communicate the meanings of terms to others (linguistic behavior might be the most common example of this).[112] Quine's primary dispute with the museum theory concerns whether or not there ought to be a distinction between (2) and (3). Although museum theory does distinguish these, Quine does not. For the museum theory, meanings are in the head—they are mental states which our behavior (e.g., linguistic behavior) describes, conveys, or reproduces in others.[113] Meanings are "private" in the sense that they exist in the minds of individual persons as mental entities. Words have meaning in virtue of being associated with these private mental states—i.e., in

---

[110] Quine, William Van Orman. Ontological Relativity and Other Essays. New York: Columbia University Press, 1969. p. 27.

[111] The "museum" view as Quine characterizes it is really applicable to a whole class of theories about how language acquires meaning. It shares many features with the class of picture theories of representation, although it is a somewhat narrower class as it is characterized by Quine. Quine refers to this class of theories with the term "museum" both because of the analogy with labels on exhibits and, I believe, in order to convey his idea that this view is outdated.

[112] Although this description of the museum view makes it a view about how *words* get their meaning, the analogue of this three-part description in the general case would be (1) the things in the world to which representations refer (the referents of representations), (2) the content of representations, and (3) the behavior through which we communicate the content of our representations to others (linguistic behavior might also be the most common example of this).

[113] Commonly, the museum view is associated with the idea that meanings are mental states in the mind of the speaker. Although Quine's discussion of the museum view is mostly consistent with that trend, he does acknowledge that the meanings of terms can be, in this view, "...Platonic ideas or even ... denoted concrete objects" (Quine (1969), p. 27).

virtue of labeling them—and are useful only insofar as they manage to reproduce in others the same mental states (i.e., the same meanings) as those I possessed when writing or speaking them. In this way, words are "public" whereas meanings are not. In addition, the museum view is an "atomistic" theory of meaning, in that it implies that terms (and representations generally) possess meaning independently of the linguistic and/or conceptual scheme into which they fit. In order to establish the meaning of the term, 'dog,' I need only establish that (1) it is the term I utilize to label a specific mental state; namely, a mental state which constitutes my concept of dog, and (2) it produces in others the same mental state it does in me when so utilized. I do not need to take account of how the mental state the term labels is related to other mental states I may possess. In this sense, terms are utilized atomistically—they are simply labels—to be associated directly with a mental state or set of mental states without regard for how these states are related to others (see **Figure 4-A**).

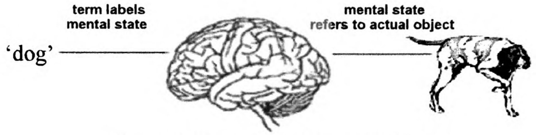

Figure 4-A: Terms As Labels for Mental States

Finally since, on the museum view, terms are simply labels for mental states which refer to concrete objects, it is a consequence of this view that meanings can be gotten right or wrong—that there is a fact of the matter concerning any given term's meaning. To be correct in one's understanding of a term, one must imagine that the term applies to some mental state. Whether one is correct in his/her application of the term to a mental state must be judged relative to the convention governing the use of the terms in one's own native language (this is part (2) above of the two part "museum view" of meaning). However, whether or not a mental state *means* what we say it means is a matter of whether there is a relationship between that mental state and the object it presumably means (part (1) of the view described above). Whether such a relationship exists is, in the museum view, distinguishable from the behavioral conventions of a culture. It is the second part of the relationship, depicted in **Figure 4-A** above, that must be "gotten right," in this view, in order to translate terms accurately.

In contrast to both structural semantics and the museum view of meaning, Quine holds that terms do not admit of fixed, determinate meanings.[114] Although it may make sense to say that one has a meaning wrong (say, if everyone who responds to his usage of the term behaves in strange and unanticipated ways), a term's meaning is never, *can* never be determined solely by its relationship to a concept or referent. Rather, meanings are had by a subject in virtue of that subject's "dispositions to overt behavior."[115] I understand, I *possess* the meaning of 'chair' if I am disposed to behave in such and such a way (for example, if I am disposed to sit rather than stand on the object in question when it is

[114] At least, not in any absolute sense. Meanings may be fixed, for Quine, only relative to a background theory.

[115] Quine (1969), p. 27.

placed at a table, if I am disposed to answer "Yes" when someone points at such an object and inquires, "Chair?", etc.). Although I may be in a certain mental state when (and maybe even only when) I see a chair, knowing the meaning of 'chair' does not amount to possessing a mental representation, or to being in *any* kind of mental state for that matter. Rather, knowing the meaning of 'chair' is to be disposed to certain kinds of behaviors. And there is nothing more to the meaning of 'chair' than knowing its meaning. For Quine, there are no "private" meanings.

4.2 Quine's Translational Indeterminacy

4.2.1 Meaning and Translational Indeterminacy

Central to understanding the motivation for Quine's rejection of the museum theory of meaning in general, and the notion that a term's referent bears a relationship to its meaning in particular, is the concept of translational indeterminacy. Quine has argued that possessing the meaning of a word amounts to getting acquainted with the behaviors of others who use the word and mimicking those behaviors in such a way as to obtain social sanction for one's performance. He describes this process in detail in Ontological Relativity:

> The word refers, in the paradigm case, to some visible object. The learner has now not only to learn the word phonetically, by hearing it from another speaker; he also has to see the object; and in addition to this, in order to capture the relevance of the object to the word, he has to see that the speaker also sees the object. Dewey summed up the point thus: 'The characteristic theory about *B*'s understanding of *A*'s sounds is that he responds to things from the standpoint of *A*' (178). Each of us, as he learns his language, is a student of his neighbor's behavior; and

conversely, insofar as his tries are approved or corrected, he is a subject of his neighbor's behavioral study.[116]

Because my knowledge of meanings depends, in this sense, on observations of my neighbor's behaviors, and because such behaviors must then be *interpreted by* me as, for example, confirming my correct usage of the term in question, I can never be certain that my understanding of a term is consonant with my neighbor's. There will always be, in other words, a degree of indeterminacy in my attempts to translate my neighbor's words into my own. This is part of what Quine refers to as the "indeterminacy of translation." Note however, that the indeterminacy of translation is not merely an *epistemological* problem for Quine—that is, it is not the case that only our ability to *know* that the referent of our's and our neighbor's term is the same is threatened here (although this is, in fact, threatened). Rather, it is the very *existence* of a uniquely correct translation that is problematized by translational indeterminacy. Any attempt to pick out a unique referent for a term will fail, argues Quine, because the behavior of native speakers in the community in which the term is utilized will always be interpretable as consistent with any number of distinct referents. Indeed, Quine argues that it will not even be possible to determine whether the referent is a concrete object or an abstract universal based on the native speaker's behavior alone.[117] Moreover, this difficulty is not limited to the would-be translator of an unfamiliar language. The difficulty of associating words with unique referents applies to the process of learning a language as well—indeed, even one's own

---

[116] Quine (1969), p. 28.

[117] Quine, W.V.O. Word and Object. Cambridge: MIT Press, 1960. sect. 12 pp. 51-57 and Quine (1969), pp. 30-37.

native language.

Quine's views about translational indeterminacy implicitly invoke a distinction between what are merely *epistemological* obstacles to overcoming indeterminacy, and what obstacles are *ontological.* Quine has identified other contexts in which epistemological indeterminacy may be found, that is, contexts in which our claims "go beyond" the available evidence. For example, one form of epistemological indeterminacy arises, according to Quine, when we realize that our scientific theories are "underdetermined" by the available empirical evidence. A scientific theory is underdetermined when there is no experience we can produce which would count as evidence for it and against an alternative incompatible theory. As in the case of translation, this sort of scenario is problematic because it seems that both options cannot be correct, and yet, we have no method of choosing between them. But this may be where the similarity between translational indeterminacy and the underdetermination of theory by fact ends. In an article entitled "Translation, Physics, and Facts of the Matter,"[118] Roger Gibson argues that it is Quine's view that the underdetermination of theory by evidence presents an epistemological dilemma since our only way of knowing which is the best choice is by consulting the evidence, but the evidence is equally consistent with both options. However, Gibson argues that Quine does not conclude, from this, that *there is no fact of the matter* regarding which of the two competing theories is the correct one. In the case of the underdetermination of theories by fact, Quine believes there is a right or wrong

---

[118] Gibson, Roger. "Translation, Physics, and Facts of the Matter." in The Philosophy of W.V. Quine. eds. L.E. Hahn and P.A. Schilpp. LaSalle: Open Court, 1986.

answer to the question of which of the two competing theories is in fact true. In contrast, Quine holds that translational indeterminacy is more than just an epistemological problem about evidence. According to Gibson, Quine holds that meanings are *inherently* indeterminate.

> In effect ... *there is a fact of the matter* to the question of which of two physical theories, both of which are consistent with all possible observations, is *the correct one* but ... *there is no fact of the matter* to the question of which of two translation manuals, both of which are consistent with the speech dispostions of all parties concerned, is *the correct one.*[119]

Quine does seem to suggest exactly this in The Pursuit of Truth, when he writes:

> There is evident parallel between the empirical underdetermination of global science and the indeterminacy of translation. In both cases the totality of possible evidence is insufficient to clinch the system uniquely. But the indeterminacy of translation is additional to the other. If we settle upon one of the empirically equivalent systems of the world, however arbitrarily, we still have within it the indeterminacy of translation.[120]

Gibson correctly suggests that Quine's views about the special status of translational indeterminacy are heavily influenced by his commitments to both physicalism and empiricism. A closer examination of translational indeterminacy will reveal why Quine thinks that it is more than just an epistemological dilemma. In addition, understanding how Quine's physicalism and empiricism motivate his convictions concerning the indeterminacy of translation is essential to appreciating how his perspective differs from that of structural semantics and how this perspective may pose a threat to any theory that attempts to fix the meanings of terms by associating them with their referents. Indeed, if well founded, Quine's views are particularly troublesome to a theory such as structural semantics. As we have seen, structural semantics has it that mental states mean or represent what they map

[119] Gibson (1986), p. 140.

[120] Quine, W.V.O. The Pursuit of Truth. Cambridge: Harvard University Press, 1990. p. 101.

into via homomorphic functions. Objects thus mapped can quite correctly be called the "referents" of these mental states. While Quine's focus is on language, and not on mental states, his arguments concerning the difficulties of uniquely fixing the referents of terms ring familiar as a kind of analogue to the problem of uniqueness. If it can be argued that there is no fact of the matter as to what my mental state refers to, then presumably there is likewise no fact of the matter concerning what my mental state represents. We will certainly need to examine whether the obstacles inherent in fixing the translation of a term, of linking it uniquely to another term in another language, and therefore of linking either to a unique referent, are obstacles which face the attempt to map mental states uniquely to referents in the external world via homomorphism.

4.2.2 Translational Indeterminacy and Non-Uniqueness

In order to appreciate the relationship of translational indeterminacy to the problems of uniqueness, let us first examine some specific examples of the phenomenon of translational indeterminacy in the works of Quine. Quine provides this example in "Ontological Relativity":

> To see what such indeterminacy would be like, suppose there were an expression in a remote language that could be translated into English equally defensibly in either of two ways, unlike in meaning in English. I am not speaking of ambiguity within the native language. I am supposing that one and the same native use of the expression can be given either of the English translations, each being accommodated by compensating adjustments in the translation of other words. Suppose both translations, along with these accommodations in each case, accord equally well with all observable behavior on the part of speakers of the remote language and speakers of English. Suppose they accord perfectly not only with behavior actually observed, but with all dispositions to behavior on the part of all the speakers concerned. On these assumptions it would be forever impossible to know of one of these translations that it was the right one, and the other wrong. Still, if the museum myth were true, there would be a right and wrong of the matter; it is just that we would never know, not having access to the museum. See

language naturalistically [and meaning behavioristically] on the other hand, and you have to see the notion of likeness of meaning in such a case simply as nonsense.[121]

Let me represent pictorially the difficulty Quine is describing here. Let *n* be a term in the native language, e_1 and e_2 terms in English. The functions f_1 and f_2 are translation functions which facilitate a mapping from *n* onto e_1 and from *n* onto e_2 respectively (see **Figure 4-B**).

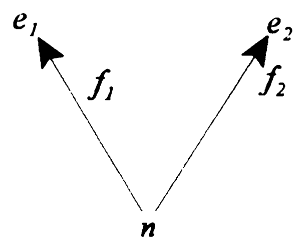

Figure 4-B:
Translational Indeterminacy

Given this picture, we can easily see how the meaning of *n* in English is indeterminate. It might either mean e_1 or e_2. Observing the behavior of English speakers when *n* is translated as e_1 rather than as e_2 will not resolve the indeterminacy, for by hypothesis, either usage will allow a speaker to behave in ways consistent with other English speakers' expectations. Indeed, even in cases where e_1 and e_2 are semantically related to[122] other terms in the language, Quine imagines that *n* can be mapped to both in

---

[121] Quine (1969), pp. 29-30.

[122] By "semantically related to" here I mean only that e_1 and e_2 bear relations to other terms in the language. In particular, you may think of e_1 and e_2 as elements in the domain of a relational system. The relations which order the domain may be characterized, in this example, by speakers' dispositions. For example, if e_1 is the term 'mammal,' it may be related to the term 'dog' as long as the speaker is disposed to behave in ways that would suggest his or her belief in the proposition, "Dogs are mammals."

such a way that either mapping would "accord equally well with all observable behavior on the part of speakers of the remote language and speakers of English." All that is required is that the translations of the semantically related terms be adjusted as well. This more complicated, but more likely possibility is illustrated below, in **Figures 4-C** and **4-D**.

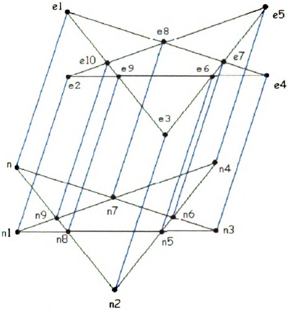

Figure 4-C: *n* Translated As e_1

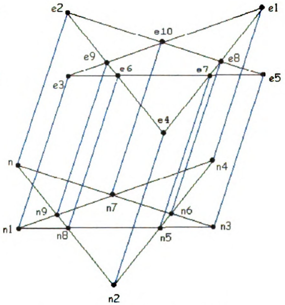

Figure 4-D: *n* Translated As e_2

In these figures, two semantic networks are portrayed. In one (**Figure 4-C**) the term *n* in the native language is mapped to the term e_1 in English. Other terms of the native language are mapped to other terms of English and in both languages, individual terms bear relations to others. The idea is that it should be possible to map *n* into e_2 instead of e_1

When terms are semantically related in this way (i.e., related in virtue of their meanings or related in virtue of the speaker's disposition to behave as though they referred to objects which bear a relation to one another) it seems harder to imagine that either might be one of many possible translations for a native term. This is because the translation of *n* as 'dog' would carry with it the implication that *n*s are mammals and an alternative translation for *n* might not preserve the truth of this implication. Quine's example is meant to show how a persistent indeterminacy is, nonetheless, possible under these circumstances.

without anyone noticing so long as accommodating adjustments are made in the mappings of the other native terms as well. This possibility is illustrated in **Figure 4-D** where *n* is now mapped to e_2 instead of e_1 while preserving the structure of the networks.[123]

Although the homomorphisms f_1 and f_2 (see **Figure 4-B**) are homomophisms between terms in one language and terms in another, and not between mental states and entities in the world, the scenario Quine describes nonetheless bears a striking resemblance to the uniqueness problem for structural semantics. To see why, let us examine two concrete cases of translational indeterminacy.

4.2.2.1 Two Cases of Indeterminacy

Quine's concrete example of the sort of indeterminacy he has been describing involves the French language construction 'ne...rien.' The term 'rien' can, by itself, be translated correctly either as "anything" or as "nothing" in English.

---

[123] The claim that the structure of the graph is preserved between **Figures 4-C** and **4-D** is based on the assumption that there are only two properties of this graph which constitute its structure and hence, only two which need to be preserved. The first of these is the *degree* of each of the graph's nodes (a node's degree is determined by counting the number of nodes to which it is directly connected). The second is the specific set of nodes to which each node is connected. The reader may wish to verify that **Figure 4-D** maintains 5 nodes of degree 4 and 4 nodes of degree 2, just as **Figure 4-C** did and that (excepting *n*, e_1 and e_2 themselves) each node in **Figure 4-D** is connected to the same nodes it was connected to in **Figure 4-C**, despite the fact that most nodes are mapped into by different native terms in each case. Although providing this example is made easier by the deliberate symmetry of the graphs, Quine's indeterminacy could be illustrated with non-symmetric graphs as well, even if one allows that additional kinds of properties (such as weightedness, directedness, etc.) ought to be counted as relevant to the graph's structure. An interesting and relevant question pertaining to the problem of uniqueness concerns whether larger, less symmetric, and/or more complex graphs would make it systematically more difficult to obtain alternative, structure-preserving mappings of this sort. Randall Dipert discusses this issue as well as what types of properties are relevant to the judgment that two or more graphs are structurally equivalent in an article entitled "The Mathematical Structure of the World: The World as Graph." The Journal of Philosophy Vol. XCIV, No. 7, July 1997.

Figure 4-E: Translational Indeterminacy for 'ne rien'

What is required is that the 'ne' is translated accordingly as either "not" or as a redundant phrase. In the one case, 'rien' is mapped to "anything," in the other, to "nothing." **Figure 4-E** illustrates the relevant relationships. Quine calls this example "disappointing" because he believes it is easy to argue that the indeterminacy it describes is circumventable:

> You can object that I have merely cut the French units too small. You can believe the mentalistic myth of the meaning museum and still grant that "rien" of itself has no meaning, being no whole label; it is part of "ne...rien," which has its meaning as whole.[124]

Quine anticipates the proposal that the reason you can't get a determinate translation of 'rien' into English is because 'rien,' by itself, has no (unique) meaning. But it is arguable that 'rien' is part of a larger linguistic unit and that it is *that* unit which gets mapped uniquely onto the terms of English—it is *that* unit which has unique meaning. The terms 'ne' and 'rien' only *look* as though they have non-unique meanings when they are considered apart from the unit to which they belong (as shown in **Figure 4-F**).

[124] Quine (1969), p. 30.

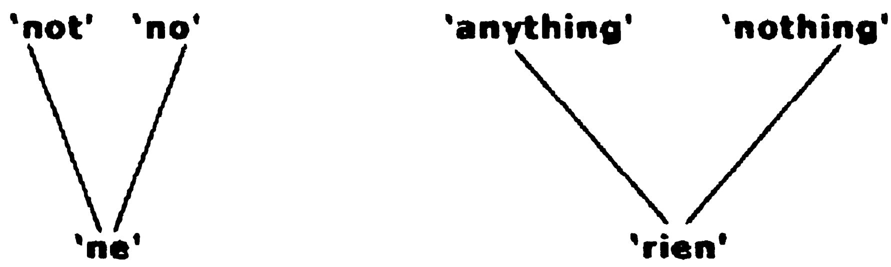

Figure 4-F: 'ne rien' Treated Atomistically

But if treated as part of an indivisible unit, the indeterminacy disappears (see **Figure 4-G**).

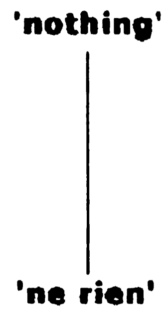

Figure 4-G: 'ne rien' Treated Holistically

Attempting to resolve the uniqueness problem through an appeal to holism is an attractive option to the structural semanticist. Since both representations and external objects are thought of as complex relational systems in this view, it seems intuitively true to think that the more complex the cognitive relational system we are considering, the harder it will be to specify functions which map it to more than one external object and visa versa.

Part of a resolution to the uniqueness problem would thereby be achieved by conceiving of each representation (here analogous to terms in a language) as possessing content in virtue of its role in a larger unit (specifically, a larger relational system). It is the mapping of the sum total of our representations to a network of external objects that provides individual representations with the content that they have. If we adopt a holistic strategy, and consider first how the entire representational relational system best maps onto the external world, taken in its entirety, then the result will be that certain constraints are placed on the relationships between individual substructures of each relational system which were not there when we considered them outside of the context of the more complete relational system of which they are a part.

The result is that uniqueness of representational content is achieved, for all practical purposes on this view, by finding a homomorphic relationship between one's entire network of representations and the objects in the world. Ultimately, to achieve uniqueness for individual representations will require one to acknowledge that the contents of each of these representations are derived from a[125] homomorphic relation between an external relational structure and the network of representations one possesses.

But in order for this tactic to resolve translational indeterminacy altogether, it would have to be the case that *every* instance of indeterminacy could be handled by showing that the indeterminate elements are part of a larger unit and do not themselves have independent meaning. If not, then there will be some cases of indeterminacy that are intractable, and analogously, some cases of representational non-uniqueness which remain unresolved by the "holistic" approach.

Quine believes that an appeal to holism is *not* effective for resolving indeterminacy,

---

[125] 'the' if the mapping is in fact unique.

and gives his famous 'gavagai' example as evidence. When the native points to a rabbit and utters 'gavagai,' we will not be able to tell, he argues, whether the native term maps to the English term 'rabbit' or to the English phrase 'undetached rabbit part.' This is because the native's pointing action is simultaneously a pointing toward a rabbit, and a pointing toward an undetached rabbit part.[126] Which is the intended referent of the native when he utters 'gavagai?' So far, this example is similar to the one involving 'ne rien':

> This artificial [gavagai] example shares the structure of the trivial earlier example 'ne ... rien.' We were able to translate 'rien as 'anything' or as 'nothing,' thanks to a compensatory adjustment in the handling of 'ne.' And I suggest that we can translate 'gavagai' as 'rabbit' or 'undetached rabbit part' or 'rabbit stage,' thanks to compensatory adjustments in the translation of accompanying native locutions. Other adjustments still might accommodate translation of 'gavagai' as 'rabbithood,' or in further ways. I find this plausible because of the broadly structural and contextual character of any considerations that could guide us to native translations of the English cluster of interrelated devices of individuation. There seem bound to be systematically very different choices, all of which do justice to all dispositions to verbal behavior on the part of all concerned."[127]

To understand why the 'gavagai' example is different, and specifically, why, in Quine's judgment, the appeal to a more holistic context is ineffective for it, we must understand something about Quine's views concerning the nature of language, and how, like his views on translational indeterminacy in particular, they are informed by his deep commitment to physicalism and empiricism. We need also to understand what standard would generally have to be met, in Quine's view, in order for translational indeterminacy to be

---

[126] In fact, the situation is even worse, as Quine points out in Word and Object. The native's ostensive act would also be compatible with the translation of 'gavagai' as "all and sundry undetached parts of rabbits" or as "a single term naming the fusion ... of all rabbits" or still further as, "a singular term naming a recurring universal, rabbithood." Hence not only is the particular referent of 'gavagai' at issue, we cannot even be sure of whether the term itself is a general or singular one, or of whether it refers to a concrete or abstract object. Quine, W.V.O., Word and Object, Cambridge: MIT Press, 1960. p. 52.

[127] Quine (1969), p. 34.

surmountable—or put in slightly different terms—for there to be a "fact of the matter" regarding the translation of terms.

4.2.2.2 Quine on Language and Meaning

As we have already seen, Quine believes there *is* a fact of the matter where scientific theories are concerned and believes this despite the underdetermination of theories by empirical evidence. We have also seen that Quine contends there *is no fact of the matter* concerning matters of translational indeterminacy. If we are to understand why Quine believes that underdetermination is surmountable, but indeterminacy is not, then clearly we will need to appreciate his notion of a "fact of the matter." Understanding this will provide a deeper look at what motivates Quine's belief that translational indeterminacy is intractable in the 'gavagai' case, and ultimately will more clearly identify the obstacles translational indeterminacy poses to a theory of representation such as structural semantics.

According to Gibson, many of Quine's commentators have gotten his notion of a "fact of the matter" all wrong. In particular, Noam Chomsky, Richard Rorty, and Dagfinn Føllesdal have all assumed that Quine has a epistemological or methodological understanding of the notion, such that facts of the matter are those things for which there is a rational procedure for reaching agreement about what to assert (Rorty), things for which we have ample empirical evidence (Chomsky), or things for which we have ample evidence and which meet the standards of simplicity in addition to other extra-evidential concerns (Føllesdal). Still others (Bruce Aune) have assumed that Quine's notion of a "fact of the matter" is more transcendental, implying that there are facts which may be beyond all evidence.[128] None of these captures Quine's own view about facts of the matter, according to Gibson. Instead, he contends,

> Quine's understanding of this term is decidedly *naturalistic* and *physicalistic*. When Quine says that there is a fact of the matter to physics and no fact of the matter to translation, he is talking about physical facts, and he is talking from within an already accepted naturalistic-physicalistic theory.[129]

So facts of the matter, where they exist, must be understood as physical facts. This is to be distinguished from the *evidence* we have for what is true about the objects in our ontology, as surely as epistemology is to be distinguished from ontology. According to Gibson's Quine,

> ... despite the fact that it is meaningful to say of alternative theories that they are *equally warranted* by the same sensory evidence, it makes no sense to say that they are *equally true*. It makes sense to say they are equally warranted, because we are speaking from *within* the same (physicalisitic) theory of evidence: given all the (possible) evidence, *this* ontology is warranted, *that* ontology is warranted, and so on. However, it makes no sense to say they are equally true, because we are *not* speaking from within the same theory of objects (i.e., ontology). The trouble isn't with 'equally', it is with 'true'.[130]

Why can't the same be said for translational indeterminacy? That is, why can't it be said that though one or another translation is equally warranted by the same sensory evidence, and in particular by observations of the same behavioral dispositions on the part of speakers, there is yet but one correct translation of a term? Taking the 'gavagai' case as our example, it would go something like this: "Although the available empirical evidence

---

[128] Aune, Bruce. "Quine on Translation and Reference." Philosophical Studies. Vol. 24. April 1975. pp. 221-236.

[129] Gibson (1986), p. 143.

[130] Gibson (1986), p. 152.

(that is, the totality of behavioral dispositions to stimuli and to others' speech behavior) may underdetermine which of one or more referents is the intended referent of 'gavagai,' there is nonetheless a physical fact of the matter concerning which referent is the correct one." For Quine, the reason that the translational indeterminacy for 'gavagai' is intractable is rooted in his decidedly naturalistic views about language. If there were going to be facts of the matter concerning language and concerning the synonymy of linguistic expressions in particular, then these would have to be found *in* observable behavior.[131] This follows from what Quine believes language *is*. When we think about what distinguishes a language from a mere collection of meaningless sounds and physical distortions or from an unorganized scribbling on a medium, we recognize that it is a collection of linguistic *behaviors*. In attempting to discover from what source the collections of terms in a language derive their meaningfulness, one cannot, if one is a staunch physicalist such as Quine, point to things like abstractions, Ideas, or even non-physical mental states. The behavior of a linguistic community around a term is a quantifiable, naturalistic alternative. But by hypothesis, the behavioral dispositions of this community are insufficient for distinguishing between the possible referents of the term. And since there is nothing else to do the job which also meets Quine's strict empiricist standard, no distinction can, even in principle, be made. That there is no *physical-behavioral* fact of the matter concerning which of the available translations is correct, means that there is no fact of the matter concerning which is correct at all.

[131] Gibson (1986), p. 144., Aune (1975), p. 224.

> The current theory of the world is physicalistic—and with good reason, Quine thinks. And this physicalistic world-view settles, for the present, the physical facts of the matter and thereby what can be said to be true or false given these facts. Any putative meanings, therefore, that fall between the cracks of the physical facts just aren't meanings at all. And, further, since there just aren't any facts for such putative semantical statements to be about, it follows that such statements are indeterminate, that is, they are neither true nor false. In this straightforward, naturalistic-physicalistic sense, then, there is no fact of the matter to the question of which of two manuals of translation is *the* right one.[132]

Quine thinks that "putative meanings ... fall between the cracks of the physical facts" quite simply because the only physical facts available where language is concerned are facts about behavior, and behavior does not always distinguish between the content of one term and another. This conviction, that what counts in matters of meaning and translation is the predictable, conditioned responses of speakers, is also expressed clearly in Chapter 1 of Word and Object.

In Word and Object, Quine portrays a natural language as consisting of words and sentences which, when uttered within a social context, evoke predictable and regular behavioral responses from other speakers. This predictable response to linguistic stimuli on the part of other speakers is what distinguishes mere noise from meaningful linguistic units of speech. Specifically, the meaning of a term is given in the disposition to a behavioral response which that term elicits in others. Quine thinks that most of the time, the kinds of linguistic units that consistently elicit the same socially sanctioned responses are not individual words, but sentences. Take the English word "Red" as an example: For Quine, when a subject applies this word to a specific stimulus, it is more natural to think of the subject as having uttered a one-word sentence, and in particular, to have uttered a shortened version of the sentence, "This is a red thing." After all, it is the *application* of

[132] Gibson (1986), p. 153.

the term to a stimulus which is approved or disapproved by the subject's linguistic peers, and therefore it is the *application* of the term to a stimulus which has meaning. "Red," *qua* one-word sentence captures the intention to apply the adjective to a stimulus, and hence captures what it is that is meeting with the community's approval. "Red" *qua* isolated term, captures no such application to a stimulus. As an utterance untethered to experience, it leaves us with no basis for the approval or disapproval of its usage. And since linguistic units derive their meaning from such social sanction of their usage, "Red" is best understood as a one-word sentence, if it is to be understood at all.

The interesting and relevant thing to note here is that the way Quine divides up the units of a language has everything to do with what elements of the language reliably covary with conditioned responses on the part of the linguistic community. Specifically, just in case a unit of speech can be mapped consistently to a set of behavioral responses, then it counts as a *meaningful* unit of speech. Understanding this is part of what is necessary to appreciate why the 'gavagai' example is an intractable case of translational indeterminacy for Quine, while the 'ne rien' case is not. In the latter case, it is arguable that there is no consistent and identifiable behavioral response to either of the two units 'ne' or 'rien' that compose the French phrase when taken in isolation. That they are therefore ambiguous in any translation attempt is far from problematic—on the contrary, it is exactly what Quine's view would predict. As in the case of "red," it is more appropriate to think of the "longer" unit, 'ne rien', as the unit of meaning, rather than to attempt to translate the phrase compositionally. For it is 'ne rien' that covaries with specific, predictable responses on the part of French speakers. When we do so, we find that translation comes easily and the ambiguity drops away.

'Gavagai,' on the other hand, *does*, by hypothesis, elicit specific, predictable responses on the part of the community and does so all by itself. Taken as a one-word sentence implying the application of the term to a stimulus, it will elicit approval or disapproval from other native speakers in predictable ways. The problem arises when we recognize that these predictable, conditioned responses are consistent with a myriad of translations for the word 'gavagai' into English. As we've seen, it might be translated as 'rabbit' or possibly as 'undetached rabbit part' among other options. The utterance, 'Gavagai,' taken as a one-word sentence applying to a stimulus, is compatible with either translation, making determinate translation impossible.[133]

Why not narrow down the options, as in the 'ne rien' case, by considering the relationship of 'gavagai' to other terms in the language in hopes that its relationship to these terms might constrain its translation into English? Certainly Quine would not deny, especially in light of his association with meaning holism,[134] that fixing the meanings of

---

[133] The fact that it is even possible to imagine the term 'gavagai' having a whole host of different referents might lead one to conclude that indeterminacy of translation, far from being a problem for a view such as structural semantics, is quite consistent with the view. This is because translational indeterminacy seems to provide a counterexample to the view that meanings are given exclusively by the behavioristic trends of a linguistic community. If 'gavagai' fails to evoke behaviorially distinct responses from speakers whether taken as referring to rabbits, Rabbithood, or undetached rabbit parts, then meaning must not be a matter of behavioral responses after all. But this line of reasoning makes the mistake of equating meaning with reference, something which Quine would not allow. The fact that 'gavagai' can be taken to have a number of different referents, while still evoking the same behavioral responses from speakers, would suggest that meaning and reference are *not* equivalent, so long as one accepts Quine's account of meaning.

[134] It is certainly arguable that Quine is a meaning holist given his position in "Two Dogmas" wherein he critiques the logical positivist's view that sentences can be confirmed or disconfirmed individually, without the benefit of considering their relationship to other claims in a scientific theory. In this work, he writes, "But the dogma of reductionism has, in a subtler and more tenuous form, continued to influence the thought of empiricists. The notion lingers that to each statement, or each synthetic

some terms in a language constrains the available interpretations of those which remain. And indeed, he does not deny this. If we consider all of the native sentence utterances containing the word 'gavagai,' and require that they all be translated into true (or at least plausible) English sentences, Quine's claim is that there will be more than one translation/interpretation (actually an infinite number) which will satisfy the whole corpus. Interpretations of individual terms *within* a translation might constrain the interpretations of other terms within that same translation, but there is presumably no way of deciding between translations on a language-wide level. In theory, doing so would require that one be able to find a decisive reason for choosing one over a multitude of available mappings between a term in the language and a specific stimulus. This initial assignment of term to stimulus might then constrain the mappings of other terms to other stimuli, assuming that something like meaning holism is true. But the gavagai example is meant to illustrate the inherent difficulty in establishing even one such initial, constraining mapping between term and stimulus. For apart from interpreting 'gavagai' relative to an already-established translational framework, there is simply no behavioral criterion for assigning it to the set

---

statement, there is associated a unique range of possible sensory events such that the occurrence of any of them would add to the likelihood of truth of the statement, and that there is associated also another unique range of possible sensory events whose occurrence would detract from that likelihood. This notion is of course implicit in the verification theory of meaning. The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My counter-suggestion ... is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body." ("Two Dogmas of Empiricism" in The Theory of Knowledge. ed. Louis P. Pojman. Belmont: Wadsworth, Inc., 1993. p. 404).

Quine's view, that it is only our entire system of beliefs that can be tested against experience, and not each individual belief taken in isolation, might be better described as a kind of "verification holism" or "evidential holism." However, since Quine seems to retain, to some extent, the equation of verification and meaning here, it is not entirely implausible to view him as adhering to some version of meaning holism as well.

of rabbits instead of to the set of undetached rabbit parts. Indeed, there is no such criterion for assigning *any* arbitrary term to one stimulus rather than to another, and hence no way of establishing constraints which would recommend one translation of the whole corpus over another translation. What this seems to show, is that unique translations of terms are indeed possible, but only relative to an arbitrarily chosen translational framework. One can have unique translation of individual terms which is "internal" to this framework, but cannot have unique translations which go beyond, or transcend such frameworks, and this despite wholeheartedly embracing a holistic approach to meaning.[135]

4.2.2.3 Two Kinds of Non-Uniqueness

The 'gavagai' example bears some extended scrutiny, because it reveals much more about the extent to which the indeterminacy of translation is a special case[136] of the problem of uniqueness for structural semantics. In effect, the 'gavagai' case seeks to recreate indeterminacy (and hence, non-uniqueness) *in spite of* the insistence that linguistic units may sometimes derive their meanings from the role they play in a larger unit (i.e., in spite of the "holistic" response to indeterminacy suggested above.

If Quine is correct in assuming that an appeal to holism fails to eliminate translational indeterminacy in this case, then this bears on one of the responses which structural

[135] Compare this to "internal realism," the view that though we can talk about a reality independent of us, this talk must always take place within a rational framework of beliefs and concepts: There is no God's Eye view of the world.

[136] In particular, it covers the case of *linguistic* representation while the problem of uniqueness is a problem for accounts of representation generally.

semantics has to offer to the uniqueness problem. In particular, if the structural semanticist intends to eliminate as many competing mappings as possible by arguing that it is the mapping of *entire networks* of representations to *entire networks* of objects which confers meaning, then any claim that indeterminacy persists in such a scenario, or that individual representations in the network ought to possess meaning independently of it, is a threat to her account of representation.

One way of interpreting the difficulty that Quine's arguments raise is to see them as demonstrating the persistence of Cummins' version of the uniqueness problem. Recall that the standard version of the problem arises when a single representation can be mapped to more than one external object. Each mapping between representation and external object is facilitated by a distinct homomorphic relation.

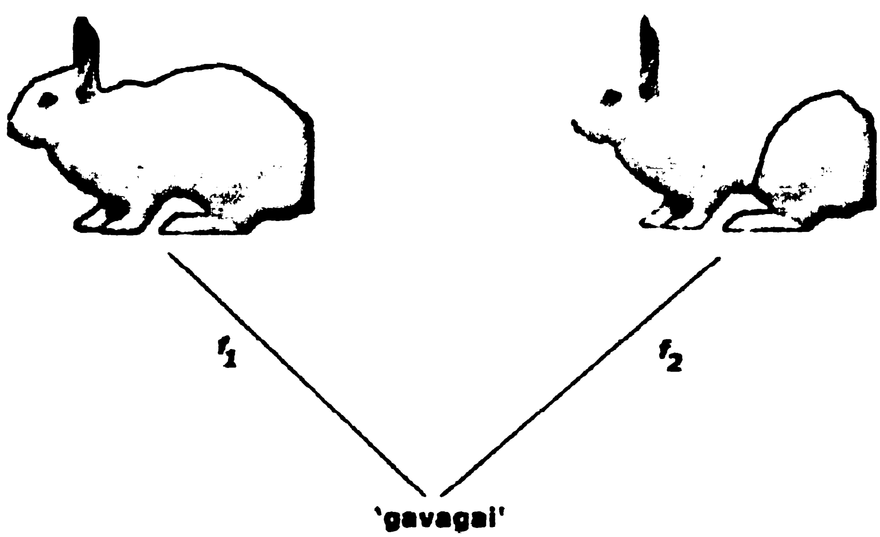

Figure 4-H: Standard: Rabbit vs. Undetached-Rabbit-Part

Cummins version of the problem, in contrast, posits the existence of several homomorphic mappings between a single representation and a single external object. **Figure 4-H** illustrates the 'gavagai' case as an example of the standard uniqueness problem for structural semantics.

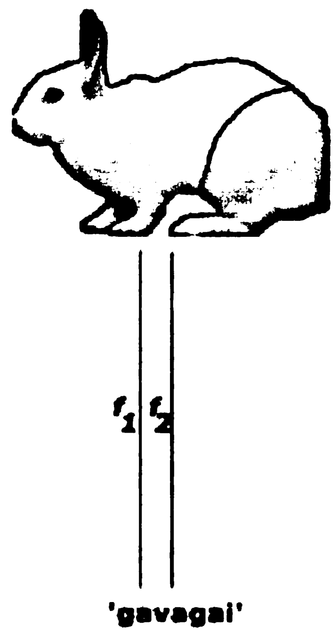

Figure 4-I: Cummins: Rabbit vs. Undetached-Rabbit-Part

It shows multiple mappings from the term 'gavagai' to objects in the world (the translation functions f_1 and f_2 map 'gavagai' to rabbits and undetached rabbit parts, respectively). **Figure 4-I** demonstrates the problem as an instance of Cummins' brand of non-uniqueness.

The standard version of the uniqueness problem illustrates a situation in which two or more distinct structures are viable candidates for the translation/meaning of a single term/representation. When the 'gavagai' case is considered to be a problem about the translation of a single term as either one of two distinct sets, rabbits or undetached rabbit parts, it makes sense to see the case as an instance of the standard uniqueness problem. However, when we recognize that the real importance of the 'gavagai' case lies in its demonstration that a single language can be mapped in multiple, non-equivalent ways to a single ontology, something like Cummins' version of the uniqueness problem comes more to mind. Ultimately, which version of the uniqueness problem you take the 'gavagai' example to illustrate depends on how you cut up the world. Quine's deep physicalist convictions lead him to conclude that the only patterns and structures that can serve as a means of establishing constraints on the interpretation of linguistic units, and on the establishment of content for representations generally, are physical, empirically accessible structures and patterns. And the only kind of structure which meets the physical-empirical standard according to Quine, is a structure formed out of the totality of behavioral dispositions on the part of a specific community of speakers.

> When with Dewey we turn thus toward a naturalistic view of language and a behavioral view of meaning, what we give up is not just the museum figure of speech. We give up an assurance of determinacy. Seen according to the museum myth, the words and sentences of a language have their determinate meanings. To discover the meanings of the native's words we may have to observe his behavior, but still the meanings of the words are supposed to be determinate in the native's *mind*, his mental museum, even in cases where behavioral criteria are powerless to discover them for us. When on the other hand we recognize with Dewey that 'meaning ... is primarily a property of behavior,' we recognize that *there are no meanings, nor likenesses nor distinctions of meaning, beyond what are implicit in people's dispositions to overt behavior* [emphasis mine].[137]

Since the totality of these behavioral responses would not, by hypothesis, distinguish

---

[137] Quine W.V.O., "Ontological Relativity", p. 29.

between rabbits, undetached rabbit parts, and rabbithood, there is no empirical basis for distinguishing between these things at all—as far as meaning and representational content are concerned anyway. This is why Quine's translational indeterminacy is something more akin to Cummins' version of the uniqueness problem: There is but one empirical relational system to which 'gavagai' can be mapped in several different ways, and no ontological basis for characterizing the external world as being composed of more than one such system in this particular example. Although, in Cummins' case, these several mappings are facilitated by the complexity of the relational system, the structure responsible for this complexity is not something we can exploit in the service of uniquely specifying a mapping, because it is not, again by hypothesis, something which makes any difference to the behavioral responses of speakers using the representation. If we see translational indeterminacy as a property of whole systems in this way, then it is arguable that Quine's critique shows the holistic approach to be effective only against non-uniqueness *within* translational systems (read: non-uniqueness of the standard variety) but not against non-uniqueness of the sort which Robert Cummins describes.

4.3 Where Does This Leave Structural Semantics?

By way of summary, it seems that a convincing case can be made for the intractability of translational indeterminacy if we can grant Quine a few of his more basic assumptions. In particular we must start out, with Quine, from a decidedly physicalist perspective, and adopt with him the further view that the only physical facts of the matter relevant to the meaningfulness of terms are facts about the behavioral dispositions of speakers.[138] Adopting this point of view is easier to do if one is already sympathetic to empiricism. A physicalist who was not sympathetic in this way, might seek to identify other physical facts about language which, perhaps, are not as readily apprehended by the senses as speech behavior undoubtedly is.

In order to assess the effect of Quine's critique on structural semantics, we must first situate that critique relative to these assumptions, for while structural semantics shares some of them, it does not share them all. And a departure from Quine on any of these points—physicalism, empiricism, or behaviorism—might change the degree to which his arguments concerning translational indeterminacy are compelling.

Regarding the commitment to physicalism: I have always thought of structural semantics as a position which is sympathetic to physicalism and most of the examples I have used of "structures" which could be exploited in the service of discovering mappings between representations and the external world, have been physical structures, or, at least, structures which could be realized in physical media. In point of fact, however, there is nothing that has been said so far regarding structural semantics' account of how representations have content, which would commit it exclusively to a physicalist metaphysics. I see this as an advantage of structural semantics since it allows the view to recommend itself to a range of metaphysical positions. However there may be an advantage to this flexibility which we can only now fully appreciate in light of having

[138] It's also possible that a belief in the intractability of translational indeterminacy requires a commitment to a strong version of verificationism, such that the meaning and evidence are one and the same.

considered Quine's views on translational indeterminacy. For if translational indeterminacy is a consequence of the belief that the only properties of language which count in the attempt to fix meanings are the physical properties of language, and in particular the behavioral dispositions of speakers of the language, then translational indeterminacy does not necessarily follow from a non-physicalist starting point. It certainly seems easier to accept that there are no distinguishable *physical-behavioral responses* to rabbits when compared with responses to undetached rabbit parts, than to accept that there are no means of distinguishing them whatsoever.

Of course, there are, at least in my view, a significant number of more serious problems which structural semantics might have to contend with if held in conjunction with a non-physicalist metaphysics. So I have defended, and will continue to defend the view under the assumption that some form of physicalism is true. What I do not believe, and will not defend, is the implication of a behaviorist approach to meaning from a physicalist metaphysic combined with an empiricist epistemology. Implicit in my discussion of structural semantics until now has been the assumption that we are capable of appreciating structural properties of networks of representations which go beyond just the patterns of behavior exhibited by the people who possess them. Specifically, I've talked about the formal properties of such structures and have assumed that these types of properties are what individuate representations from one another and which ultimately enable them to stand in relation to that which gives them the content that they have. Even assuming physicalism, translational indeterminacy is only a threat to the structural semanticist's account of meaning if it can be shown that, in addition to there being no way of using *behavior* to distinguish between two or more proposed referents of 'gavagai,' there is no appreciable physical difference whatsoever between them such that one would recommend itself above the other as the most appropriate translation for the term. In the next chapter, I will examine whether or not such a case can be made against structural semantics.

It is worth noting that there is starting to emerge a very significant difference between the specific kind of view that translational indeterminacy threatens (namely, a view about the meaningfulness of terms) and the view that is represented in structural semantics. I have been assuming that Quine's arguments for the indeterminacy of translation carry over to mental representations, thereby posing a new argument for the problem of uniqueness. But this assumption may be false. For it is possible that the structural semanticist could admit all of what Quine assumes, *even that there are no significant properties beyond speakers' behavioral properties to consider when learning the meanings of terms in a language*, and still maintain her affirmed account of representational content. This is because structural semantics is a theory about the representational content of mental states and not a theory about the meaningfulness of terms in a language. Perhaps, insofar as natural languages are uniquely social phenomena, the properties of natural languages (including the meanings of terms in such languages) *are* primarily characterized in terms of dispositions toward behavior on the parts of speakers. And if so, perhaps translational indeterminacy is an interesting phenomenon which Quine has identified as inherent in natural language. This does not necessarily imply that the same is true for mental states. Once again, as a physicalist, I would urge that mental states are physical (possibly physical-functional) states of the brain. As such, there ought to be a whole host of properties which can be discovered about them and which go beyond just the behavioral evidence for their existence. In my view, mental states are arguably more like the objects of a scientific theory than like the terms in a language and, as we have seen, even Quine would admit to there being physical facts of the matter concerning what is true of the objects of a scientific theory. For structural semantics to maintain viability as a theory of mental representation in the face of translational indeterminacy via Quine, we have only to contend, minimally, that there is a fact of the matter as to which mental states are structurally similar to which external, physical states. Next, let's consider whether or not this minimal condition can be met by the theory of structural semantics.

CHAPTER 5: *Formal and Metaphysical Criticisms of Structural Semantics*

Much of the preceding discussion has relied on the assumption that the notion of "similar structure" can be successfully understood in terms of homomorphic relations between cognitive and external relational systems. Though Quine's discussion of translational indeterminacy reveals a potential obstacle to the acceptance of this assumption, I've argued that a departure from any of Quine's basic starting assumptions could do much toward removing the force behind the obstacles his version of non-uniqueness presents to structural semantics. Thus far, I've also portrayed the main positive response to non-uniqueness problems as involving the following sort of move, which, for convenience, I will refer to as "the holism solution":

> **The Holism Solution:** Attempt to eliminate as many alternative mappings from a representation to objects in the external world as is possible by locating the representation within a network of other representations, themselves mapped to a network of objects. Which is the appropriate mapping for individual representations will then be determined by deciding which of the original competitors is compatible with the mapping relation that holds for the entire representational network.

In this chapter, I'll consider some reasons for thinking that a solution to the uniqueness problem will not be this easy. Already, Quine has made a convincing case for the claim that there is a persistent version of the uniqueness problem, namely Cummins' version of the problem, which threatens the project of fixing unique referents for representations, and thus, for structural semantics, threatens the possibility of assigning unique content to them. Furthermore, he has argued effectively that the holism solution does not adequately address this version of non-uniqueness. Now we must consider whether there is a way for indeterminacy to persist even where physical-behavioral distinctions are not the only kinds of distinctions we allow between two or more content structures. As background to this discussion, let's first examine an interesting consequence of the Löwenheim-Skolem theorem.

5.1 The Löwenheim-Skolem Theorem

The Löwenheim-Skolem theorem states:

> *LS*: Let Γ be a satisfiable set of formulas in a language of cardinality $\varkappa$.[139] Then Γ is satisfiable in some structure of cardinality $\leq \varkappa$.[140]

The theorem says that there can be a model (an interpretation) for the sentences in Γ which is lower or equal in cardinality to that of the language from which Γ is constructed. Since by the "cardinality of a model" we simply mean the size of that model's domain, we can restate the Löwenheim-Skolem theorem as follows:

> Let Γ be a satisfiable set of formulas in a language of cardinality $\varkappa$. Then Γ has a model whose domain is of a cardinality less than or equal to $\varkappa$.[141]

This theorem, in conjunction with the compactness theorem, can be used to prove many interesting results. This version of the theorem, usually referred to as the downward Löwenheim-Skolem theorem (DLS), has a counterpart (the upward Löwenheim-Skolem theorem or ULS) which provides that there exist models for Γ which are ***higher*** in

---

[139] By "a language of cardinality $\varkappa$" we mean, a language containing at least $\varkappa$ terms (expressions which can be interpreted as naming an object).

[140] Enderton, Herbert B. A Mathematical Introduction to Logic. New York: Academic Press, 1972. p. 141.

[141] This is what Quine refers to as "the strong version" of the theorem in "Ontological Relativity." Quine (1969), p. 60.

cardinality than the language from which Γ was constructed. Although it is arguable that both versions of the problem present a threat to the uniqueness of models, Quine describes an argument which relies primarily on the DLS result, as we shall presently see.

Let us consider one result of the DLS. Let Γ be a satisfiable set of formulas in a language L of cardinality $\aleph_0$, the cardinality of the natural numbers. One may think of Γ as some set of true sentences of L, in particular, as some *theory* constructed in L. Suppose the intended interpretation of this theory has an uncountable domain, e.g., the real numbers.[142] The downward version of the L-S theorem tells us that there will be an alternative model that has a domain of cardinality $\leq \aleph_0$.

This is a remarkable consequence because, given that the domain of any acceptable theory is, in effect, reducible to a denumerable ontology (via the DLS), it suggests that any domain "... is reducible in turn to an ontology specifically of natural numbers, simply by taking the enumeration as the proxy function, if the enumeration is explicitly at hand."[143] In effect Quine is saying that, no matter what the intended model of a theory, this theory's ontology is mappable, via a proxy function, to the natural numbers (or some subset of them), and hence, that its ontology is ***reducible*** to the natural numbers or one of its subsets.

What effect, if any, does this chain of reasoning have for the theory of structural semantics? Conceive of the "theory" in Quine's argument as equivalent to the network of

---

[142] A set is countable if it is either finite or countably infinite. A set S is countably infinite provided that S and the set of natural numbers, $\mathbb{N}$ have the same cardinality. A set S is uncountable if it is not countable.

[143] Quine (1969), p. 59.

interpreted representations possessed by some entity (e.g., a person) with representational capacity. Suppose this network has only $\aleph_0$ elements (as is plausible for any finite being). This theory (call it Φ) is interpreted in virtue of a homomorphism to a network of objects (and the relations between them) in the actual world—such objects are the intended model of the theory. However, given the DLS together with Quine's point about proxy functions, we know that Φ may also have, as a model, the natural numbers and the relations between them. DLS provides that a model equivalent in cardinality to the natural numbers exists. Quine's proxy function provides a mapping from the elements of this model to the natural numbers. The result is that there are at least[144] two models for Φ, that is, two interpretations of it. Since Φ is just a network of interpreted representations, another way of putting this result is that DLS together with Quine's proxy function implies *persistent* non-unique interpretations of the same representational structure (see **Figure 5-A** below).

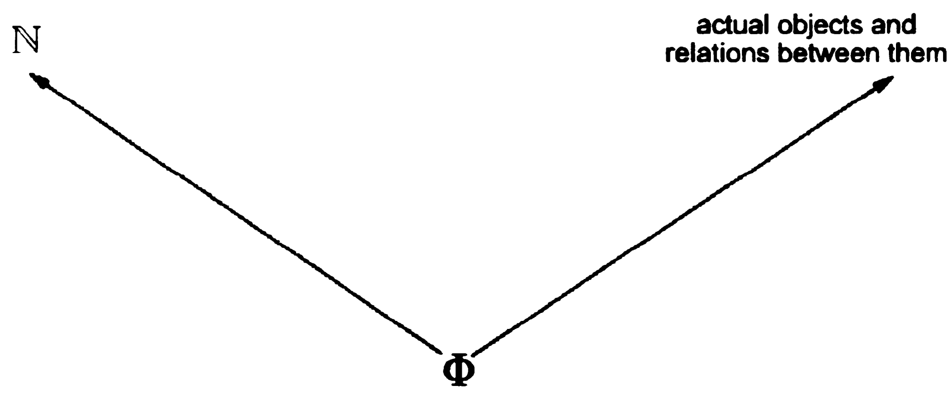

Figure 5-A: The Pythagorean Problem

[144] Actually, there will be an infinite number of competing interpretations, because if a theory can be mapped to the natural numbers, then it can be mapped to the evens, the odds, etc.

But that is not all. The prospect of having more than one possible model for a representational network is not a new one. DLS perhaps allows a more formal illustration of the problem than has been offered until now, but the problem itself is simply what I have been addressing all along. What DLS shows that is new, is that among the competing interpretations of Φ are models which would make the content of representations in Φ nothing more than numbers and relationships between numbers, when what we'd wanted was a representation of the objects and relationships in the actual world with which we are familiar. Quine recognizes this result in "Ontological Relativity" and thinks it is a kind of Pythagoreanism regarding representational content that ought to be avoided, if possible.

Notice that the problem of Pythagoreanism arises regardless of whether or not one embraces an ontology of external structures which are manifest only through behavioral response. Even if it is true that allowing for a richer ontology of structures will eliminate many of the competing translations of a given term that a more spartan ontology would retain, the DLS shows that there will, nonetheless, always be more than one model for any given representational structure. The assumption behind the holism solution is that where two or more models exists for a single representation, it will always be possible to discover finer structural detail in the external world which will provide the justification for considering one competitor a better model than the other. What DLS shows is that this assumption is false: Even if the holism solution may eliminate competing models up to a point, there will always be at least one structure which can model a representation as well as any other, and that is the model provided by the ordered set of natural numbers.

5.2 The Pythagorean Problem

So how can competing mappings be eliminated in spite of the DLS result? The answer is that they cannot be eliminated and happily, do not need to be. But before explaining why this is so, let us take a closer look at how the reduction described in the DLS would proceed.[145]

The downward Lowenheim-Skolem theorem shows that "all but a denumerable part of a [theory's] ontology can be dropped and not be missed."[146] If we begin with a theory whose universe of discourse is non-denumerable, then we can imagine that the reduction proceeds, in theory, as follows:

> [We begin by] partitioning the universe into denumerably many equivalence classes of indiscriminable objects, such that all but one member of each equivalence class can be dropped as superfluous; and one would then guess that where the axiom of choice enters the proof is in picking a survivor from each equivalence class. If this were so, then with the help of Hilbert's selector notation we could indeed express a proxy function.[147]

Figure 5-B represents a procedure for the construction of a proxy function. On the first line we have the set of all real numbers x such that $1 \leq x \leq 2$, $2 \leq x \leq 3$, and so on. This is the non-denumerably infinite set of real numbers between 1 and 2, 2 and 3, etc. If a survivor is picked from each subset, then we are left with the set of natural numbers $\mathbb{N}$.

---

[145] Note that this "proof" is not intended to be a proof of DLS or of the method by which an appropriate proxy function is constructed either. The DLS only shows that a model exists which is equivalent in cardinality to the natural numbers. It does not provide the proxy function which maps the elements of this model to those of the intended model or any other specific model. This illustration is simply one example of how such a proxy function might work.

[146] Quine (1969), p. 60.

[147] Quine (1969), p. 60. Note that the proxy function reduces a homomorphic function between relational systems to that of isomorphism, where the latter is a special case of the former.

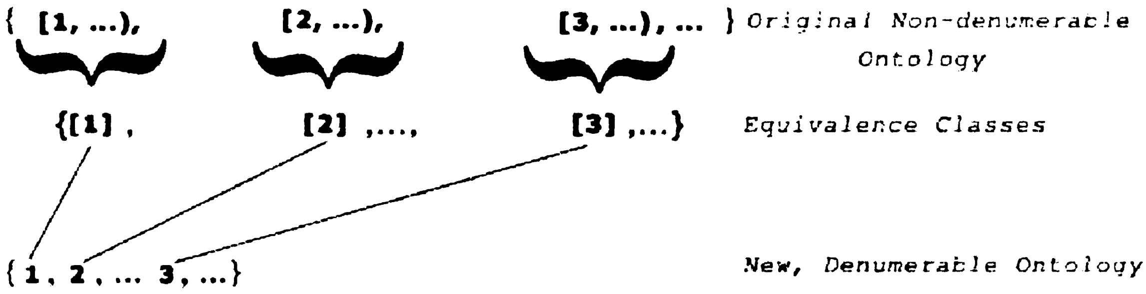

Figure 5-B: Construction of a Proxy Function for DLS

There doesn't seem to be much point in disputing the truth of the Lowenheim-Skolem theorem or of the possibility for constructing proxy functions like the one Quine describes. So our available options seem to be these:

(1) accept the consequences of Pythagoreanism outlined above or,

(2) add an additional constraint to the structural semanticist's account of representation. That is, insist that there is something more to the representation relation besides homomorphism and make sure that the extra condition doesn't allow representations to be mapped to the natural numbers but only to the universe of objects you really want to be your model.

Of these options, (2) seems to compromise the main advantage of structural semantics—namely, the beauty and simplicity of the account. Structural semantics makes no appeal to causality, evolutionary history, or behavior in order to explain representational content as do most all other theories. Indeed, there are also well known difficulties with all of these alternatives which are, in my opinion, much more daunting than those associated with structural semantics. Moreover, to appeal to one of these alternatives now would make the account of mental representation I am defending here *ad hoc*.

This leaves us with option (1), which, on the face of it, sounds unacceptable too. But is it really? What happens if we accept that a homomorphic account of representational content will always leave at least one competitor (and in fact, an infinite number of them) to the mapping between representations and objects in the world, namely, a mapping between representations and the natural numbers or some subset of them? What seems to happen is that objects and numbers become ontologically indistinguishable relative to the theory. For example, it seems that there would be no difference between a representation whose content is actual dogs and a representation say, of the number 4.

Remember that, according to structural semantics, the content of a representation can't be determined by looking at the way it individually maps to an element of the external world, but instead is determined first by considering how the entire representational system maps to the external world, and deriving the contents of individual representations from this. And we have seen that even when the representational and external networks are considered holistically, we are still committed to Pythagoreanism. But it is a very unusual sort of Pythagoreanism to which we're committed if it is a form of Pythagoreanism at all: What constitutes a mapping between two structures is not that they contain certain *kinds* of objects or the same *number* of objects or really anything to do with the objects themselves. Rather, what constitutes a mapping is the relationships that the objects (or numbers, as the case may be) bear to one another (and of course, whether the representations to be interpreted enter into the equivalent[148] relationships). That is, whether or not a network of representations can be mapped to another structure depends on whether or not that whole structure serves as a *model* for the whole representational network.

Now here is the crux of the point: If any model consisting of the natural numbers is a model for a network of representations, then it would have to be the case that the natural numbers were related to one another in this model in precisely the "same way" that the representations in the network are related, otherwise it would not be a model. But this is all that was ever required of a model consisting of objects![149] Hence, though there may be two potential models for the representational network, both "say" the same thing about the content of representations in that network. What provides the representations in the network with content is exactly the same in both models.

Paul Benacerraf makes a similar point in "What Numbers Could Not Be," an article that probes the question of what kinds of things (e.g., sets, objects, relations, etc.)

---

[148] By "equivalent relationships" here, I do not mean to suggest that the relations which order the elements in the two structures mapped to the representational network must be the same relationships. Rather, I mean to say that whatever relations order the two structures, they must be equivalent in respect of what makes them models for the representational network. Hence "Structure A and structure B are equivalent" should be read as shorthand for "A and B are equivalent in respect of which makes them both models for a representation R."

[149] It is possible to argue that this begs the question. After all, couldn't somebody (e.g. Fodor) say that for the content or meaning of language or thought, more is required, namely that it be the right set of actual objects, not just some homomorphic model? Even if what is claimed here is correct when applied to models, when applied to content or meaning in general, does it not beg the question? I guess it does beg the question, strictly speaking. But only if taken in isolation from the other arguments contained in this defense of structural semantics. I see this argument as addressing a problem which is *internal* to structural semantics. For even if one believed the theory wholeheartedly, you would still be confronted with the problem of avoiding Pythagoreanism.

numbers might be.[150] Benacerraf claims that, given a set of sufficient conditions for explaining the concept of number, we can yet arrive at interpretations of the concept of "number" which are contrary to one another. In effect, Benacerraf illustrates a version of the uniqueness problem regarding the content of "number." To take a particular number as an example, suppose that the dispute is over the matter of the proper interpretation of '3.' **Figure 5-C** illustrates some of the possible ways in which '3' might be interpreted:

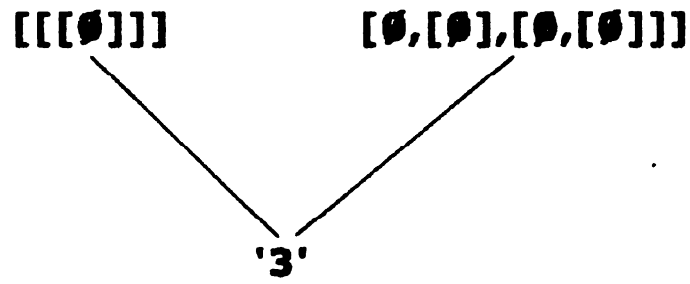

Figure 5-C: Interpretation of Numbers as 'Objects'

Both interpretations suppose that the number '3' should be understood as the successor under some relation *R* of the number 2. The problem is that in one case (right) the successor under *R* of 2 is understood as the set consisting of 2 and all the members of 2, while in the other (left) the successor of 2 is simply [2], the unit set of 2—the set whose only member is 2.[151] Which interpretation is correct?

The curious thing is that there can be such a high degree of concurrence regarding what (relational) properties the numbers have (e.g., that each is the successor under *R* of

---

[150] Benacerraf, Paul. "What Numbers Could Not Be" in The Philosophical Review. Volume 74, Issue 1. January, 1965. pp. 47-73.

[151] Benacerraf (1965), p. 54.

the number 2), and yet so much disagreement over what kinds of objects (e.g., which set) the numbers *are*. And indeed, Benacerraf thinks that the problem arises from the assumption that numbers must be characterized as some kind of set-defined *object* at all. Numbers, on Benacerraf's view, ought to be characterized much as representations are interpreted on the structural semanticist's view: they are individuated exclusively on the basis of their role in an ordered (relational) system.

> *Any* object can *play the role of* 3; that is, any object can be the third element in some progression. What is peculiar to 3 is that it defines that role—not by being a paradigm of any object which plays it, but by representing the relation that any third member of a progression bears to the rest of the progression. ....
>
> So what matters, really, is not any condition on the *objects* (that is, on the set) but rather a condition on the relation under which they form a progression. To put the point differently—and this is the crux of the matter—that any recursive sequence whatever would do suggests that what is important is not the individuality of each element but the structure which they jointly exhibit. This is an extremely striking feature. One would be lead to expect from this fact alone that the question of whether a particular "object"—for example, [[[⌀]]]—would do as a replacement for the number 3 would be pointless in the extreme, as indeed it is. "Objects" do not do the job of numbers singly; the whole system performs the job or nothing does. I therefore argue ... that numbers could not be objects at all; for there is no more reason to identify any individual number with any one particular object than with any other (not already known to be a number).[152]

While Benacerraf reaches the conclusion that numbers could not be objects, I say (above) that numbers and objects are arguably more or less indistinguishable. These may sound like very different points of view but they are really not. Both positions embrace an ontology in which relations are what are real. Benacerraf sees the notion of "object" as one which suggests singularity and isolation and therefore discards the idea that numbers,

[152] Benacerraf (1965), pp. 69 & 70.

qua essentially relational entities, can ever be "objects." But I see "object" as just another way of labeling a cluster of relations. This is why I don't mind when Quine argues, effectively, that one might just as soon interpret some appropriately structured representation as '3' rather than as a dog. Assuming that both really are models for this representation in respect of the relational properties which they instantiate, it doesn't really matter to me what name they are given. And indeed, it seems to me that so long as *the same relations* are being modeled in each case, there is no sensible way in which to understand this as a case of non-uniqueness. Benacerraf writes:

> Why so many interpretations of number theory are possible without any being uniquely singled out becomes obvious: there is no unique set of objects that are the numbers. Number theory is the elaboration of the properties of *all* structures of the order type of the numbers. The number words do not have single referents. Furthermore, the reason identification of numbers with objects works wholesale but fails utterly object by object is the fact that the theory is elaborating an abstract structure and not the properties of independent individuals, any one of which could be characterized without reference to its relations to the rest.[153]

This seems to me to be the right argument but I would put the conclusion in a slightly different way: Indeed, why so many interpretations of a single representation are possible without any being uniquely singled out is that there is no unique object that is the representation's referent. Contrary to Benacerraf, I would argue that representations *do* have single referents nonetheless. Their referents are the *sets of relations* which each competing model instantiates.

Though Quine may be right in demonstrating a persistent variety of non-uniqueness—one which the holism solution is of no use to eliminate—it seems that the

[153] Benacerraf (1965), pp. 70-71.

sort of non-uniqueness we are left with is completely irrelevant to the task of fixing representational content, if it is even fit to be described as a version of "non-uniqueness" at all. Those characteristics which distinguish alternative external structures from one another are not the characteristics in respect of which both constitute models for a particular representation. As a result, the fact that there may be more than one kind of object that can realize a model for a particular representation does not imply that the representation has more than one model.

You may be thinking that an approach such as this resolves the problem DLS posed for *uniqueness* of models in interpretational semantics, but that it still leaves us with something too close to a Pythagorean view. After all, what is a "Pythagorean view of content" if not the view that numbers and objects can be used to "say" the same things about the content of a representation? I would argue, however, that saying this no more makes one a Pythagorean than does accepting the principle of the identity of indiscernibles. For someone who accepts this principle, things do not differ unless their properties differ. Hence, the only real difference between objects and numbers must reside in a difference in their properties, including relations.

To articulate this position in still another, final way: If a set of numbers and a set of objects are both models for a system of representations, then there can be no difference in their properties (or at least, no difference in the properties which are represented by the system). So it seems just as legitimate to call what gives representations their content "objects" on this scenario as it does to call them "numbers." Indeed, it must either work out that in our ordinary language (where it *seems* we can distinguish objects from numbers) we can also express some property of one that is not a property of the other, in which case they could not both be equally good models, or else we are wrong in thinking we can distinguish the two in ordinary language. In the first case, both objects and numbers would not turn out to be equally good models of the representational network we seek to explain, with the result that the uniqueness problem for that network never arises. In the latter case, where both options turn out to model the representational network equally well, the models are unique in respect of that which makes them models and hence do not provide conflicting answers to the question of what content representations in the network may possess.

Of course, insofar as being a Pythagorean means that one believes numbers are what really exist, this is anything but a Pythagorean view. For in this view, numbers are not what really exist and give content to representations, *relations* are.[154] It's just that "numbers" can assume these relations as easily as anything else. If this view is nonetheless a kind of Pythagoreanism in the reader's opinion, then it is a kind which the structural semanticist ought not to regret embracing.[155]

---

[154] This may seem like a curious claim to make given that, in mathematics, relations are just sets of ordered pairs. If we think of relations as understood in mathematics, then it seems like there must exist "elements" or "individuals" which belong to ordered pairs in order to make sense of the concept of relations. However, we can acknowledge that the notion of "relation" requires that there are "elements" which form sets of ordered pairs while still thinking of these elements as relational systems themselves rather than as discrete individuals. (See Chapter 3 for a more in-depth discussion of "levels" of representation.) This way of thinking about relations may ultimately only push back the question of which n-tuples of elements constitute the relations on representational and external relational systems to a lower level of analysis. If so, then it may make sense to ask whether an alternative model of relations might suffice for the structural semanticist's account of representation (cf., Randall Dipert's "The Mathematical Structure of the World: The World As Graph." and graph theory).

[155] This sort of "pythagoreanism" is already finding its way into mainstream thinking. Consider the following case, reported by *The New York Times* in early March of 2001: Professor David Touretzky, a computer scientist at Carnegie-Mellon, sponsors a website containing representation of software written to

5.3 The Role of A Background Theory Revisited

Finally, it would be nice to know what bearing the preceding discussion may have on the earlier question, raised in connection with Quine, of the degree to which a background theory is required in order to make sense of talk of the relations between theories and between theories and their models.

In Chapter 4, we saw that Quine's views concerning translational indeterminacy are compatible with the view that terms can have fixed reference, so long as one understands that the reference of such terms is relative to a framework. In the 'gavagai' case, the "framework" or "background theory," as Quine calls it, is the native language, the terms of which are translated/interpreted as the terms of English. For structural semantics, the background theory consists of a network of mental representations, a cognitive relational structure which is mapped to an external relational structure, thereby interpreting it. According to Quine, it makes sense to speak of the interpretation of a term or representation only relative to this background theory: Outside of any overarching

---

allow users of the Linux operating system to play DVDs on their home computers. In 1999, when the software was initially developed, no software was available for viewing DVDs on Linux machines. Creating the software required developers to crack the encryption scheme protecting the discs. The new software, called DeCSS was published on sites like Professor Touretzky's as part of the open-source software movement, causing it to run afoul of the Digital Millenium Copyright Act, a law passed in 1998 that makes it illegal to offer a way to gain unauthorized access to a copyrighted work protected by encryption. In February of 2001, the Motion Picture Associate of America contacted Touretzky's ISP in attempt to convince them to terminate his website on the grounds that it was illegal. In response, Touretzky invited open source programmers from around the world to participate in a "contest" to develop shorter and shorter versions of the code that would break the MPAA's encryption scheme. He then published the results on his website. One submission came from a programmer and mathematician in Finland, who contributed a submission which hides the DeCSS code in the digits of a very long prime number. The MPAA is contending that this number is illegal under the Digital Millenium Copyright Act. (Source: *New York Times Online*, March 2001, http://www.nytimes.com/2001/03/30/technology/30CYBERLAW.html). It seems clear, therefore, that from the perspective of the MPAA, there is no difference between the code which breaks their encryption scheme and this particular prime number.

theory, the notion of an interpretation is nonsensical, since there is no criterion for assigning one interpretation to a term in the theory over any other.

It certainly seems that at least one result of the preceding discussion is the realization that if we can distinguish (i.e., describe the differences between) two competing structures, C_1 and C_2, both represented by *R*, then we must be doing so in a "background" representational structure which is more "complete" (or richer) than *R*. Were this not the case, then the uniqueness problem would not arise in the first place. For in order to see C_1 and C_2 as competitors, we must have some way of representing their differences. The uniqueness problem arises because *R*, by itself, represents both C_1 and C_2 equally well, but unless our representation scheme is richer than *R*, there would be no way to know that C_1 and C_2 were different objects, and therefore no way to articulate any "problem" about which one of them *R* should be taken to represent.[156]

Having said this, it should be clear that at any point in time, our most complete *R* (our most complete representation of the structure of the external world) has only one model, *C that we can describe*. The progress of our knowledge consists in discovering that our most complete *R* is in fact *not* complete, and in introducing new representations into *R* that bring it back into harmony with the structure of the external world. In this very specific respect, whether or not a unique model is available for any given representation *is* something which must be decided from within the context of a richer background theory. In particular, the background theory fixes our assumptions about what sorts of things

[156] A similar observation arose in Chapter 1 in connection with Cummins' theory of error and the possibility of distinguishing between alternative targets.

(e.g., relations, objects, behavioral dispositions, etc.) are legitimate members of our ontology. From there we can do our best to establish which representations are mappable onto which elements of this ontology. Perhaps another way of putting Quine's point, the point that a background theory is required for the project of fixing content, is to say that a theory of content cannot be defended independently of a set of ontological commitments.

Although Quine's ontological commitments and those implied by structural semantics are very different, it seems that, in the end, his views are not entirely different from my own. Quine would seem to agree that there can be no basis for distinguishing between competing external structures (here analogous to the referents of terms in his discussion) so long as they do not differ in respect of that which makes them models for the representational structure: This is really just another way of putting the idea behind translational indeterminacy. For Quine, what characteristics are allowed to be considered in determining a thing's "structure" are undoubtedly very different, and indeed, more limited, than those characteristics considered by the structural semanticist. However, beyond this difference in point of view, Quine and the structural semanticist have more in common that it originally seemed.

CHAPTER 6: *The Problem of Error and Indirect Representation*

Having attempted to provide an account of how *successful* representation works and to defend this account against some of its more serious obstacles, finally it is necessary to consider how one should explain representational error. Typically, explaining how representations err has been a daunting task for theories of representation. Philosophers often find that having spent considerable time perfecting an account of how representation works when it works, there is no room left for representational failures. But clearly we *do* make mistakes which we describe as "misrepresentations," consequently, a good theory of representation ought to be able to account for this.

In addition to misrepresentation, a related phenomenon requiring explanation is the representation of objects that do not exist in the external world. This too has been a challenge for most theories of representation, but it presents a special challenge for the theory of structural semantics. Given that structural semantics grounds representational content on a mapping relationship between properties in the representing entity and properties of the thing which is represented, it might seem that the theory is hard-pressed to explain how "things" that don't even exist are to be represented. After all, there will presumably be no entities which can stand in the relation that is supposed to characterize representation in this view.

6.1 The Problem of Misrepresentation

Since structural semantics has it that representations have the content that they do in virtue of the existence of a homomorphic mapping between the representation and some relational structure in the external world, there is a real sense in which no representation is ever "inaccurate." Insofar as any external relational structure is a model for a structure in the cognitive system, the structure in the cognitive system can be said to represent that external relational structure. There is really no way out of this for structural semantics given its account of what representation is. So if there is no sense in which a representation *R* fails to be a representation of *C* when *R* is mapped to *C* homomorphically, then how can the concept of misrepresentation even be well defined in a structural semanticist's theory, let alone explained?

The answer is that for structural semantics, misrepresentation cannot be defined simply as a failure to represent. And accordingly, freedom from error is not simply a matter of achieving successful representation. This claim should sound familiar. Cummins argues precisely this in his own discussion of error and interpretational semantics, as we saw in Chapter 1. But if there is something more to error and accuracy than the failure and success of representation respectively, then what is it?

You'll recall that Cummins thinks that the answer to this question can be found in making a distinction between what a representation *ought* to represent, namely, its *target*, and what it *actually does represent*: the *content* of the representation. For Cummins, and for structural semantics, the content of a representation is never in error. This follows from the fact that the content of a representation is, in both views, given exclusively in virtue of the presence of a mapping relation between representation and external object. If such a mapping relation obtains between representation and external object, then the content of the representation is the external object and, in general, there is nothing more to determining the content of representations. There is no sense in which a representation can have content which is "incorrect." Representations either have the content that they do, or they fail to have it. Therefore, rather than equate misrepresentation with a failure to represent, Cummins claims that error arises whenever the content and "target" of a representation are not the same.

For Cummins, the "target" of a representation is something which can only make sense relative to a particular representational system with a particular architecture. According to him, the target of a representation *r* is what a system Σ which tokens *r* "expects" to find when it accesses *r*.[157] Substructures of Σ called "mechanisms" or "intenders" function to produce tokenings of representations, which in turn inherit their functions or "targets" from the function of the mechanism that tokens them. Mechanisms have the functions they have in virtue of having the structure they do, that is, in virtue of "incorporating a design assumption" inherent in the architecture of the system.[158] As a way of illustrating part of what Cummins means when he says that a mechanism "incorporates a design assumption," consider the phenomenon of "reverse engineering" as applied to software programs. Reverse engineering projects rely on the assumption that it is possible

[157] Cummins (1996), p. 18. Recall that Cummins explains the notion of what a system "expects" to find by stating that it is equivalent to what the system, or more specifically, the system's *mechanisms*, *function* to represent.

[158] Cummins (1996), p. 18.

to discover the function or purpose of a piece of software by reasoning "backwards" from the code in which the program is implemented. Other projects involve "decompiling" machine code in order to obtain the original source code in which the program was designed. In the first case, if enough information is known about how the program functions in relationship to its environment (usually a specific operating system) and its input, one can use this, together with what is known about the program's own structure, to determine what the program does. In many cases, there is only one, unique, behavior the program exhibits when executed in a specific environment with specific input. This behavior is arguably the "function" of the program insofar as it produces output or behavior which systematically covaries with specific input, together with the program's own internal architecture. In Chapter 1, we saw that Cummins wants to argue that mechanisms of Σ possess functions in exactly this sense of the term—in the sense that their physical architecture determines their function, and ultimately, the function of the representations which they serve to token.

> Just as programs are designed around assumptions about what will be accessed when a given variable is evaluated, cognitive systems are designed around assumptions about what will be represented by various intenders. Specifying a system's design or functional architecture involves specifying what intenders are possible, and hence, on the current conception, what targets are possible. The targets that are possible for a given system are thus fixed by its functional architecture.[159]

One advantage of Cummins' conception of function is that, unlike adaptational role theory, for example, it does not imply that every type of representation has the same

[159] Cummins (1996), p. 18.

function/target. In his view, mechanisms or "intenders" possess functions implied by the structure or architecture of the system. Representations *inherit* their functions from the mechanisms which token them, so, for example, if a mechanism with function f_1 should produce a representation | elm | at time t_1, and a mechanism with function f_2 should produce | elm | at time t_2, then the result is that | elm | has function f_1 at t_1 and function f_2 at t_2. This is a feature of Cummins' account which is all but essential if proper and improper exploitation of function is going to underwrite an account of representational error in the way he has envisioned. For Cummins, error occurs when there is a mismatch between target and content. The content of a representation is always "accurate" in Cummins' view—that is, a representation always has the content that it does in virtue of an isomorphic mapping between it and the structure it represents. In the absence of such a mapping the representation isn't "incorrect"—it simply doesn't represent at all.[160] This implies that for error to occur (i.e., for there to be a variance between target and content) it is the *target* of the representation which has to change. But since the target of a representation *just is* its function, the *function* of a representation must be allowed to vary from token to token. Cummins therefore needs a notion of "function" which is compatible with this requirement.

Another advantage Cummins seeks for his notion of "function" is that correct functioning will not end up being understood as equivalent to, or as implying, "successful" representation. It is now easy to see why Cummins can say that it is mistake to equate correct representation with freedom from error, or alternatively, to equate error with

---

160 ... and therefore isn't really a representation.

incorrect representation. If Σ tokens | elm | on an occasion when the target of the tokening was beech, then there is error borne of a content-target mismatch, although there is nothing whatsoever the matter with the content of | elm |. It represents elms just as surely as it would have had the target and content of the representation been the same. Likewise, if Σ tokens | elm | on an occasion when the target of the tokening was an elm, it is the consistency between target and content that is responsible for this success and not the content of the representation, which remains constant in both cases.

In Chapter 1, I argued that the biggest challenge to Cummins' strategy for accounting for error lay in his efforts to separate the notion of function from representational success while maintaining that mechanisms' functions can be understood in a way that does not rely on antecedently meaningful concepts involving the goals or intentions of a system or its designer. There, I distinguished between two senses of the term "function." One of them, which I termed $function_1$, is the sense of function one employs when one intends to talk about the behavior a system exhibits in virtue of facts about its physical architecture. In contrast, $function_2$ is the term reserved for talk of behavior consistent with the maximization of a system's purposes or goals. For a given system, its $functions_1$ and $functions_2$ may be, but are not necessarily, the same.

I want to argue that, in order to distinguish successfully the content of a representation from its target, Cummins needs a conception of function which is closer to that of $function_2$. However, I believe Cummins does not (and possibly cannot) provide an account of $function_2$ while still maintaining a strict distinction between function and representational success.

In Chapter 1, I argued that a notion of function which is close to function$_1$ cannot do the work of distinguishing target and content. In that chapter, I described a grading algorithm which, while designed to compute the average of a set of five grades and set the data object *Average* equal to this value, nonetheless prompted the user for six grades and set the value of *Average* to a lower quantity than was expected. In Cummins' terms, though the *target* of *Average* was the average of a set of five grades, its *content* was the average of six.

Putting aside what we know about the grading algorithm from its use in Chapter 1, suppose that this piece of code was offered to us for the purpose of reverse engineering it. If we want to know the target of the data object *Average*, then we need to know what its function is. In the absence of any access to the program's designer, our best way (and arguably our only way) of determining this is to examine the architecture of the algorithm. We examine the algorithm and find that its structure, combined with the input we provide to it on request functions$_1$ to calculate the *Sum* divided by 6 and to assign this value to *Average*. This provides us with the content of *Average* but, by hypothesis, not with its target. In fact, it is hard to see how examination of the architecture or function$_1$ of the grading algorithm will *ever* reveal what the true target of *Average* had been. The functions$_1$ of systems seem good only for revealing the *content* of the representations they employ. This leaves function$_2$ as the only alternative for understanding target in terms of function. But I believe that Cummins cannot describe a sense of function$_2$ that does not involve an appeal to representational success, contrary to what he claims he must do.

As a matter of fact, Cummins does acknowledge that correct functioning might be understood in terms of the types of performances that promote the *system's* survival and fitness.[161] A system's mechanisms might arguably have the function of producing representations of edges, given that edge-detection may be an important skill to possess for the purpose of avoiding falls and collisions, for example. Representations tokened by these mechanisms would then have edges as their targets, but might have anything whatsoever as their actual contents. That is to say, in Cummins' view, it should be possible to identify a representation as having a specific target, even if no representation has ever been tokened whose content matches this target!

> ...selectionist theories assume that something must actually perform its function successfully in order to contribute to its replication. Whatever you may think of this assumption in other contexts, it won't do in the current one, for it amounts to the assumption that something is a *t*-intender only if it sometimes successfully represents *t*, and that is just the assumption we must avoid for the very good reason that it is not true.[162]

If this is the case, then performative success, understood by Cummins as the convergence of target and content, might never be observed in the system. But if observance of successful performance is a prerequisite to the identification of targets/functions (as I have argued it is), then we cannot use the concept of a "target" to evaluate successful performance.[163] To do so is to be involved in a kind of vicious

---

[161] Cummins, 1996, p. 115.

[162] Cummins, 1996, p. 115.

[163] In addition, so long as targets/functions are considered to be those things which would facilitate the "right" performance of the system Σ, it would seem that for Σ to be in error, it must be the case that the content of one of Σ's representations *r*, is not consistent with that which facilitates right performance. This would seem to imply that Cummins cannot allow for errors which actually promote right performance. This is a consequence which he clearly wants to avoid, since he believes that there are

circularity.

This leaves us to wonder if there might be an alternative way of identifying targets, namely, one which does not tie the notion of a target so closely to the system's correct performance, or one which maintains that connection, but can evaluate what it means for a system to perform correctly without appealing to the notion of a target antecedently. I believe there is an alternative, but accepting it comes at a price.

If "misrepresentation" is not going to be understood as incorrect representation (i.e., as having a representation, the content of which is somehow "wrong"), then I believe that the only alternative is to understand it as a species of performance error. In addition, conceptualizing performance error may indeed require the identification of a target, application, or "planned use" for representations, which is violated whenever misrepresentation occurs, as Cummins has argued. But contrary to Cummins, I do not believe that it will be possible to identify the targets of representations exclusively through an appeal to a system's physical architecture. In fact, it is possible that the only way of identifying targets is to allow their identification to be informed by considerations of value. The notion that one object is better understood as the target of a representation than another is a judgment which seems to invariably require a normative component, even when the notions of "target" and "function" are explained, for example, in terms of the evolutionary fitness of systems. That a system errs when it represents the world in a way that is inconsistent with its long term survival or that it is successful when it represents the

a number of cases in which error *can* promote better performance in a system (see Cummins, 1996, pp. 27ff, 44-47, 50, 99, 116-118).

world in such a way to promote survival, is a judgment that implicitly places a higher priority on survival than extinction and on fitness above unsound health. To think that these ends should not be valued may sound ridiculous, and indeed, I am not arguing here that they ought not to be. Rather, the point I am making is that there is a non-naturalizable component to the judgment that would classify representations which do not promote these ends as "errors." The most that a naturalized theory of representation can do is explain how representations come to have the content that they do. It cannot explain why some representations are more appropriate in particular circumstances than others, or why some promote successful performance in a system and others don't. These explanations simply fall beyond the scope of a naturalized theory of representational content.

Once the intended application of a representation is well-defined, perhaps, with the help of an appeal to a set of widely accepted value-based assumptions, we *can* avail ourselves of a number of theories concerning how and why representations which do not match the intended target are activated instead of representations which do match the target. Cognitive scientists who believe that the mind is composed of representational networks, or "semantic networks" point to several reasons why one portion of the network may be more strongly activated than another. Among them is the idea that some elements of the semantic network are more strongly associated with each other than others. The strength of the association between two or more elements of a semantic network is referred to as the "weight" of that association.

> Various factors are thought to influence the weights (Smith and Medin 1981): for instance, the probability that the feature is true of an instance of the concept, the degree to which the feature

> uniquely distinguishes the concept from other concepts, and the past usefulness or frequency of the feature in perception and reasoning.[164]

In addition to perceptual influences, there is evidence that some "top-down" processing may be involved in the activation patterns of semantic networks:

> Semantic network models of concepts also predict priming and context effects. Conceptual information processing will be facilitated if a prior context has activated the pathways that are required for the task. Loftus and Cole (1974) gave subjects a superordinate category name and a feature and asked them to produce an instance of the category possessing the feature. For example, when given *vehicle* and *red*, the subject might respond *fire engine*. In some cases the category was presented before the feature, and in other cases the feature was presented before the category. Subjects were able to produce instances faster when the category was presented first; for example, they produced *fire engine* faster in response to *vehicle* followed by *red* than in response to *red* followed by *vehicle*.[165]

If top-down processing really does influence patterns of activation, this would also imply that subjects could be influenced to activate representations which are not the closest available structural matches to the stimulus with which they are presented. If there is a way to conceptualize the stimulus as the target in this example, this would be a case in which the target and content of the subject's representation were inconsistent.

There seems to be good evidence that something like Cummins' account of error is correct. In particular, it seems that there are good arguments for concluding that error is best understood as what happens when there is a mismatch between the target of one's representation and its content. The main weakness in Cummins' account is that it tries to do too much. A naturalized theory of representational content cannot account for essentially normative aspects of mental representation. But nor does it *need* to account for

---

[164] Cognitive Science: An Introduction. 2nd ed. eds. Neil Stillings, et al. London: MIT Press, 1995. p. 90.

[165] Stillings, et al. (1995), p. 91.

all of this. Naturalized theories of content ought to explain representation in such a way that allows for the possibility of error, and this can be a challenging task in and of itself. However, they do not carry the burden of providing a theory of error, insofar as error is not naturalizable.

6.2 The Problem of Representing Non-Existent Objects

Next we turn to the problem of how to account for the representation of objects that do not exist. A lot of theories of representation have problems accounting for how we represent objects that do not exist. For example, a theory of representation that claims representations get their content from the objects that cause or invoke representations in us, will need to make special efforts in order to explain how representations of unicorns are caused—there are no unicorns to cause them. Similarly, since structural semantics has it that the contents of representations are determined by what they are mapped to in the external world, it will have trouble accounting for representations of things which are not a part of that world. Nonetheless, since we clearly do have representations of objects and/or states of affairs that do not exist externally, a good theory of representation ought to be able to account for them.

6.2.1 The Axiom of Referring

Behind most of the problems associated with the representation of nonexistent things is a principle which Avrum Stroll calls the "axiom of referring."[166] In "Proper Names,

[166] Stroll, Avrum. "Proper Names, Names, and Fictive Objects." The Journal of Philosophy. 1998, pp. 522-34.

Names, and Fictive Objects" Stroll characterizes this principle as the view that *reference requires the existence of the referent*. He is not the first to identify the axiom of reference—as a matter of fact, a number of philosophers have accepted it as a starting point for discussions of reference and naming, some expressions of which are given below:

> Whatever is referred to must exist.[167]
>
> ... it would not *in general* be correct to say that a statement was about Mr. X, or the so-and-so, unless there were such a person or thing.[168]
>
> He who does not acknowledge the nominatum cannot ascribe or deny a predicate to it.[169]

Stroll himself ultimately denies that the axiom of reference ought to be accepted, claiming that it is "patently false."

> The irony of the situation is not merely that it is false, but that it is obviously so; for it is a plain fact that we do use language to refer to nonexistent (including fictive) objects by name, and to make true (or sometimes false) statements about such objects. ... I submit that we should abandon the axiom of reference in any of its forms, since it is palpably false.[170]

Regardless of the merits of those who accept (or reject) the axiom of reference, it seems that structural semantics is committed to some version of it. In particular, the positive account of representation advanced by structural semanticists implies that representation is impossible without something which is the object of the representation. This follows from the fact that representation depends on the existence of a relation

---

[167] Searle, John. Speech Acts. New York: Cambridge, 1969. p. 77.

[168] Strawson, P.F. "On Referring," in A. Flew, ed., Essays in Conceptual Analysis. New York: Macmillan, 1960. p. 35.

[169] Frege, Gottlob. "On Sense and Nominatum," in The Philosophy of Language. ed. A.P. Martinich. New York : Oxford University Press, 1996.

[170] Stroll (1998), pp. 531-32.

between mental and external structures. When one half of the relation is missing, representation cannot occur.

The upshot is that however representation of nonexistent objects is explained, it must not imply that the axiom of reference is false. As a result, I prefer to adopt a rather different terminology for this type of representation. Rather than call it "representation of nonexistent things," which seems to build in the rejection of the axiom of reference from the onset, I will call this type of representation, "indirect representation" for reasons which will become clear presently. Next, I will proceed to show how the structural semanticist can construct an account of a class of "indirect" representations which is intuitively co-extensive with those that we normally refer to as "representations of nonexistent things" and how this account is consistent with what the theory has to say about representations in general.

6.2.2 Indirect Representation

There are really two distinct issues that must be addressed in order to give a complete account of indirect representation. First, structural semantics must be able to explain how indirect representation is possible. That is, we require an explanation of what it is that forms the other half of the representation relation in cases where, intuitively, it seems that nothing is available. As we shall see, there are at least three types of indirect representation which will need to be described in order to fulfill the first of these requirements. Second, it will be necessary to account for how it is that some indirect representations are recognized by the subject doing the representing for what they are—namely, as being about seemingly nonexistent objects—whereas others are not. This is necessary because it is arguable that the content of a subject's representation would differ in each case even though the object which the representation involves remains the same. Consider, for example, Virginia's representation of Santa Claus: Speaking intuitively, it seems that there must be a difference between the content of her representation and the content of mine, who doubts his existence. If, in fact, the contents of two such representations would be distinct, then we must explain how the theory of structural semantics differentiates them.

6.2.2.1 The Classical Empiricist View

Before considering each type of indirect representation in detail, it is useful to note some similarities between structural semantics and classical empiricism on the matter of indirect representation. According to the classical empiricist view, every simple idea formed by the mind corresponds to something in the external world. No idea can be formed which is not either the result of direct experience or the mind's operations on direct experience. Concepts such as that of a unicorn, a Pegasus, or a "virtuous horse" are formed by combining simple ideas got directly from contact with the external world as is expressed in this passage from Hume's An Enquiry Concerning Human Understanding.

> Nothing, at first view, may seem more unbounded than the thought of man, which not only escapes all human power and authority, but is not even restrained within the limits of nature and reality. To form monsters, and join incongruous shapes and appearances, costs the imagination no more trouble than to conceive the most natural and familiar objects. And while the body is confined to one planet, along which it creeps with pain and difficulty; the thought can in an instant transport us into the most distant regions of the universe; or even beyond the universe, into the unbounded chaos, where nature is supposed to lie in total confusion. What

never was seen, or heard of, may yet be conceived; nor is any thing beyond the power of thought, except what implies an absolute contradiction.

But though our thought seems to possess this unbounded liberty, we shall find, upon a nearer examination, that it is really confined within very narrow limits, and that all this creative power of the mind amounts to no more than the faculty of compounding, transposing, augmenting, or diminishing the materials afforded us by the senses and experience. When we think of a golden mountain, we only join two consistent ideas, *gold* and *mountain*, with which we were formerly acquainted. A virtuous horse we can conceive; because, from our own feeling, we can conceive virtue; and this we may unite to the figure and shape of a horse, which is an animal familiar to us. In short, all the materials of thinking are derived either from out outward or inward sentiment: the mixture and composition of these belongs alone to the mind and will.[171]

Similarly Locke held that nothing contributes to knowledge save experience and reflections on experience:

Let us then suppose the Mind to be, as we say, white Paper, void of all Characters, without any *Ideas*; How comes it to be furnished? Whence comes it by that vast store, which the busy and boundless Fancy of Man has painted on it, with an almost endless variety? Whence has it all the materials of Reason and Knowledge? To this I answer, in one word, From *Experience*: In that, all our Knowledge is founded; and from that it ultimately derives it self. Our Observation employ'd either about *external, sensible Objects; or about the internal Operations of our Minds, perceived and reflected on by ourselves, is that, which supplies our Understandings with all the materials of thinking.* These two are the Fountains of Knowledge, from whence all the *Ideas* we have, or can naturally have, do spring.[172]

If we apply the empiricist approach to the problem of explaining indirect representation for structural semantics, we find that even though representations have content in virtue of sharing structure with what is represented, what is represented need not always be a single, unified object.[173] Rather, the thing represented may simply be a set

---

[171] Hume, David. An Enquiry Concerning Human Understanding. Section II: Of the Origin of Ideas, § 4 & 5. ed. Antony Flew. Chicago: Open Court Press, 2000. pp. 64 & 65.

[172] Locke, John. An Essay Concerning Human Understanding. Book II, Chapter I: "Of Ideas in general, and their Original." Oxford: Clarendon Press, 1975. lines 15–26, p. 104.

[173] Note that structural semantics *differs* from the classical empiricist view on the matter of whether all representations must come from the objects of experience. More will be said about this later, but for now, note that the similarity between these views lies chiefly in the tendency of both to allow for

of properties which are instantiated by a number of different external objects. The objects through which these properties are instantiated need not bear the relations to one another that would make them form a more complex "object" in the typical sense of the term.[174] So, a representation of a unicorn could be the result of mapping some mental structure onto selective properties of horses and horns, for example. And in general, when a thing exists, one would argue there is a co-instantiation of its relevant properties (suppose they are p_1, p_2, and p_3), and no co-instantiation of these properties when a thing doesn't exist.

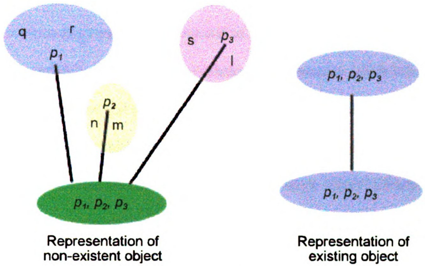

Figure 6-A: Representation of Non-Existent Objects

Happily, neither scenario prohibits the presence of homomorphic mappings between mental structures and $\{p_1, p_2, p_3\}$. For such mappings are possible notwithstanding the

---

representations to possess structure which maps to more than one external object.

[174] Perhaps the relations instantiated properties must bear to one another in order to form "objects" are relations such as "share the same proximal space," "exist in the same time," etc. I leave this question to the metaphysicians, and shall speak here simply of "co-instantiated" properties, leaving this notion undefined.

possibility that p_1, p_2, and p_3 are not instantiated in the same object; that is to say: Such mappings are possible notwithstanding the possibility that the mapping between representational and external relational systems is not one-to-one. See **Figure 6-A**.

At first, this way of explaining representation of non-existent objects may seem like a weakening of the structural semanticist's account of the representation relation. After all, representations are supposed to share a structure with the things they represent. In fact, that shared structure is at the crux of what makes mental structures "representations" at all. If I represent a unicorn, then presumably I possess a mental structure which contains elements corresponding roughly to a horn and to a horse, among other things. Moreover, these elements certainly do stand in some relation to one another which indicates that the horn belongs *on* the horse and not on some other object I might also be capable of representing. How can I explain the representational content of this mental structure by pointing to properties that are not co-instantiated by a single object? Clearly this is a case, it will be said, where some of the structure had by mental representation is missing in the elements of the external world to which it is mapped.

It is true that some of the representations we have would seem to work better if mapped to an object which possessed most of the structure present in the representation, but it is important to realize that the structural semanticist's account of representation is one grounded in a ***homomorphic*** relation, not an isomorphic one. Since homomorphism also drops the onto requirement, it is not necessary that every element of an external relational structure be mapped into by its representation. The fact that some aspects of a mental structure are not replicated exactly in the external world does not prohibit one from nonetheless identifying the mental structure as representational.

This may nonetheless seem like a hollow technical point. Surely most of what we do successfully represent, we represent in virtue of shared structural similarities that approach isomorphism. So even if the structural semanticist can provide a theoretical defense of less "strict" forms of representation, we may feel that in many cases we are owed an account of representation of non-existent objects where there is more shared structure than the theory minimally requires.[175] I believe structural semantics can provide this more robust account. Next, let us examine in detail some different types of indirect representation in order to see how.

6.2.2.2 Representation of Fictions

In cases such as the representation of unicorns, it is arguable that our representations do closely mirror the structure of certain objects in the external world. Pictures, sculptures, and other renditions of unicorns are likely to instantiate many of the properties related to one another in our representation. Of course, someone had to have a representation of a unicorn without the presence of these artifacts, else they would not exist themselves. But this does not in any way obfuscate the claim that the content of our representations of unicorns is usually a function of the relation between them and any

---

[175] Note also that until now, the failure to achieve a 1-1 mapping between representation and represented object has always stemmed from the fact that there have been fewer elements in the representing structure than in the represented one. In this case however, the failure stems from the opposite circumstance.

number of drawings, paintings and statues of the creatures. Remember, structural semantics is not a theory about *why* we have representations or about what caused the first mental representation of some object or other. Rather, it is a view about how representations have the content that they do. And this question is answered equally well by structural semantics whether the representation in question is caused by a real unicorn, or by a picture of one.

I will call representations which map to artifacts such as stories, pictures, myths, and works of fiction in general, *representations of fictions*. There is a significant body of work available which deals with representations of this sort. I'll draw on some of that work in what follows in order to illustrate how fictional representation can be transparent to the subject.

6.2.2.3 Representation of Non-Physical Objects, Contradictory Objects, Hallucinations and Ideas

Not all representations which seem to involve nonexistent objects can be explained by pointing to an image or other representation of the object concerned. For example, there are no pictures of the present king of France, nor of MacBeth's dagger.[176] Similarly, it is certainly possible for us to have ideas which have not been committed to record, or to develop mathematical structures which currently *cannot* be.[177] We have also to consider

---

[176] Although there will presumably often be linguistic representations of these objects: for example, 'the present king of France.'

[177] I am thinking here of a 1,000,000,000-sided polygon or Gabriel's horn—the parabolic conical object with infinite surface area but finite volume.

representations of abstractions such as the general preferences and feelings of a culture or of predicted and hoped for events. What will structural semantics have to say about these?

Some of the these examples can be handled by the classical empiricist view, if they do not qualify as representations of fictions. For example, though there may in fact be no present king of France, it is possible for us to form the representation just in case its constituent parts may be found externally. Similarly, if I hallucinate a dagger, it arguable that a representation is activated in me which maps quite naturally to external objects (namely, real daggers). What makes it a hallucination is that presumably there is nothing which is responsible for the *activation* of this representation. But as we have seen already, structural semantics is a theory about the *content* of representations, not a theory about the mechanism through which representations are activated.[178]

What about contradictory objects, such as Gabriel's Horn? We can represent Gabriel's Horn as a parabolic conical structure which has an infinite surface area but a finite volume. If one thinks of Gabriel's Horn as an enormous paint pot, the idea is that the Horn cannot contain enough paint to paint its own surface, no matter how thin a coat of paint is used (See **Figure 6-B**). Surely such an object cannot exist in reality. But we nonetheless seem to possess a representation of it.

---

[178] Hallucinations of things never experienced by the subject, or that cannot be broken down into constituent parts which have themselves been represented by the subject may present a problem for structural semantics. But it is not obvious that such hallucinations are possible.

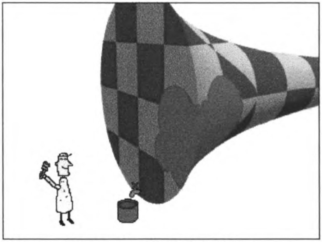

Figure 6-B: Gabriel's Horn

I would argue that what this case shows is *not* that we have representations of things which do not exist in reality and therefore which defy structural semantics' account of representation, but rather, that the "external" structures to which our representations map, do not always have to be *physical* structures. Gabriel's Horn is paradoxical only as a physical object—not as a mathematical object. There is nothing about the formal description of this solid which in any way implies a mathematical inconsistency. As a result, what we often represent when we represent Gabriel's Horn is not a physical object, but a mathematical one.[179]

Admittedly, this may seem to require a commitment to the view that mathematical structures "exist" externally to the mind—if not, then despite all, we still have a case in

[179] It is possible to represent Gabriel's Horn imperfectly—as in **Figure 6-B**, which is a picture suggestive of the actual structure of the Horn, but unreflective of all of its actual properties. This is not really an example of *mental* representation though, as the picture is what is doing the representing here, not the mind. Having said that, it is noteworthy that the representational capacity of **Figure 6-B** is consistent with the structural semanticist's theory of representational content. For although the picture is not isomorphic to the mathematical structure, it does possess properties which can be mapped *homomorphically* to it.

which the mind represents something which does not seem to exist in reality. While I don't want to deny that mathematical structures, ideas, or other sorts of abstraction might exist in reality, neither do I want to commit myself to this view here. But there is really no need to do so. For there is nothing in the structural semanticist's account that prevents structures formed in the mind from being the objects of representations themselves—nothing that equates existence in reality with existence outside of the mind. Presumably not all mental structures will be representational structures. Why not hold that these can be the objects of representation? For that matter, it seems that there is nothing in the structural semanticist's account which would prevent other representational structures from being the objects of representation. Indeed, something like this would most likely form the basis of a structural semanticist's account of introspection.

6.2.3 Meta-Mental Operators

In summary, what is usually referred to as representation of nonexistent objects is really a kind of *indirect* representation of existing objects. Some representations are really representations of pictures, myths or stories which themselves purport to describe something "out there." Others are composed of mental substructures, the elements of which individually pick out external objects, but which are collectively associated in the mind alone. Still others may be representations of other mental structures.

It is an advantage of structural semantics that it can explain different types of representation (e.g., both ordinary representation and indirect representation) using essentially the same explanatory model, but it ought not to leave us with no way of distinguishing between them. It seems like common sense to assume that I can sometimes represent the difference between an imaginary dagger and an actual one, and if so, there must be something in the account of representation to explain what distinguishes the two cases. However, considering only what has been said so far *there would be fundamentally no difference in content between "direct" representations and representations of seemingly nonexistent objects ("indirect" representations)*, according to structural semantics. To illustrate, consider the following two cases:

> **Case I:** *P* has a representation *R* of a non-existent thing and knows that what he/she represents doesn't exist.
>
> **Case II:** *P* has a representation *R* of a non-existent thing and doesn't know that what he/she represents doesn't exist.

Note that for Case II, *P* is not aware, and therefore does not represent that *R* is about something with properties which are not co-instantiated, which are instantiated in a picture, myth or story, or which are exemplified in some other mental structure. In other words, *P* does not know, and therefore does not represent that *R* is an indirect representation. This makes Case II the easier of the two to handle. For even though there surely is a difference between direct representation and indirect representation, in this case *P* does not know, and therefore does not *represent* the difference between the two. A theory of representation does not have to explain a difference where no difference (in representation) exists.

Case I, however, is more challenging. In this case, *P* knows that what he/she represents is represented indirectly. Since all knowledge involves representation, *P* must therefore *represent that R is an indirect representation.* This aspect of *P*'s representation is not explained by the classic empiricist view concerning how the mind forms complex ideas for example. For that view explains only how representations of properties which are not in fact co-instantiated might come about, not how subjects might also represent the fact *that these properties are not co-instantiated.* Similarly, we have not said how a subject can represent *that* his/her representation of Pegasus is really relative to the creature's mythological description or how he/she can know that a representation is about another of his/her own mental structures. Representing these facts seems to require another level of discourse—a way of talking *about* representations and the relationships between them. Representation such as this seems to suggest the need for a *metalanguage.*

Rod Bertolet suggests that when we knowingly represent seemingly nonexistent, predicted, or fictional objects, people, and events, we do so with the use of a kind of meta-mental operator.[180] For example, consider the following statements which seem to involve reference to nonexistent entities:

(1) Pegasus is a winged horse.
(2) Santa Claus has a white beard.

According to Bertolet, there are circumstances in which these statements are semantically equivalent to the following:

(1') [According to Greek mythology] Pegasus was a winged horse.
(2') [According to legend] Santa Claus has a white beard.

---

[180] Bertolet, Rod. "Reference, Fiction, and Fictions." Synthese. Vol. 60, 1984. pp. 413-37.

Specifically, it is when the subject *knows* that his or her reference to Pegasus or Santa Claus is indirect that the claims in (1') and (2') are semantically equivalent to those in (1) and (2). When the subject does not realize that this is the case, the claims are not necessarily equivalent. I say, "not necessarily" here because a subject who is not aware that (1) is false, may nonetheless believe (1'). That is, a subject may believe that according to myth, Pegasus was a winged horse, but may also believe that it is true *in reality* that Pegasus was a winged horse. The bracketed portion of the statement is the meta-mental operator. The job of the meta-mental operator in this case is to relativize the statements in (1) and (2) to a certain context. The presence or absence of a meta-mental operator is the difference between representations which are recognized by the subject as indirect, and those which are not.

As an additional example, consider the following statement:

(3) The earthquake will be here in three hours.

This statement can be translated into:

(3') [According to the prediction] the earthquake will be here in three hours.

In *this* case, it is arguable that the meta-mental operator is doing more than it did in the previous example. Here, the operator, when present, seems to relativize the truth of the statement to a prediction. When absent, the sentence implies a belief that there *will* be an earthquake, rather than merely the belief that there is a standing prediction. In this case, the meta-mental operator alters the semantic content of the statement in a way not seen in the first example.

> But then, the objection continues, suppose that no earthquake comes: won't (3), and what the speaker says by uttering it, be *false*? And won't it be false *because* no earthquake comes? And doesn't this show that with predictions, unlike perhaps the Pegasus myth, we are dealing not just with what the prediction *is*, but with its *accuracy* as well—whether the world is, or comes to be, as it claims? Doesn't the parallel then break down, since whether what one says by uttering 'Pegasus is a winged horse' is true depends only on what the story says, whereas whether what one says by uttering 'The earthquake will be here in three hours' is true is not settled by the content of the prediction alone?[181]

The problem seems to be that (3) carries with it a certain *endorsement* of the prediction which (3') does not. But if this is correct, then (3') and (3) are semantically equivalent only in cases where the speaker *suspends belief* concerning whether or not the prediction is correct. Hence, we ought to apply the meta-mental operator only in such cases. Where the subject transparently believes the prediction to be true, the meta-mental operator is not applied. Once again, this suggests that the role of the meta-mental operator is to distinguish those cases in which the subject knows that his/her representations are indirect from those in which the subject has no such knowledge.

Meta-mental operators should be thought of as functions which operate on the representational structures to which they are applied. In general, they will take representations as arguments, and return representations. Meta-mental operators which take a representation ***r*** and return a result which describes ***r*** as indirect are the kinds of operators which pronounce negatively or neutrally on their input. They are therefore the kinds of operators which are needed for describing Case I above. Case II can be thought

[181] *ibid.*, p. 428.

of as the "default" kind of case and can be described with operators of either of the two remaining types—those which pronounce positively on their input, and those which are neutral (see **Figure 6-C**).

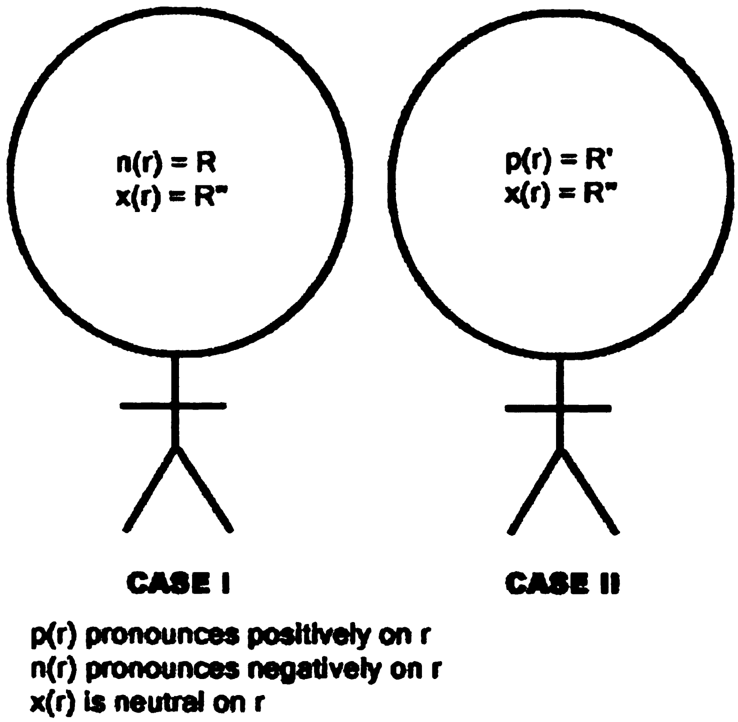

Figure 6-C: Meta-Mental Operators As Functions

Although the notion of a meta-mental operator does introduce another level of representational complexity to the structural semanticist's account, it fits in nicely with the overall strategy of the view. Using the basic idea that representation is a matter of homomorphic mappings between structures as a starting point, structural semantics, with the use of meta-mental operators, can account for the mechanics of indirect representation while explaining how it is different from representation in the standard case.

BIBLIOGRAPHY

Antony, Louise. "Meaning and Mental Representation." (Review) in Mind. October, 1998.

Aune, Bruce. "Quine on Translation and Reference." Philosophical Studies Vol. 24. April 1975. pp. 221-236.

Benacerraf, Paul. "What Numbers Could Not Be." The Philosophical Review, Vol. 74, Issue 1, January 1965. pp. 47-73.

Bertolet, Rod. "Reference, Fiction, and Fictions." Synthese. Vol. 60, 1984. pp. 413-37.

Boolos, George and Richard C. Jeffrey. Computability and Logic. 3rd ed. New York: Cambridge University Press, 1989.

Block, Ned, ed. Readings in the Philosophy of Psychology. Vol. 2. Cambridge: Harvard University Press, 1981.

Bruner, J.S. "On Perceptual Readiness." Psychological Review. Vol. 64, 1957. pp.123-152.

Bruner, J.S. , J.J. Goodnow, and G.A. Austin. A Study of Thinking. New York: Paperback Wiley Science Editions, 1962.

Bruner, J.S., R.R. Olver, and P.M. Greenfield, Studies in Cognitive Growth. Wiley, New York:1966.

Chisholm, Roderick. "The Problem of the Speckled Hen." Mind. Vol. 51, Issue 204. October 1942.

Churchland, Paul M., Matter and Consciousness. revised ed. Cambridge: The MIT Press, 1993.

Copleston, Frederick S.J. A History of Philosophy Vol. I, Greece and Rome. New York: Doubleday, 1993.

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

Cummins, Robert. "Conceptual Role Semantics and The Explanatory Role of Content." Philosophical Studies. 65, 1992.

Cummins, Robert. Representations, Targets and Attitudes. Cambridge: The MIT Press, 1996.

Cummins, Robert. "Reply to Millikan." Philosophy and Phenomenological Research. Vol. LX, No. 1. January 2000.

Dennett, Daniel. "The Nature of Images and the Introspective Trap." Readings in Philosophy of Psychology. Vol. II. ed. Ned Block. Cambridge: Harvard University Press, 1981.

Dipert, Randall. "The Mathematical Structure of the World: The World As Graph." The Journal of Philosophy Vol. XCIV, No. 7, July 1997.

Enderton, Herbert B. A Mathematical Introduction to Logic. New York: Academic Press, 1972.

Fletcher, Peter and C. Wayne Patty. Foundations of Higher Mathematics. 2nd ed. Boston: PWS-Kent Publishing Co., 1992.

Fodor, Jerry. "Imagistic Representation." Readings in Philosophy of Psychology. Vol. II. ed. Ned Block. Cambridge: Harvard University Press, 1981.

Fodor, Jerry. Psychosemantics. Cambridge: The MIT Press, 1987.

Fodor, Jerry. A Theory of Content and Other Essays. Cambridge: The MIT Press, 1994.

Fraleigh, John B. A First Course in Abstract Algebra. 5th ed. Reading: Addison-Wesley Publishing Company, 1992.

Frege, Gottlob. "On Sense and Nominatum." The Philosophy of Language. ed. A.P. Martinich. New York : Oxford University Press, 1996.

Gallagher, David F. "Movie Industry Frowns on Professor's Software Gallery." The New York Times Online. March 30 2001. url: http://www.nytimes.com/2001/03/30/technology/30CYBERLAW.html

Gibson, Roger. "Translation, Physics, and Facts of the Matter." The Philosophy of W.V. Quine. eds. L.E. Hahn and P.A. Schilpp. LaSalle: Open Court, 1986.

Hempel, Carl. "Geometry and Empirical Science." American Mathematical Monthly 52, 1945.

Hume, David. An Enquiry Concerning Human Understanding. Section II: Of the Origin of Ideas, § 4 & 5. ed. Antony Flew. Chicago: Open Court Press, 2000.

Hungerford, Thomas. Graduate Texts in Mathematics: Algebra. New York: Springer-Verlag, 1974.

Jacobson, Nathan. Basic Algebra I. 2nd ed. New York: W.H. Freeman and Company, 1985.

Jacquette, Dale. "Intentional Semantics and the Logic of Fiction." British Journal of Aesthetics. Vol. 29, No. 2. Spring 1989.

Jastrow, Joseph. 1900 "Duck-Rabbit Drawing." Henry Gleitman, Psychology New York: W. W. Norton and Co., 1981.

Kim, Jaegwon. Philosophy of Mind. Boulder: Westview Press Inc., 1996.

Locke, John. An Essay Concerning Human Understanding. Book II, Chapter I: "Of Ideas in General, and their Original." Oxford: Clarendon Press, 1975.

Logue, James. "A Wittgenstein Reader."The Oxford Companion to Philosophy. ed. Ted Honderich. New York: Oxford University Press, 1995.

Millikan, Ruth. "Biosemantics." Journal of Philosophy, Vol. 86, No. 6, 1989.

Millikan, Ruth. "Representations, Targets, and Attitudes" (Review) Philosophy and Phenomenological Research. Vol. LX, No. 1, January 2000.

Quine, W.V.O. "Two Dogmas of Empiricism." 1953. The Theory of Knowledge: Classic and Contemporary Readings. ed. Louis P. Pojman. Belmont: Wadsworth Pub.Co., 1993.

Quine, W.V.O. Word and Object. Cambridge: MIT Press, 1960.

Quine, W.V.O. Ontological relativity, and other essays. New York: Columbia University Press, 1969.

Quine, W.V.O. The Pursuit of Truth. Cambridge: Harvard University Press, 1990.

Rich, Elaine & Kevin Knight. Artificial Intelligence. 2nd ed. New York: McGraw Hill, Inc., 1991.

Rosenthal, David, ed. The Nature of Mind. New York: Oxford University Press, 1991.

Searle, John. Speech Acts. New York: Cambridge, 1969.

Simmons, Alison. "Changing the Cartesian Mind." The Philosophical Review. January 2001.

Strawson, P.F. "On Referring." in A. Flew, ed., Essays in Conceptual Analysis. New York: Macmillan, 1960.

Stich, Stephen and Ted A. Warfield, eds. Mental Representation: A Reader. Cambridge: Blackwell Publishers, Ltd., 1994.

Stillings, Neil et. al., eds. Cognitive Science: An Introduction. 2nd ed. London: MIT Press, 1995.

Stroll, Avrum. "Proper Names, Names, and Fictive Objects." The Journal of Philosophy. 1998, pp. 522-34.

Suppes, Patrick and J. L. Zinnes, "Basic Measurement Theory." R.D.Luce, R.R. Bush, and E. Galanter, eds. Handbook of Mathematical Psychology. Vol. I, New York: Wiley, 1963. pp. 1-76.

Tarski, A. Contributions to the Theory of Models. I, II. Indagationes Mathematicae. 1954.

Walton, Kendall. "Fiction, Fiction-Making, and Styles of Fictionality." Philosophy and Literature.

Wilson, Rob. "Representations, Targets, and Attitudes." (Review) The Philosophical Review. Vol. 107, No. 1. January 1998.

Wittgenstein, Ludwig. Tractatus logico-philosophicus. eds. B. F. McGuinness, T. Nyberg and G. H. von Wright, trans. D. F. Pears and B. F. McGuinness. Ithaca: Cornell University Press, 1971.

Wittgenstein, Ludwig. Philosophical Investigations. 3rd ed. trans. G.E.M. Anscombe. New York: MacMillan Publishing Co., Inc., 1958.