The Measure Of Meaning

By

Simon Carl Pollon

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2007

Author's Declaration for Electronic Submission of a Thesis

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.
I understand that this thesis may be made electronically available to the public.

Signature

Abstract

There exists a broad inclination among those who theorize about mental representation to assume that the meanings of linguistic units, like words, are going to be identical to, and work exactly like, mental representations, such as concepts. This has the effect of many theorists applying facts that seem to have been discovered about the meanings of linguistic units to mental representations. This is especially so for causal theories of content, which will be the primary exemplars here. It is the contention of this essay that this approach is mistaken. The influence of thinking about language and mental representation in this way has resulted in the adoption of certain positions by a broad swathe of theorists to the effect that the content of a concept is identical to the property in the world that the concept represents, and that because of this a concept only applies to an object in the world or it does not. The consequences of such commitments are what appear to be insoluble problems that arise when trying to account for, or explain, misrepresentation in cognitive systems. This essay presents the position that in order to actually account for misrepresentation, conceptual content must be understood as being very much like measurements, in that the application of a content to an object in the world is akin to measuring said object, and that conceptual content ought be understood as being graded in the same way that measurements are. On this view, then, concepts are the kinds of things that can be applied more, or less, accurately to particular objects in the world, and so are not identical to whatever it is that they represent.

Acknowledgements

I would like to acknowledge and thank my advisor Chris Eliasmith for his patience, theoretical guidance, and general support. I would also like to thank the readers of this essay, Paul Thagard, and Tim Kenyon, for taking the time to read, and provide feedback upon, this small work. I would further like to acknowledge the University of Waterloo for its generous financial support.

For Liese.

Table of Contents

Chapter One:

Mental Representation and Language

1.1 Introduction

Concepts are mental representations. Since they are representations, they have representational content. Representational content needs to be explained. This essay is concerned with the explanation of representational content, and is also intended to be a part of a much larger body of discussion concerned with the naturalization of the content of mental representations. Naturalizing anything requires that whatever it is that is being naturalized be ultimately analyzed, or explained, in terms consistent with the vocabulary of the natural sciences. The naturalization of mental representation, or content, is the project of formulating what the relationship is between a mental state and something else, such that the mental state represents the other thing, i.e. what makes it the case that a mental state has a particular representational content. The fact that this is a naturalistic project means that our explanation of mental representation cannot itself rely upon references to representation, or occult forces, to explain how it is that a mental state has the content that it does. The first of these options would leave us with a circular account, which is a problem, and the second would result in a supernatural account, rather than one that is naturalistic, which is also a problem. Further, we postulate mental representations in order to explain the cognitive mechanisms that in turn explain the wide range of behaviour that complex neurological/cognitive systems exhibit on a regular basis.[1] [2]

[1] See Fodor 1987, 1998; Dretske 1981, 1988; Cummins 1996 for various formulations of this condition.

The contents of concepts are, most generally, understood as being categories/classifications, or properties or predicates that can be applied, or attributed, to existent and non-existent entities (Eliasmith 2005; Murphy 2002; Fodor 1998; Dretske 1981). The application of a concept is, basically, the application, or attribution, of a property to an entity, or understood in a slightly different way: the placing of an entity into a category. Say, if the concept DOG were tokened in response to some object being present in the environment, the property of being a dog, or doghood, is being applied, or attributed, to the object. In other terms: the object in question has been placed in a category of 'dog,' i.e. has been classified as a dog.

The general approach taken in most work on mental representation and conceptual content that tends to understand concepts as being such that their meanings are the same as the words we use to refer to them. So, the content of 'dog' and DOG are the same. And so concepts can be combined according to a syntax to form propositions in the same way that words can be arranged, according to a syntax, to form sentences. And so, the meaning of propositions and sentences will be the same, in just the same way that the meanings of their respective constituent concepts and words have the same meaning.

This tendency to understand conceptual content as being identical to that of the content of words results in the application of whatever semantic facts that are discovered

---

[2] Names of concepts will appear in all capital letters, like DOG is the name of the concept that picks out dogs. The meanings, or contents, of mental states or words, will appear in italics, so the meaning of DOG, *being a dog*, or *doghood*, will appear in italics as they just have. The property of being a dog, or doghood, will appear in regular type. Mentionings and references to tokens of sentences will be designated by single inverted commas, e.g. 'Jesse is sleeping'. Propositions, the contents/meanings of sentences, and complex thoughts, which will be discussed below, will appear as meaning in general will, in italics. So, the proposition that *Jesse is sleeping* would appear as it just has.

about words being applied to conceptual content. I think that this way of approaching the content of concepts is a fundamentally flawed one, because it results in some very serious problems for accounting for the conceptual content.

This beginning-with-the-word way of thinking about conceptual content has resulted in a tendency to think of conceptual content as being such that it is identical to the property in the world that it represents, and to further understand conceptual content as either applying to a particular object, or not (and not both). I will attempt to illustrate that this has likely occurred due to the influence of two widely influential ways of thinking about language: Denotational theories of the meanings of words, and Conventionalist ways of thinking about the meanings of words. I will also try to show that understanding mental representation as being language-like is also, naturally, accompanied by thinking that propositions are like sentences, and that what is taken to be the case for the semantics of propositions will also apply to the semantics of concepts. And a consequence of this is that concepts are understood as being true of, or false of, some thing in the world, which reinforces the notion that conceptual content either applies correctly to an object, or it does not. Establishing the plausibility of the view of the situation I have just outlined will be the project of the remainder of this chapter.

In the second chapter, I will further attempt to show that understanding conceptual content as being such that it is identical to what the concept represents, i.e. the property in the world (the referent), and that conceptual content either applies to an object, or it does not, results in extreme difficulties in explaining misrepresentation in cognitive systems. This is a serious problem, because explaining misrepresentation is going to be required of any successful theory of mental representation, and this is because

misrepresentation is ubiquitous in representational/cognitive systems, and any worthwhile theory of mental representation is going to have to account for the common features of representational/cognitive systems. I will then go on to contend that in order to account for misrepresentation, it will be necessary for us to abandon any commitments to content identity, and an all-or-nothing, or absolute, understanding of the application of concepts to objects in the world. In the stead of these positions we will have to accept a graded notion of content, that the content of a concept can be applied to an object in degrees, such that the application of a concept to an object can be more-or-less accurate, and we will have to accept that the content of a concept has to be distinct, i.e. non-identical, to what it is that the concept represents.

In Chapter Three I will take up demonstrating what I have just said about misrepresentation and graded content is, in fact, the case. To do this, I will need to show just how it is that a graded notion of conceptual, i.e. that conceptual content comes in degrees, can explain misrepresentation, where content identity cannot. This graded notion of content that I intend to employ will be based upon the notion that mental representation has more in common with measurement than with it does with natural languages. This position will be taken because a neurological/cognitive system can easily be seen as a sort of measuring device, and measurement provides us with a very clear model for understanding how it is that conceptual content can be graded. This I will also argue for. It is likely obvious to the reader that a graded notion of content is going to require that we accept that there is some way in which our concepts of the same thing (my DOG, and another person's DOG) are similar, so that it is the case that we can be said to, in some way, have "the same" content. However, this point does not come free.

Because there is a well-known argument that any notion of the similarity of representational content is going to necessarily presuppose, and depend upon, a robust notion of content identity, and therefore content identity is our only viable option for the nature of the relationship between the concepts that belong to distinct cognitive systems that are about the same thing; as well as content identity being the only viable option for the relationship between a conceptual content and the property it represents. I will then make the case that content similarity is itself basic, or that content similarity is at least as basic as content identity. I intend to do this by showing that content identity brings with it some very likely unwanted metaphysical baggage, namely the Platonic Forms (which have their own very well known problems), and further demonstrating that it is not as obvious as it might initially seem that content similarity is dependent upon a robust notion of content identity. And this is because content identity very likely also requires some robust notion of content similarity for its even being reasonably postulated.

An aside: I would like to demonstrate this conclusively for all theories of mental representation that seem to assume that mental representation is like language, but unfortunately this essay does not allow for enough space to do so. So this essay will take as its primary targets causal theories of conceptual content that postulate the identity of a concept and what it represents.

1.2 The Focus on Language in the Study of Mental Representation

Inquiry into the nature of meaning usually begins, as mentioned in the previous section, with language, where "inquiry into the nature of meaning" refers to the study of how it is that some things in the world represent, or "stand for," other things in the world

that they are not, i.e. how is it that some things carry, or possess, representational content about other things that they, themselves, are not? The word and the sentence, as you are no doubt aware, are rather significant elements of human language. For most of us, intuitively, at least, language is our model for meaning. This should not be surprising given its ubiquitous presence in human life. But as a consequence of this, language is what many of us (including many philosophers[3]) tend to think of when we think of things in the world that have, or carry, meaning. And increasingly, language is being conceived of as natural phenomena (Millikan 1984; Chomsky 2006; Fodor 1975, 1998; Devitt and Sterelny 1999), if it is not already universally conceived of in this way. So meaning, according to this general view, is to be understood and explained in naturalistic terms, where by "naturalistic" is meant "consistent with the natural sciences."[4] Thus, meaning is not to be understood as something that is without a physical, and law-like-generalization governed, nature. These concerns have resulted in the large and difficult project of trying to provide a naturalistic account of the meaning of the elements of natural languages.

However, despite this tendency of inquiry into the nature of semantics to begin with, and focus upon, language, such inquiry almost as often relies upon to the capacity of the mind, and its constituent mental states, to represent parts of the world such that these mental states will provide us with the meanings of linguistic units. On such a conception of language the meaning of linguistic units is "derived," or otherwise dependent upon the meaning, or representational content (as I will like to refer to it in the bulk of this essay), of mental states (Grice 1957/2000; Dretske 1981; Harman 1982;

[3] See Loar 1981; Evans 1982; Block 1986; Harman 1982; Lycan 1984; Fodor 1998; Fodor 1987; Dretske 1981; Fodor and Lepore 1999; Lepore 1997.
[4] This is the case with mental representation, which was mentioned above.

Searle 1983; Block 1986; Cummins 1989; Lepore 1997; Fodor 1998; Devitt and Sterelny 1999)[5].

This view of linguistic content is a common theme in philosophical work on language, meaning, and mind, and likely finds it motivation in the idea that linguistic units, like words and sentences, express thoughts, and that it is the primary function of language to do this (Devitt and Sterelny 1999). Language exists primarily for the purpose of expressing thought, and, as such, is engaged by human beings in order to communicate the content of our thoughts to other human beings (Devitt and Sterelny 1999; Fodor 1998; Cummins 1989, 1996; Dretske 1981; Grice 1957/2000). Though, as the reader may be aware, language is not engaged by us solely to express our thoughts to others. We often find ourselves making use of language to express our own thoughts, ideas, etc. to ourselves. When we think without speaking our thoughts out loud, it seems to us as though (feels as though), the phenomenology of such a process is as if, we are speaking to ourselves "inside of our heads." Language and thought *feel* tightly coupled. So, it is, in many ways, obvious to think that we think in a language, that is by employing a language, and that even if it is the case that we are capable of thought before acquiring a language, it certainly seems plausible that whatever medium we think in "naturally" is going to be very much like language, i.e. that language and thought are going to be very much alike in structure and content. A natural end to this line of thinking is the notion that language will map on to thought, mental representation, nearly perfectly (if not perfectly), and so it further follows that the semantic and syntactic discoveries made with

[5] For a dissenting voice see Speaks 2006.

regard to natural languages will apply *de facto* to the semantics and syntax of thought. And vice versa.[6]

Further, this equating of the semantic facts of natural languages and those of mental representation seems to be the general line of thinking that guides philosophical work concentrating upon the semantic content of elements of natural languages and the semantic content of mental representation upon which the content of language is identical to. It is the contention of this essay, that such an approach is fundamentally misguided. It is not obvious, upon further reflection, that the meanings of the components of natural languages, particularly that of words, and that of mental representations, particularly concepts, will be identical, even if it is the case that linguistic meaning is dependent upon mental meaning (which I am inclined to believe). So, our mental states are not words and sentences. And this paper, to repeat a bit, is concerned with arguing that mental representations are not linguistic in nature, and that understanding mental representation as being language like is a mistake. This essay holds this point contrary to the point of view of, for example, Fodor (1998), where it is held that, mental representation is going to be a lot like language, for "how could language express thought if that were not the case?" (25).[7] The consequence of this, as I have hinted above, is that words and sentences can be swapped with concepts and propositions with regard to their representational content, because the meaning of the word is the meaning of the concept, and the meaning

[6] See for examples of philosophers who explicitly hold such a position: Fodor 1975, 1987, 1998; Field 1978; Sterelny 1991; Davidson 2001(a), 2001(b), 2001(c). See also discussion in Dennett 1987.

[7] See also Fodor 1987 pages 135-154 on "Why There Still has to be a Language of Thought, as well as Fodor and Lepore 1999 page 383 for the equation of a theory of meaning for mental representations with the theory of meaning that will hold for language. See also Harman 1982 and Block 1986 for attempts to provide the semantics for mental states and linguistic entities in one fell swoop.

of the sentence is the meaning of the proposition. The word and the sentence each get their content directly from the concept or proposition to which they, respectively, correspond.

1.3 Propositions?

Though this essay's primary focus is upon the nature of concepts, there is an issue having to do with the nature of propositional content, which came up briefly in the previous section, that must be dealt with before we can move on to a detailed discussion of the nature of conceptual content. Particularly, we must first address arguments for the non-gradability of propositional content, that propositions either have a particular content, or they do not, and that because propositions cannot have graded content, no mental representation can possess gradable content, and so concepts can not have gradable/graded content, because they are mental representations. However, I will argue that it is quite likely the case that propositional content can be graded, and that even if it is the case that propositional content is not graded, and, in principle, cannot be, this does not obviously, or necessarily, lead to the conclusion that concepts cannot possess graded content, because it is possible for semantic facts that apply to sentences/propositions that do not apply to words/concepts.

In order to take up the task I have just set out in a serious and rigorous manner, it will be required that we get clearer on what propositions are understood to be, because there is no consensus view of propositions (Dennett 1987, 120). I wish to make clear which view of propositions I will be working with for the purposes of this essay, and

why, as well as how the theory of propositions of choice is relevant to the central topic of this paper: conceptual content.

1.4 What are Propositions?

Propositions and their contents tend to play a large role in theories of mind with regard to being the contents of propositional attitude states. Propositional attitude states are those like Beliefs, Desires, etc., which are understood as "attitudes" taken toward propositions. The theory of the mind that qualifies over such states, as most readers already likely know, is most often identified by the names "Propositional Attitude Psychology," or "Folk Psychology." So, when one has a belief[8], one has the belief that *such and such is the case*, (Example: the belief *that it is raining*) where the proposition is the construct that follows the "that" in the belief attribution sentence. The sentence structure following the 'that' is the expression of the proposition in a particular language, and what is being attributed to the belief state of the mind in question is the content of the proposition (which is what the proposition is). The proposition is its content. There is nothing to the proposition other than its content. Since propositions are the contents of thoughts, and the contents of thoughts can be expressed by sentences, a proposition/thought is quite naturally understood as also being the meaning, or content, of a sentence.[9] However, I would like to stress that it is usually held by those concerned with propositions, that propositions are distinct from the sentences of a natural language that express them. The rationale for this position is that different sentences uttered by different people at the same time, or at different times, still express the same proposition,

[8] I will take beliefs as my standard example of propositional attitude psychology state for the purposes simplifying exposition in this essay.
[9] This was mentioned above.

e.g. if I say, “It is raining,” and you say, “It is raining,” in the same context (time, place, etc.) both of our sentences express the same proposition, that *it is raining*. A further consideration in favour of the distinction between propositions and the sentences that are said to express them is that it seems to be the case that different sentences in different languages can still express the same proposition, and so express the same thought (because the proposition is the content of the thought). For example, “It is raining,” and “Il Pluit” express the same proposition and thought

With this general picture in mind we can now briefly list and describe the three dominant views in the literature with regard to the nature of propositions. One is the idea that propositions are sets of possible worlds. “Two sentences express the same proposition just in case they are true in exactly the same set of possible worlds” (Dennett 1987, 121). So, the content of a proposition is the possibly existing conditions under which the proposition would be true, or, in other words: the truth conditions under which the proposition would be true (Dennett 1987, 120; Stalnaker 1999, 678). An important fact to note about the possible worlds interpretation of propositional content is that the proposition is in no way structured or built out of smaller contentful units (Dennett 1987, 120; Stalnaker 1999, 678). Whereas the remaining two most popular theories of propositional content do postulate smaller units out of which propositions are constructed, though they each postulate very different sorts of structural components.

A second theory of propositions postulates that the contents of propositions are states of affairs, or the arrangements of, and relations among things in the world, where the objects and the relations among them are the components out of which propositions are constructed (see Dennett 1987, 120; Stalnaker 1999, 678). I would like to draw

attention to the fact that the components of propositions under this view are not themselves representations, but are in fact, the actual parts of the world[10], which makes this theory quite different from our third option, which proposes that propositions have a composite nature.

The third theory postulates that propositions are sentential in nature, in that they are constructed from smaller semantic units according to a set of rules, or a syntax, that governs how it is that the smaller semantic units are to be put together and arranged in such a way that the larger sentences have meaning, or representational content, about something. The smaller semantic units out of which propositions are built are most often postulated to be concepts (Dennett 1987, 120; Stalnaker 1999, 678). So, it may be obvious at this point that this is most likely going to be the understanding of propositions that we will be considering here, given that they postulate a need for concepts, which are the focus of this essay.

There are more reasons than mere expedience, however. A further reason for why we will be focusing upon the "sentential" interpretation of propositions, as the content of thoughts (propositional attitude states) like beliefs, desires, etc. is because of arguments provided by theorists we will be considering[11] to the effect that human thought is language-like, because it is systematic and productive to a nearly infinite degree, and is so because it is compositional, that is, composed from smaller semantic units. The argument proceeds as follows. It is generally accepted, because it seems to be fairly obviously the case, that any cognitive system that can entertain the thought that *Bill loves*

[10] Though, it perhaps ought to be noted that the relations between objects, and the objects in the world map onto the arrangements of words in sentences of natural languages, which leads us to our next candidate . . .

[11] E.g. Fodor 1987, 1998; Dretske 1981; Sterelny 1991; Fodor and Lepore 1999.

Mary can also entertain the thought *Mary loves Bill*, because each thought contains the same semantic parts, namely BILL, the relation LOVES, and MARY, arranged in a specific way that either reflects or does not reflect reality—i.e. is true, or false (see Fodor 1998, 94-100, Sterelny 1991, 177-185). This interpretation of propositional content, aside from creating the possibility that propositions exist only in minds as sentences in the head/mind composed of smaller semantic mental states called concepts (Dennett 1987, 130), and allowing for the possibility that propositions are not abstract objects to which minds are somehow related (this relation, of course, needing to be explained), also allows for a relatively straightforward explanation of how it is that minds might instantiate propositional content.[12] And this seems to be the construal of propositions chosen by the bulk of the theorists that we will be considering, and especially that of Jerry Fodor (1987, 1998), who will be serving as our exemplar of the view that I am arguing against, because he is the most explicit with regard to claims regarding thought being language-like, and postulating content identity very clearly.

With regard to this essay, I would like to remain as neutral as possible regarding what the true nature of propositional content really is, primarily, because, I am, in this essay, focused upon concepts. However, the sententialist perspective is the theory of propositions on offer that explicitly postulates and requires concepts theory of, and it correctly, it seems, postulates compositionality in cognitive systems. Further, it is the

[12] It should be mentioned, before we move on, that the compositionality of thought is often marshaled in support of the position that language and thought are isomorphic, because language is fairly obviously compositional (larger, more complex meaningful units are constructed from smaller, simpler semantic units and the meaning of the larger units is dependent upon the meanings of smaller units). See Fodor 1998 pages 25-28; Fodor and Lepore 1999 page 383. However, this is beyond the scope of this essay, since it is focused on conceptual content and its implications with regard to understanding thought as being non-linguistic. Compositionality is another battle for another day.

theory of choice for theorists of mental representation that argue for the following positions (which directly concerns the proper way to understand conceptual content) that needs to be dealt with here: that because sentences/propositions cannot have graded content, neither can their conceptual components. It is to this issue that we will turn in the next section.

1.5 Reasoning from Propositional Content to Conceptual Content

There is a great deal of argumentative slippage between the contents of beliefs and other attitudes (propositions), and the content of concepts that compose to generate propositions. It is not uncommon to find in the literature on mental representation arguments that proceed as though the general semantic facts that one might find holding for the contents of beliefs (propositions) will also hold, necessarily, for concepts, and vice versa.[13] Thus, often, when one finds arguments about beliefs, or more accurately, their propositional content, one also finds claims that the conclusions of these arguments apply to all mental representation. I am of the opinion that this argumentative strategy is misleading, and will endeavor to convince my reader that this strategy is misleading.

In particular I have in mind arguments run along the lines of claiming that the propositional content is such that a belief that-P, either is, or is not the belief that-P[14], say, and as such mental representation as a whole, either are, or are not correctly applied when they are applied to a state of affairs (in the case of propositions), or an object (in the case of concepts). Dretske (1981, 57-62) holds that the content of propositions are either

[13] See Fodor 1975, Fodor 1987, Fodor 1994, Fodor 1998, Dretske 1981, Harman 1981, Block, 1986, Putnam 1975. See Cummins 1989 Chapter 1, and Cummins 1996 pages 15-16 for observations regarding the ill-foundedness of this symmetry.

[14] That the belief cannot be the belief more-or-less-that-P.

possessed, or believed, or entertained, in their entirety, or they are not possessed, etc. at all. And so all of their content is possessed by the mind in possession of the proposition, or it is not. One cannot only possess, or believe, etc. a proposition in part, or more or less.[15] This position likely finds some of its motivation in thinking about propositions in terms of their truth or falsity. And quite clearly, propositions are entities (on an interpretation consistent with classical logic) that are either true or false. Truth and falsity are a binary set of values (that can be represented as '0' for false and '1' for true), and in classical logic, believed to be exhaustive[16]. So, it is easy to see how other semantic qualities of propositions, like the nature of their contents, could be understood as being absolutely one, or the other, as well. In order for a proposition to be true or false, it must apply or not apply to the world. This idea seems to spill over into thinking that conceptual contents can be true or false, or truly or falsely applied, in that they either apply to a referent, or they do not. Cummins (1996) attributes to Jerry Fodor the idea that a concept ought to be understood as correctly applying to (representing) a referent when it can be said of the referent, to which the property is ascribed by the concept, that it is true of the referent that it possesses the property ascribed to it by the concept (8). Further, Fodor (1987) discusses the application of a concept to a referent as being either "veridical" or "*un*veridical" (101). Fodor (1998) puts matters the following way, "Greycat the cat, but not Dumbo the elephant, falls under the concept CAT. Which, for present purposes, is equivalent to saying that Greycat is in the extension of CAT, that 'Greycat is a Cat' is true, and that 'is a cat' is true of Greycat" (24). This quotation from

[15] See below for a more detailed discussion of Dretske's point.
[16] I wish to remain neutral with regard to the gradability of truth and falsity, though it would not surprise me if truth and falsity were graded sorts of things. I also wish to remain neutral in this paper with regard to what the One True Logic will be.

Fodor (1998) illustrates nicely the spilling over of thinking about propositions to thinking about concepts. Dretske (1981) discusses the truth or falsity of mental representation in total (195), and Dretske (1981) formulates concept application in terms of being correct, or incorrect (225), of either applying or not applying.

1.6 An Argument for the Absolute Content of Propositions

The idea that propositional content can be graded may well strike one as counter-intuitive. And it is relatively easy to see why this might be so. We very often express propositions to one another by using the same sentences. For example, we say, "It is raining" to each other to convey, what seems to us to be the same idea, because we use the same sentence. Though, it also happens that we also seem to use different sentences to convey the same thought, as is frequently thought to be the case with regard to two different sentences in two different languages that mean the same thing, that is express the same propositional content, like "It is raining" and "Il pluit" both express the proposition, thought, that *it is raining*. What I would like to suggest, is that we tend to believe this, and that it has a great deal of intuitive appeal, because we who use the same language use the same sentences to express our beliefs etc. I am inclined to believe, and will argue that this intuition is one that we should be less certain of[17].

It seems to me that it is this intuition to which Fred Dretske (1981, 57-62) is appealing when he makes the claim, in the context of developing an informational

[17] The idea that we should be less certain that the content of every person's belief that is expressed with the same sentence should not be assumed to have, in fact, identical propositional content has been promoted by many who think that language is holistic in nature with regard to the meanings of its components, or that mental states are holistic with regard to their meanings. See Harman 1982; Davidson 2001(a), 2001(b), 2001(c); Block 1986; See Fodor and Lepore 1992 for a detailed discussion of holistic theories of content. Interestingly, holists about meaning tend to think that if it holds for language it will hold for the mind, and vice versa.

semantic theory, that the content of a belief cannot be transmitted via an informational channel (whatever it may be, speech, perception, etc.) only in part. The information that the proposition that-P can only be transmitted completely, for if it is not transmitted completely, it is not transmitted at all. Suppose we were to place the success of the transmission of information on a scale of 0 to 1. On this scale, 0 is stands for no-information being transmitted, and 1 stands for all of the information that can be transmitted being transmitted. So, the information that-P can only either be transmitted completely, where all the information contained in the proposition that-P is transmitted (a value of 1 on the transmission scale), or the transmission fails when it does not communicate all of the information that-P. So, if the transmission of a proposition does not transmit all of the information contained in the proposition, (at a value of 1 on our scale), then none of the information contained in the proposition is transmitted (and so we get a value of 0 on our scale for the transmission of information in the proposition). Thus, the information that-P is not transmitted, and it is not transmitted even a little bit. The transmission of propositional content is an all or nothing affair. And that's just the way propositions are. There exists no possibility, on Dretske's account, of a partial transmission of the proposition that-P. Thus it also follows, one either believes that-P, or one does not believe that-P, but one cannot believe more or less that-P, because one cannot believe that-P, or possess any of the content of that-P, without possessing all of the content of that-P.

1.7 Further Arguments for the Absolute Content of Propositions

Jerry Fodor agrees with Dretske on this point. And in his (1987), Fodor makes an argument for the position that the content of a belief (the proposition that it is an attitude toward) is not something that can be graded. So, Fodor argues that propositional content cannot be a more-or-less sort of thing, and also claims that because belief content is not gradable no mental representational content is gradable, and so mental representations as a whole, and so concepts, do not and cannot have graded content. Fodor's argument occurs in the context of refuting Holism with regard to representational systems, where holism is the idea that the meaning of a representational state in a system is dependent upon certain other representations—that certain other representational states are required in order to have certain other representational states. This results in an eventual slide down a slippery slope to the position that the meaning of any given representation (concept or proposition) is dependent upon every other representational state in the system for its meaning. Fodor calls the representational states that determine the meaning of a given representation the determined representation's "Epistemic Liaisons." While the project of this essay is not directly concerned with Holism, the argument Fodor offers against Holism has broader applicability than merely to Holistic theories of content.

Fodor's argument against the sensibility of a graded[18], "more or less," notion of (mental) representational content proceeds in the following way: Let us focus upon a proposition, and an attitude of which the proposition is the content, that most of us would

[18] This is how Fodor's argument can be more generally applied to non-holistic theories of content.

understand, say the belief that *Tibaldi is a better singer than Callas*.[19] Now, it is fairly obvious, according to Fodor, that we can grade epistemic commitment. You can be certain of the truth of a proposition/sentence like *Tibaldi is a better singer than Callas* to a greater, or lesser, degree—anywhere between absolute certainty and total doubt (which is usually represented as 1 for the former, and 0 for the latter). However, it is not obvious, according to Fodor, that you can believe more-or-less that *Tibaldi is a better singer than Callas* (let us shorten this to *Tibaldi is better than Callas*—nothing hinges on 'singer,' or 'a' here). You cannot believe that *Tibaldi is better than Callas* to a degree between 1 and 0, i.e. you either believe that *Tibaldi is better than Callas* or you do not believe some proposition that would fall under the English expression 'Tibaldi is better than Callas,' if you have, in fact, any belief about the matter at all.[20] So, a graded notion of content for, a version of which is offered by Holistic theories of content, will not work, because the meaning of the proposition *Tibaldi is better than Callas* is dependent upon its epistemic liaisons, and these epistemic liaisons will be different to varying degrees in every human being.

This is a problem, because it has the consequence that, if we were to idealize (set the value between 0 and 1, or at least, a minimal amount of meaning required to possess the meaning of the proposition), with regard to the epistemic liaisons a particular proposition has; we have to pick one set of epistemic liaisons—those that a particular person has, or a particular set of homogeneous minds have, and Holistic theories of

[19] Fodor's example is "Callas was a better singer than Tibaldi" (Fodor 1987, 55), but that I do not believe that the order of the names in the example has much bearing upon the argument Fodor makes.

[20] This is not meant to be about cases in which one believes that *Tibaldi is better than Callas*, or one believes that *it is not the case that Tibaldi is better than Callas* (i.e. believes that *Tibaldi is better than Callas is False*), and not both.

content provide no principle means with which to do so (Fodor 1987, 55-57). And this would require that propositional content be, at base, identical, and so cannot be holistic for the purposes of holding that each person's beliefs, etc. are not exactly the same (i.e. not identical).

Further, a more generally motivated notion of graded content will not work either, because the move to idealization (the move to pick out the meaning against which the instances of more-or-less that P can be "measured") is itself fraught with difficulties. Idealization is understood to be like scientific idealization (frictionless planes, pure chemicals, etc.) setting an asymptotal value (or meaning) that the values (meanings) of the more-or-less content Holistically determined propositions approach, but never meet. So, if we claim that you believe proposition *Tibaldi is better than Callas* $n<1$, or believe *Tibaldi is better than Callas* $n>0$, then we would have to have a determinate idealized value, or meaning, for n, some value greater than zero and less than-equal to one. This would mean that we would have to have an account of what it is to believe a proposition *tout court* (whatever value greater than zero and less than-equal to one that we set n to). And if we have this, then we have the criterion for whether you believe that *Tibaldi is better than Callas* or you do not believe that *Tibaldi is better than Callas*. Thus, there is no need for a graded notion of propositional content, because a graded notion would presuppose that there already are determinate conditions for believing n, or not believing n. And since well defined conditions would exist for what it is to believe n, the work that graded notions of content would be doing would be unnecessary.

It ought to be mentioned that some of the motivation for Fodor's argument against content "more-or-less-that P," is that he is an intentional realist who holds that Intentional

Attitude Psychology is mostly true, and if propositional content is not absolute and identical among the holders of a particular proposition, then the explanatory and predictive power of Propositional Attitude Psychology is not very high. Because if it were the case that propositions were "more or less that P," then your and my belief that P would be different. Hence generalizations with respect to the role that the proposition that-P play in minds with regard to behaviour, etc. would not likely be very successful generalizations, admitting of a large number of exceptions, which would, more or less, defeat the purpose of making generalizations for the purposes of explaining human behaviour at large.

1.8 Why it is Not Obvious that Propositions Have Absolute Content

However, I do not think that Fodor's conclusions regarding the contents of propositions follow as obviously as he would like. One reason for thinking this is that he begins with the postulate that there is a more-or-less-that P (where P is some proposition), which pretty clearly assumes that there is, or could be, believing that-P *tout court*. However, believing that P may not exist, since it is possible that the sentences that express propositions pick out, or label, a set of mental states or relations between objects, where the proposition is just a label for a number of similar, but non-identical mental states. However, pursuing this line of inquiry would lead us into a discussion of matters having to do with similarity and identity of content, and arguments, put forward by Fodor (1998) (and by Fodor and Lepore 1992, 1999), that any notion of similarity of content will depend upon an already presupposed robust notion of content identity. The issue of similarity versus identity will be dealt with in the necessary detail in Chapter 3, and

before getting to this issue I would like to focus upon the idea that it is possible that there are groupings of relations for which phrases like "better than" are merely labels for and in so doing, tease out more about the implications propositions are thought to have for concepts and the gradability of content. So, for the present discussion I will skirt the issue of the relationship between similarity and identity of content.

It is likely that my reader would agree that 7>4, and also that 5>4, and that these two strings of numbers joined by '>' mean different things, namely that *seven is greater than four* and that *five is greater than four*, respectively. So, suppose I were to think that Tibaldi possessed a singing quality value of 7 (if we could easily quantify such things and let us suppose for the sake of argument that we can), and you believed that Tibaldi had a quality value of 5. Suppose further that we each believed Callas to have a singing quality value of 4. It would be true that that the sentence "Tibaldi is a better singer than Callas" applies to both of us, but it seems that despite this, it is not the case that our beliefs have identical content, because of the difference in degree of our quality assessments of Tibaldi. My belief that "Tibaldi is a better singer than Callas" differs by a certain degree from the content of your belief that "Tibaldi is a better singer than Callas" in much the same way that a belief that *7>4*, and a belief that *5>4* differ in content with regard to how much more the first argument is than the second, which would seem to affect the semantic content of "greater than." As such, I believe it would be fairly accurate to say that we both believe, more-or-less-that "Tibaldi is a better singer than Callas."

To continue to address Fodor's (1987) argument: it is not obviously the case that there exists a proposition with the content that *Tibaldi is a better singer than Callas tout court*, against which we could compare our more-or-less beliefs that "Tibaldi is a better

singer than Callas," unless, of course, we were just willing to stipulate one. But then we would need a good argument for why we are stipulating what amount of better counts as *the* "better than" relation, and this is especially difficult given that real numbers/values could be employed, and so there is no least number, or amount, larger than another. So, any person's valuing of Tibaldi's ability to sing, or its quality, could be infinitesimally smaller better than whatever the quality of whomever Tibaldi is being compared to, like real numbers/values. So, it is difficult to find a reason why my "better than" relation with regard to Tibaldi and Callas would be identical to yours. This would imply that there exists no particularly good reason to believe that the phrase "better than" is not just a generalization over a great many states of evaluative comparison between entities to which differing values, and differing differences between the differing values, are assigned, across many individuals.

Though, there is certainly an existent sentence that can be used to communicate most of what I believe, and most of what you believe, but this does not imply that we believe the same thing *tout court*. It is not obvious that, because we have a phrase that each of us can use to express some state of our minds, the two states being expressed have to have identical representational content. All this really implies is that we have some rough and ready equivalence of meaning with regard to our propositional mental states such that we are capable of getting the gist of whatever it is the sentence is being used to communicate. The sentence 'Tibaldi is a better singer than Callas,' is a generalization that covers at least two different, but somewhat consistent, sets of measurements made by two individual neural/cognitive systems with regard to the same two objects (Tibaldi and Callas).

However, there is the possibility of a content identity preserving response that runs along the following lines: in these instances where we have different beliefs that can be understood as different proposition, e.g. the belief that *Tibaldi is a better singer Callas to a degree of 3* and your belief that *Tibaldi is a better singer than callas to the degree of 1*. However, this would mean that we have different beliefs. And though it might be possible that at least two people both have an identical belief with regard to the degree to which Tibaldi is better than Callas, it is unlikely that this would be widespread. But given a fine enough scale of singing judgment, it is also unlikely that anyone would even have the same belief with regard to the degree to which Tibaldi is better. But, even if there were two people, or a few more, of the same belief with regard to degree, there would still be an identity of belief content. And so in order to say that two cognitive systems would have the same (i.e. identical) belief.

It should be kept in mind, as discussed above, that the chances of this being widespread are slim, and it would make the identity theorists victory a little hollow, since such a small circle of cognitive systems with identical propositional contents would have very limited applicability with regard to their generality with respect to more than a few agents, and as such would prove only useful in explaining and predicting the behaviour of a very few individuals. This should give a content identity theorist like Fodor some pause, given that one of the primary reasons the identity of propositional attitude content is hoped to be identical is to facilitate the plausibility of intentional attitude psychology, with regard to explaining and predicting the behaviour of people in general. So, there would be little advantage to holding on to content identity over the gradability of content.

1.9 The Analytical Priority of Conceptual Content over Propositional Content

So, if it is the case, as I have argued, that propositional content is graded, then we cannot argue that concepts cannot have graded content because propositions do not. This is not what I intend. I intend to merely to make an argument for the plausibility of propositional content similarity (gradability) in a manner cautious of the argument for content identity, and by doing this block the move from propositional content identity to conceptual content identity outright by denying propositional content identity[21]. But what if I am wrong about the gradability of propositions? In case this is so, further arguments to the effect that it is a mistake to argue that if something holds for propositional content, then because propositions and concepts are both mental representations, what holds for propositions will hold for concepts.

One reason not to accept that what will hold for propositional content will hold for conceptual content is that it seems to get the order of explanation wrong. Concepts are postulated as those meaningful elements from which the meanings of propositions, the contents of beliefs, desires, etc., are derived. Given the fact that the meaning that arises from the composition of concepts/words into propositions/sentences is going to be dependent upon the meanings of the terms, and that the meanings of these concepts are going to determine, at least in part, the meaning of the whole proposition. This is just the systematicity, productivity, and compositionality, point again, and should not strike anyone as too controversial. So it should be relatively obvious that in order to properly understand the nature of propositional content one will have to first determine the nature of the content of its content providing components (concepts).

[21] The more intricate matters regarding the relationship between content identity and content similarity will be indirectly addressed until Chapter Three where this matter will be addressed in detail.

Further, propositions may enter into significantly different relationships with the world than their contentful elements might. The proposition/sentence *There is a dog*/'There is a dog' will generally be regarded as being true or false in a given context. And even if we were to suppose that the isomorphism between language and thought held, most of us, I do not think, would not be willing to say that the word 'dog' is true all by itself. The mere utterance of 'dog' (unless it is shorthand for the sentence 'There is a dog', or some such, and then it is the sentence that has been shortened that bears the truth value) does not make the word true or false, and so the same would follow for the concept DOG. DOG could not be true or false all by itself; a property, in and of itself, is not a thing that can be true or false. But it is fairly obvious that it is the case that we can reasonably hold that sentences (of the declarative sort), and propositions, to be true and false.

Despite this, theorists of mental content tend to understand not only propositions as true or false, right or wrong (that is, bivalently, as they are traditionally interpreted, intuitionistic and sub-structural logics to the side for the moment) but also individual mental representations/concepts as being tokened rightly or wrongly, correctly or incorrectly, i.e. bivalently. Essentially, concepts are viewed as being tokened truly or falsely. This, it seems to me (and was argued above), is a common motivation for understanding mental representations as being bivalent on the linguistic isomorphism view. The concept DOG is understood as either applying to an object in the environment, or not applying, but this probably is not accurate, because the tokening, or the occurrence, of DOG, is not something that can be evaluated as representing the world truly or falsely, because it does not, as a matter of fact, represent the world as being any particular way,

like a proposition is usually conceived as doing[22]. This issue will be addressed in more detail in the next chapter.

So, what I would like to suggest at a minimum, is that even if propositions and the propositional attitudes are not gradable with regard to their content, for reasons governing their composition, perhaps, or maybe the way propositions "hook up with" the world, (though I doubt that propositions are not graded), it may still be possible for individual representations (concepts) to be graded.

So, in the next chapter we will turn to more directly addressing the primary subject of this essay, namely, the matter of whether or not conceptual content ought to be understood as being graded, or gradable, in the context of problems that arise for linguistically inspired content identity theories with respect to handling misrepresentation. In doing so we will address further difficulties that arise from the assumption that concepts work like words with regard to their semantic properties. I will make the case that concepts cannot be understood as entities that possess contents that are identical to their referents, and so cannot be understood as being such that they either apply or do not to an entity in the same way that words do.

[22] See Cummins 1996 for multiple arguments for appropriately understanding representation as dissociated from matters regarding truth.

Chapter Two:

Misrepresentation

2.1 Words, Concepts, and The Disjunction Problem

The central notion that I am arguing against in this essay is that concepts function in the same way that words do (as described in detail in the last chapter), where both a word, and a concept, either apply to an object or do not. Difficulties with such a view begin to arise when we consider that the standard of correct application, in the case of words, is plausibly determined by the community of language users, whereas in the case of concepts, the linguistic community is not obviously available to play this role. So, something else needs to do this work for mental representation. Take the word 'dog,' for instance. 'Dog' applies only to dogs when we use the term literally. We use 'dog' to refer to dogs, and this referring relation is underwritten by the larger language using community that is able to say something like, "all of these things here (dogs) that are like this (doggish) in the relevant ways, these things are what we are going to call 'dogs' and use the word 'dog' to refer to them." 'Dog' refers to dogs by convention, an agreement, spoken, or unspoken, among a group of people about what we are going to be referring to when we use the word 'dog' (literally). And 'dog' names, or refers to, all dogs, in every user of the word's mouth, regardless of whether or not the user of the word can, or would, recognize every member of the set of dogs as a dog, and so be able to correctly apply 'dog' to every entity to which the word could be correctly applied. The extension, or referent, of the word is all of the entities a community would classify or group together 'under' the designating term. This is, in fact, a rather broadly espoused theory of how it is that arbitrary symbols like words come to represent the things they represent, i.e.

become conventional (Strawson 1964; Lewis 1969; Kripke 1982; Stainton 1996, 183-194).[23] Words can work this way, relatively unproblematically, because there are people around us who will tell us when we find ourselves applying an utterance of 'dog' to something that is not a dog, when we make an error of attribution, or a labeling error.[24][25]

In these instances, the term 'dog' is either applied correctly,[26] or it is not, because there is a you-independent standard upheld by the community of language users, in which one has found oneself long enough to learn a particular language, that determines whether or not you are using the correct word. The reference of the word is established independently of you. This, however, does not conflict with the notion that the meaning of linguistic terms are dependent upon the mental representations of language users, since it seems that there can not be conventions without minds generating, and following the conventions (Fodor 1998, 149). However, the details of this dependency are not exactly clear; it is not clear how complex and mediated the relationship between 'dog' and DOG is. When one learns the meaning of a word, it is likely a mapping of the word, 'dog' say, onto the mental representation DOG, or some such thing, that in some way represents dogs (or the properties of dogs) in the mind independently of the word 'dog'. It seems

---

[23] See Russell 1905/2005 for a rather famous formulation of the view that the meaning of the word is its reference (the denotational view, see Stainton 1996). I am suggesting here that the denotational theory exemplified by Russell, and the Convention view, are two ways of thinking about language that are intuitively compatible, and seem to both influence thinking about mental representation as being language-like.

[24] This is not meant to be *the* way that theorists in general understand linguistic meaning. It is merely meant to draw out some ways of thinking about language that have had a rather large impact upon theories of mental representation. Though, see Putnam 1975 page 247 for an explicit formulation of this combination of views.

[25] See Fodor 1987 page 73 for the influence of this theory of linguistic meaning upon his account of mental content.

[26] Here we might understand the word is being understood as denoting a particular property, or set, and this denotation is possibly determined by the community of language users. Where "denotes" is the technical term in such theories for 'refer to', or 'pick out'. See Stainton 1996.

somewhat natural to think that when one applies ‘dog’ to dogs, when dogs cause DOGS, that you are then correctly applying the word (more or less). However, it is possible that one could apply the word correctly even if one’s DOG did not get tokened, but one uttered ‘dog’ in the presence of a dog, like when one is/were learning the word and do/did not yet know the meaning of the word. So, it would seem, at least, on the face of it, that ‘dog’ and DOG are dissociable.

Notice that the meaning, or, at least, the determinant of the correct use, of the word in the linguistic case seems to be the set of all things to which the community applies the word literally, or, at least, is the property that determines the set. This “direct reference” or “denotational theory” carries over into many theories of mental meaning, as we will see below (see Fodor 1987, 73). An error in language is the misapplication of a word to something it does not represent, or apply to when you are trying to apply it correctly. And since the community maintains the criterion of rightness of the application of the word and they are able to correct/punish you when you get it wrong. This is a common experience of nearly every child, I imagine.

There is a comparable, and quite nearly parallel case of application error that is discussed in the mental representation literature from which the linguistic case above is modeled). It is often called misrepresentation, or even just “error.” It is very much like the word application case above, but does not necessarily, or even likely, involve words. However, it does maintain, as the linguistic example does, the view that one either correctly applies a property, or category, to an object, or does not. Here is how it works.

In a case of error (with respect to mental representation now), there are whatever conditions, or sets of relations, that instantiate or underwrite, depending upon one’s

theory, the 'representation relation' between states of one's brain/mind (representational states) and things in the world that one is capable of representing. One finds oneself in a situation in which one's content underwriting conditions are cheated in some way so that one ends up representing some thing as some other thing. Perhaps an illustrative example: suppose I find myself in a situation of low light, and compromised audio fidelity—whatever mechanisms I have that play a role in instantiating the content, or meaning, of my mental states—and I detect, or spot, some four legged creature moving about in the distance, and it causes/results in the tokening of my mental representation of a DOG. But there is a problem: what I saw was not a dog, it was a large cat with a doggish body structure (or something) that was not, and is not, a dog. And so DOG is misapplied to the cat, because DOG does not correctly apply to the cat because cats are not dogs, and DOGs are 'supposed to' represent (are about) dogs not cats. Thus, in virtue of tokening my DOG representation, instead of my CAT representation in response to the detection/seeing of a cat, I have *misrepresented* the cat as a dog.

No doubt, this happens quite often to creatures that represent the world around them. And because misrepresentation is a regular feature of representation, in order to generate a satisfactory account of mental representation one must, when specifying the content instantiating relationships between mental states and things in the world (whatever those instantiating conditions are), do so in such a way that the DOG representation means *dog*. So, one's content instantiating conditions need to be set up in such a way that there are things that are not dogs, e.g. cats, that cause DOG under these instantiating conditions, are excluded in a principled way from the content of DOG. Further, such an account must also not only exclude cats contributing to the content of

DOG, but they must also explain how it is that cats can cause DOGs without cats being part of the content of DOG, which they should not be. If cats, etc. cannot be excluded from the content of DOG in a principled way, then what we end up with for the meaning of DOG is not *dog* as it should be, but *dog or cat or whatever else happens to cause the tokening of DOG*. This problem is known as the "disjunction problem" (after Fodor 1987), because it is a problem of a disjoint content existing where there should not be such a thing, and has the consequence of making it possible that nearly every concept has as its content nearly everything, because nearly everything could go in the disjunction under these circumstances. This is an especially large problem for the views we are considering, because what causes the representation in the specified conditions of content instantiation *is* the content of the representation[27], and so anything that triggers the representation/mental state must be a constituent/part of the content of that state (as would be the case in a disjunction).

So, the content instantiating relations must be set up in such a way that misrepresentation is explicable in a way that does not allow for the sorts of disjunctions just discussed to be the content of the representation, unless they ought to be. Most theories of mental representation are structured in such a way so that occurrences of things in the environment cause tokenings of their corresponding representation, as mentioned above, is a major concern when attempting to construct, or properly describe, the relations that instantiate meaning (Fodor 1987, 1998; Dretske 1981; Millikan 1989; Cummins 1996; Ryder 2004; Eliasmith 2005).

[27] This is like the linguistic case in which the word labels or denotes a property, or class, which is the content of the word.

To build upon the immediately preceding: It seems to me, and I will attempt to demonstrate in what follows, that a good deal of the difficulty in dealing with the disjunction problem for a number of accounts of mental representation is due to thinking that mental representation is like propositions/sentences in the sense that they are either true or false; and thinking that mental representations work like words in the conventional/direct reference case are tokened or applied, correctly or incorrectly (when used literally) as described above. Recall, though, that content is of the world is taken to be the set of objects to which we agree the word refers/represents. So, in what follows an argument to the effect that the practice of identifying the content of a representations with what it represents, as is often assumed because of the acceptance of the isomorphism between thought and language does not, in fact, even allow for error to occur. Following this an argument will be put forth for what is needed to allow for misrepresentation to occur, and explain the occurrence, in a cognitive system with regard to its mental representations.

2.2 Content Identity and its Difficulties with The Disjunction Problem

I have made the point in previous sections that the identification of the content of a word with what the word represents is very often accepted by theorists of mental representation to apply also to concepts and their contents. Theories like these tend to postulate some law-like, or "nomic" relation between the thing(s) represented and the content of the representation, and this relation not only determines, in some way, the content of the representation, but makes it so that what is represented *is* the content of the representation—that is the content of the representation is not distinct from the thing

represented. The content of the representation=the thing represented. The law-like, or nomic, conditions that instantiate the representation relation are some set of natural(istic) conditions that are the representation relation. In the case of Fodor (1994; 1998) and Dretske (1981) it is an informational relation between a thing in the world and things in the head that instantiates the representation relation. The informational view is a version of the broader collection of theories that postulate some sort of causal relation between object and representation that instantiates the content relation, e.g. Fodor (1987), Stampe (1979), and Kripke (1980) (if you believe Sterelny (1991, 116-118)). There exist also teleological views like that of Dretske (1988, 1999) and Millikan (1984, 1989), who hold that what instantiates the representational relation is biological function of mental states of neurobiological systems to represent certain parts of the world determined by natural selection in an organisms evolutionary past, that is the mental state has a history such that it has adapted to represent a particular type of thing.

Cummins (1996) also lumps all of the causal-content identity postulating theories into a category he labels "Use Theories" of meaning (see page 53 in particular). He does this, because in all of these theories correct representation is determined be the correct use of the representation, that is, the correct application of the representational state to the object it represents.

> To use a representation is to apply it to a target [some thing in the world]. Uses, then, are simply applications. To specify how a representation is used on a particular occasion is to specify a particular target[. . . .]The fundamental idea is simple: in a case of correct use [of a representation], content = target. So, if we

know what a representation is applied to, and we know that the use is correct, we know the content (29).

Basically, all of the theories listed above contend that content is determined by correct use, and the content of a representation is identical to that which it represents. Not every theory listed above is clear about these commitments. However Fodor (1987) and (1998) is very clear about his theoretical commitments in these matters. So, Fodor (1987) and (1998) will be employed as the foil(s) here, because he most explicitly demands and makes the case for a causal relation between referent and content and the identity of content and referent.

2.3 Causal Relations, Content Identity, and Not Solving The Disjunction Problem

It seems that causal theories of content that postulate the identity of the content of a representation and the representation's referent have a bit of a problem. Fodor (1987) calls it "an embarrassment".

> It seems that, according to [causal theories like this], there can be no such things as *misrepresentation*. Suppose, for example, that tokenings of the symbol 'A' are nomologically dependent upon instantiations of the property *A*; viz., upon A's. Then, according to the theory, the tokens of the symbol denote A's (since tokens denote their causes) and they represent them *as* A's (since symbols express the property whose instantiations cause them to be tokened). But Symbol tokenings that represent A's as A's are ipso facto veridical. So it seems that the condition for an 'A'-token meaning *A* is identical to the condition for such a token being

> true. Hence, then, how do you get *un*veridical 'A' tokens into the causal picture? (Fodor 1987, 101).

Attempting to determine how it is that we get "unveridical" 'A's result in the disjunction problem, and it arises for familiar reasons. We might say, in order to account for misrepresentation, that some B causes an 'A', which is fair enough. But because we are working with a causal theory that identifies the content of a representation with what it represents, we run into the disjunction problem. Anything not-A, B for example, that tokens 'A' will therefore be part of the content of 'A', and so *mis*representation does not occur. What has happened is that a representation 'A' whose content is actually *A or B or*. . . has simply been tokened by one of its disjuncts, and this is not misrepresentation. It is representation. Thus misrepresentation is *prima facie* impossible for a causal-content identity theory.

Members of the causal-content identity family of theories of representation have attempted to show that misrepresentation can happen, and they have attempted to explain how it happens, and to thereby solve the disjunction problem. However, these ways of solving the problem have not yet been successful. Dretske (1981) proposes that we understand the content of a concept to be determined by the correlation that is developed between a mental state and a set of a kind of object in the world over the course of a period of time. The correlation is built up from the point of there being little to no correlation to the point where the kind of thing and the mental state are reliably correlated. This process will determine the period of time over which this is accomplished, and the period of time over which the correlation is built is called the "learning period." The mental state under discussion will then represent the property

with which it has built up a strong correlation with over the course of the learning period[28]. So, a misrepresentation occurs whenever a kind of thing that was not a member of the learning set causes the tokening of the mental state correlated with the members of the things that were in the learning set during the learning period.

The difficulties with this solution are as follows. It is not the case that we can guarantee that the learning period of a cognitive system with regard to the building up of a correlation between a mental state and a property in the world can ever be pure. It would be impossible to ensure that only the property, *P*, was correlated with the mental state that we would want to say represent *P*, and so anything else that made its way into the learning set would be part of the content of the mental state. And so, the disjunction problem returns. Further, there is no non-arbitrary point at which we can say that a learning period has ended, because the correlation of a mental state and a property will be continually strengthened throughout the existence of the cognitive system. Thus we cannot exclude future instances of non-*P*s causing the tokening of the mental state, and say that they do not count as part of the learning set (because learning has not ended), and so we cannot in a principled way not count the non-*P* as part of the content of the mental state that we want to reserve only for *P*s.

There has also been offered a teleological solution by Dretske (1998) and Millikan (1984, 1989). According to the teleological solution, the content of the mental state of a neurological/cognitive system is the kind of thing that has the function to represent some property, *P*. And this function of a mental state to represent a particular property is an evolutionary adaptation brought about by natural selection. So, a mental

[28] See Dretske 1981 pages 222-231 for his development of this idea.

state represents what it does, because that mental state has been adapted, over the course of evolution, to causally correlate with that property. So a misrepresentation occurs when an entity of a kind that the mental state was not selected to correlate with causes the mental state to be tokened, i.e. the entity that tokens the mental state is not of the kind that it is the mental state's evolutionarily determined function to represent.

This solution is not itself unproblematic. The major difficulty with this view revolves around the fact that it gets the order of explanation wrong. A representational mental state would be adaptive, selected by natural selection, because of what it means, and so would be selected because it would mean what it meant. A selective advantage could only be conveyed by a representational mental state if it already possessed a particular content. It would not then have its meaning because it was selected. The meaning would not be adaptive because it was selected. Similarly with most cases of selection and adaptation, a feature of an organism is not adaptive because it was selected, but is selected because it was adaptive. Eyes and similar structures are not good for detecting certain parts of the electromagnetic spectrum because they were selected. Eyes were selected because they were good for detecting certain parts of the electromagnetic spectrum.[29] Cummins (1996) makes this point as well (46-47, 55-57).

It seems then that there are serious problems with the standard causal theories of representation that postulate content identity. What then is a theorist who wants to hold on to causality and content identity to do, if he wants to account for misrepresentation and solve the disjunction problem?

[29] This is somewhat of a toy example, but it makes the point that we need to make.

2.4 Introducing Asymmetric Dependence

Jerry Fodor (1987; 1998) has proposed a solution to the disjunction problem and it is called the "Asymmetric Dependence" theory. He proposes that the misrepresentation of a cat as a dog (when the cat causes the tokening of DOG) is dependent upon the already existing dog-DOG representation instantiating relationship, whereas the dog-DOG representational relationship does not depend upon the existence of the cat-DOG relationship for its existence—the cat-DOG cannot have happened (existed) without the dog-DOG relationship already being in place, and so the misrepresentation of the cat as a dog, by the cat causing a tokening of DOG, is asymmetrically dependent upon the dog-DOG nomic relationship.

However, this is not a naturalistic solution, which is what we and Fodor are after, because the natural(istic) mechanisms by which this asymmetric dependence is enabled to occur are never identified. This solution, according to Fodor, "is purely formal" (Fodor 1987, 110), i.e. it restates the problem, and perhaps clarifies what is generally going on in cases of misrepresentation—that cat-DOG misrepresentations are dependent on dog-DOG representations, and dog-DOG representations are not dependent upon cat-DOG misrepresentations. And this seems right, as far as it goes. But it does not go very far, because it is not a naturalistic solution.

How, then, ought we to go about solving this problem? That is, how do we make room for error? How is it that misrepresentation is asymmetrically dependent upon representation?

2.5 Readjusting Asymmetric Dependence in Light of Content Similarity

The first step, it seems to me, given the arguments made above, has to be a relaxing of the condition that the content of concepts are identical to the objects that they represent, because, as we have seen, construing mental representation in this way does not appear to allow for misrepresentation[30]. This is because such theories merely subsume the misrepresented item as part of the content of the representation that has been inappropriately applied to it. Let us try to formulate Asymmetric Dependence in a non-content identical fashion that can be used to guide the inquiry into this problem for the remainder of the essay. This reformulation will, of course, not provide natural(istic) mechanisms that explain misrepresentation either. However, it is designed to enable a re-conception of what a solution to the disjunction problem needs to look like so that a naturalistic solution can be found.

Recall the following familiar situation: Suppose I, in a situation of poor lighting, like that described above, misrepresented a cat as a dog, and suppose that I identified the cat as the dog because the cat caused my DOG to be tokened. What might I say to someone who asks you why I thought the cat was a dog? Might I not reasonably reply that the cat looked like a dog? Might I also not be able to say in defense of my incorrect representation tokening that there was a significant degree of similarity between this particular cat, in this light, and most dogs that I have seen, say with regard to size, and movement, and even shape—and that this is the reason why I mistook the cat for a dog. Similarly, might it not be reasonable to suppose that I am only able to misrepresent a cat

[30] Or, at least, do not allow for it at al readily.

as a dog because cats are structurally, texturally, etc. similar to dogs in many ways, and that their corresponding concepts have a fair bit of their content in common, because cats are similar to dogs, though, obviously, there are a number of important differences. Hence there are different categories and concepts for them. So when a cat shares more with your DOG representation than it normally does, and more with the features correlated with DOG than with CAT (under certain lighting conditions), that these factors could result in my misrepresenting a cat as a DOG. In short: the cat has enough points of correlation in common with the content of your DOG state to cause its tokening, instead of a CAT being tokened.

This is consistent with the form of Asymmetric Dependence in the following way. I have my dog-DOG relationship, which is likely set up because features that dogs have in virtue of being dogs (I suppose) are related in certain ways to my DOG representation. So, there is that relation. Now, the cat-DOG relation will be dependent upon this relation in a way that the dog-DOG relation will not be dependent upon the cat-DOG relation. Because the cat has (under certain observation conditions), or appears to have, a number of features that dogs normally have, and so because the dog-DOG relationship is dependent upon/determined by these features, and the particular cat is similar to dogs to an abnormally large degree, it can "use" these similarities, if they are sufficient, to cause DOG to be tokened. The cat-DOG relationship is dependent upon the dog-DOG relation, because if there were not those features of dogs that are correlated with, and determined the content of, DOG, then there would be no way for a cat duplicating those features to cause the tokening of DOG. We will likely have to abandon "features" as we progress through this essay, but it serves to provide a similarity metric that is easily intuited.

Thus, for a theory that works as has just been described there has to be some similarity of content between CAT and DOG, and this may lead to another version of the disjunction problem if we are not careful. Since this fact of similarity might allow for CAT OR DOG representations of cats and dogs if the distinctions between the contents of the representations are not carefully drawn. And since the 'what they have in common' would be at least part of the content of the DOG representation, cats that token DOG would have, to some degree, whatever it is in common with dogs under these observation conditions. But if this is so, why would the dog representation not be best interpreted as merely having the what is in common as its content and effectively result in (that is to say, reintroduce) a *dog or cat* disjunction as the content of DOG? To be duly considered, certainly, but further discussion of this point will have to wait until Chapter 3.

2.6 Accuracy, Isomorphism, and the Distinction Between Content and Referent

Despite the general acceptance[31] that the content instantiating conditions of the representation relation will involve causation, some theorists are not willing to allow for any causal determinants of meaning. According to Robert Cummins (1996) (who, as mentioned above, lumps these theories together as "Use Theories," because they focus upon the use to which the representation is put, as the fact of reality that determines content, i.e. what the representation is applied to).

According to Cummins (1996), these use theories want to hold on to causation as a component off the representation relation, and account for misrepresentation while maintaining causation in representation (and so do I). But Cummins thinks that in order

[31] Also accepted here.

to properly explain misrepresentation there has to be a complete dissociation of representational content and the "target" of a representation (the thing being represented, or the thing that the representation applies to) and he suggests mental/cognitive functionaries called "intenders" that match representational contents up with represented objects in certain circumstances and particular occasions.[32] For Cummins (1996):

> Targets, then, are determined by the representational function of tokening a representation on a particular occasion in a particular context, not by the content of the representation tokened. In our example, the target on the occasion in question is P2 [the position of some item in the world] regardless of what representation is tokened or what the content of that representation is. 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. (7).

Cummins takes this analysis of error, which holds that in order for error to even be possible, there must be a complete dissociation of representational content and the thing represented (what the representation is applied to). Therefore content cannot be identical to its referent, i.e. what is being represented. Since error surely happens, a dissociation

[32] It is not important for our purposes to go into a great deal of detail about Cummins' account of intenders.

like this will be required of a theory of content in order to account for misrepresentation. These concerns guide Cummins formulation of a theory of mental representation.

Cummins, guided by these considerations pertaining to error, constructs a theory of mental representation in which structures in minds/brains/computers represent other things in virtue of being isomorphic to these other things. So, a mental state represents something else in virtue of being isomorphic to that something else. An isomorphism between two structures is present when every element and relation between the elements of a structure exist in another structure. So, if Structure One contains elements A1 and B1, and there is a relation between them, then there will be present in Structure Two, which is isomorphic to Structure One, elements A2 and B2, and a relation between A2 and B2 that instantiates the same relation that A1 and B1 have in Structure One. For Cummins (1996) a structure represents another structure when it is isomorphic to another structure. As a consequence, two isomorphic structures represent each other, and every isomorphism represents itself.

Making isomorphism the relationship between two structures that instantiates, or underwrites, the representation relation, has the effect that the content of a representation is causally independent of what it represents. There is no causal relationship that instantiates the representation relation, and so what happens to cause the tokening of a representational state has no bearing upon that state's meaning under any conditions. Thus, the content of a representation is not identical to the thing represented, and the representation (the structure) can be applied to things other than what it has been used to represent in the past, i.e. it can be used in referring to things it has not previously represented. And so representational content is effectively dissociated from what it

represents. So, for Cummins, this is the way he defines how something has meaning "*intrinsically,*" which means that a representation has meaning independent of its use, since it is irrelevant to the meaning of the mental state what it is being used by the system to represent (93).

Isomorphism does not only result in the representational content of mental states being independent of the entities they represent, but also introduces a notion of representational accuracy, which will be of great importance in chapter 3. An isomorphism can be more or less accurate, and a representation of something might, while being more or less accurate, still be able to do the job that the representation has been recruited for, despite not being completely isomorphic in some respects. Take for example a map of a city (Cummins is fond of maps). Maps, according to Cummins, have their representational content because they are isomorphic to what they represent, and that it does not matter how it is that the map came to be; so long as it is isomorphic to the city in question, then it represents the city in question. The map of the city can be more or less accurate, and still be used successfully navigate the city.[33] For example, the map could lack a number of alleyways, and put too much relative distance between streets, and as a result the city blocks might be too long relative to other structures on the map, and in these ways be inaccurate with regard to the relations in the actual city. However, this does not change what the map is a representation of, nor whether or not it can be used as a guide to get around the city.[34]

---

[33] This implicates the complexities of *map users* rather than the nature of isomorphism *per se*. I owe this point to Tim Kenyon.

[34] Cummins claims that because maps are isomorphic, they have intrinsic meaning. Though, that is a debatable claim, given that maps are quite conventional symbol laden. See Dretske 1981.

Notions of accuracy, as Cummins (1996, 26) notes, provide us with a general schema for graded notions of representational content, which, it seems (as discussed above (in Section 2.5) in terms of similarity), we very well may need in order to explain how misrepresentation occurs. While the notion of accuracy with regard to mental representation seems to me to be extremely important, and combines nicely with ideas about graded content and the similarity of mental contents mentioned above in the discussion of the disjunction problem, detailed discussion of these themes and how they might work together will have to be put off until Chapter 3, so that some important points can be drawn out about isomorphism with respect to why it is that it will not be adopted here. These reasons for not adopting isomorphism here have to do with the complete causal dissociation of representations and what they represent, and how pervasive the "all or nothing" understanding of representation is.

2.7 Why Isomorphism Will Not Work, and Why We Should Retain Accuracy

As mentioned just above, the context in which accuracy is developed by Cummins (1996), (this context being) isomorphism, has its problems. The most difficult issue for Cummins' account is likely the fact that if an isomorphism between two structures is the relation that determines that one object represents another object, then nearly every object represents every other object, because a well defined isomorphism can exist between almost any two chunks of matter, e.g. between my notebook and the atomic structure of a tombstone in Missouri. The problem then becomes that just about everything represents everything else, and it becomes how a representation is used that determines what the representation is representing. So, we need some account of the use

of the representation that determines, or picks out, the relevant structural similarities between a representation and the represented object, and this, of course, has to be specified in some non-intentional language, and this does not look easy for Cummins (1996), and he explicitly rejects use. Further, this implies that a particular representation/isomorphism could potentially represent any number of things, and so Cummins (1996) also needs an account of the matching of a representation to a particular target, and, of course, does so without relying on representational content.

Since every isomorphism is going to represent any number of things (potentially literally everything), then it is going to be what the system uses it for that is going to determine what the isomorphic structure represents at a given time, and so it will represent, essentially: whatever it can be used to represent, which makes representational content arbitrary, and if we want representations to be explanatory, like explain why it is that my DOG thoughts apply to dogs, or my MY DAD thoughts apply to my dad, then they probably should not be arbitrary, since many of the things that are represented are not arbitrary. Which suggests the problem of specifying what the relevant structural relations that are mapped between isomorphisms will be. If I continually represent wolves as a dogs (that is, subsume wolves under DOGs), because of many, many structural identities, all of which may be relevant for the purposes of representation, but not all of which make it into the "releveant" structural identities, then I may find myself in a quite a bit of trouble quite often regardless of how one construes concepts[35].

[35] Cummins (1996) holds that Concepts are knowledge structures rather than individual representations. I hope to skirt such issues here—hence the focus upon "representations" and "content" and a down playing of concepts, except where the theorists I am discussing use "Concepts" and "Mental Representations" interchangeably.

The thing is that we end up with a problem if we accept Cummins' isomorphism focused account, because representations and the things that they represent become almost completely dissociated[36], and so in order to have error even happen contents and referents have to be independent of each other. This has its problems, particularly those that show up for isomorphism, like the hard questions regarding exactly how it is that the one object represents another in any tractable way arise (as we have seen above). And I think that some of these deep problems arise for isomorphism as a consequence of the notion of error that Cummins (1996) has in mind; namely that representations either apply, or do not, to an object, because, on Cummins' scheme, representations are either useful or not, an so apply to an object or do not. So, a representation applying more-or-less to an object in the world is not employed by Cummins (1996) to address misrepresentation as it might be.[37].

So, trying to make representations completely independent of what they represent seems to defeat the explanations we are trying to construct when we invoke mental representations in the first place. And though Cummins (1996) brings up the notion of accuracy, it does not do the work that it might, because it is employed to answer the traditional understanding of misrepresentation/the disjunction problem, which holds very tightly to an "all or nothing" conception of the application of representations to what is being represented by the representation. I will address how accuracy might be applied to misrepresentation in Chapter 3.

Given the difficulties with Cummins (1996) project, we must ask the following question: What better way to determine what the representation represents (non-

[36] Associated only by isomorphism.

[37] This, as we have seen creates problems for theories of representation that postulate content identity.

arbitrarily) than to have it somewhat determined by what causes it (under certain conditions)? However, we have to, in positing a theory that involves causation be weary of falling prey to the identifying of content with its referent. It seems to me that a notion of accuracy applied to representations ("representational accuracy," perhaps) will assist in keeping us between both ditches (arbitrariness and making error impossible). The task of the next chapter will be to provide arguments for how a theory of mental representation might go about doing this.

Chapter Three:

Gradability, Measurement, and Misrepresentation

3.1 Introducing Measurement

In the previous chapter, we discussed the notion of accuracy with regard to the application of a concept, and so also its content, which is here being understood as a property, or category. So too, in the previous chapter a case was presented for the position that it is reasonable to believe that conceptual contents are graded, and that these contents are distinct from what the concept represents. As a consequence concepts ought not to be understood as merely applying, or not applying, to a particular object, but rather that they should be understood as applying to an object to a certain degree (i.e. graded). It is the task of this chapter to argue further for the position that conceptual content is graded. The argument for graded content will occur on two related fronts. The first front of the argument being that conceptual content ought to be understood as being like measurements as opposed to being like words, and so providing a way of thinking about conceptual content that makes clear its graded nature. This argument will take the following shape: I will make a case for believing that the application of a concept to an object is very much like taking a measurement (say the length of, for example) of some item in the world, in that applications of concepts are more-or-less accurate assignments of properties to objects in the same way that measurements are more-or-less accurate assignments of lengths to objects[38].

[38] Obviously, there are measurements of things other than length, like weight, mass, speed, velocity, etc. But for the purposes of simplicity, and clarity, I am choosing length as my model of measurement.

The second front pursues justification for the position that is that accepting graded content will allow us to solve the disjunction problem, which, as we have seen, content identity cannot seem to solve. However accepting the gradability of conceptual content necessitates the acceptance of a robust notion of content similarity. So an argument for the plausibility of a robust notion of content similarity need to be constructed, given that there exists a well-known argument for the position that content similarity presupposes a robust notion of content identity, and so content identity is the correct theory for conceptual content. The case for the plausibility of content similarity will attempt to show that this argument that any notion of content similarity that we come up with will necessarily presuppose content identity is not as decisive ass its proponents might like it to be. However, if this cannot be done, it seems that we will have no choice but to accept content identity, despite its difficulties.[39]

3.2 Measurement

The focus upon concepts as being like measurements may seem a bit counter-intuitive at first. However, if one considers that the nervous systems is the mind, or at least the central component of the mind, and that mental representations are formed and often tokened by the transduction of energy emanating from other parts of the world that is transduced via the exterior of an organism through the nervous system to the brain where, due to this stimuli, neural structures are shaped, or activated, that correlate in some specific way with these parts of the world, then it should not be difficult to accept, to some degree, that mental states measure the world. This is especially so given that the

[39] Some of which were discussed in the previous chapter.

energy that is transferred can be understood as being some pattern and amount of energy—frequencies, etc. And so the brain and nervous system/mind is operating on the basis of states that have received and have been augmented in certain ways by the amount of energy that was encountered by the nervous system that correlates in relatively specific ways with various parts of the world. Though, it is unlikely that any two such states of correlation will be the same, even if it is the same object correlating on two different occasions with a single specific mental state via some particular energy transmission. Since this is likely to be the case for nervous systems that are different, or, even, simpler than those of human beings, it has the promise to, perhaps, provide an account of representations that will cover more than one species featuring a complex nervous system.

Suppose we consider the relatively basic case of gauging, or measuring the length of some item in the world (let us call the item, item-x) with a stick, or rod, upon which we can mark the length of item-x. We then measure the length of item-x, and we come to a determination of its length upon our measuring rod. We have a measurement of our item's length. But is this *the* length of item-x? Can we reasonably say that this measurement we have taken is identical to the length of the item that we have measured?

In a manner of speaking, yes; the measurement of the length of item-x is the length of item-x, but there is always some degree of error in a measurement, some degree to which the measurement is 'out'. If we were to take another measurement of item-x, even with the same measuring stick, it is quite likely that we would, this time, mark a slightly different length, or with a different measuring stick, mark a length that does not match up with our first measurement. This is similar to (and is indeed modeled upon) a

situation that might occur when comparing rods in an attempt to determine their congruence, if it exists, as described in Krantz, et al. (1971)[40]. The situation runs as follows: Suppose we have three rods, R1, R2, and R3. Suppose further that we compare R2 and R1 and determine that they are congruent, i.e. that we cannot detect a difference in length between them. Suppose even further that we find the same to hold between R1 and R3. However, when we compare R2 and R3 we find that R3 is longer than R2, i.e. that they are not congruent. So, despite our observation that both R2 and R3 are the length of R1, they are observed to not be the same length as each other. The point is that measurement taking, rarely, if ever, gets things absolutely right. Krantz, et al. (1971) on the matter:

> Of course, really perfect copies cannot be prepared. Whenever physical differences become sufficiently small, any method for observing them ultimately deteriorates. In some cases, and perhaps in all, observations of two sufficiently similar entities are inconsistent when the same comparison is repeated several times. And when inconsistencies can occur, violations of transitivity may arise. (3).

Error cannot be wholly avoided in measurement.

This, I would venture to suggest, is very similar to how mental representation works. We can agree that two individuals have concepts with representational content regarding a particular thing/object (referent), and that both concepts are related via the representation relation, as the rods (R2 and R3) above are related to R1 via the measurement relation, but the contents of the two concepts, like the lengths of R2 and R3

[40] Page 3.

are not identical. The length of R2 and R3 are each similar, close in length, and bear the appropriate relation to the rod against which they were each deemed congruent.[41] Similarly with the content of the concepts and the things in the world they represent; the content of my concept of *x* and another's concept of *x* are both of the same thing, but their contents are not each the thing that they are a representation of. The concepts are both related to what they represent, or refer to, in much the same way that a measurement provides us with the length of a particular item, but the length of the measurement is almost certainly not identical to the length of the item measured, simply given the margin of error that comes along with any physical process like that of taking a measurement of some item in the world. Given that cognitive/nervous systems such as ours have been cobbled together in a slightly haphazard manner by evolutionary processes, perhaps it is not too bold to venture saying that little mistakes in, and differences among, mental representations of things in the world abound like those that do in measurement.

The question of "how strongly do I intend this analogy to be taken?" may occur to the reader at this point. So, I will address this question now. The short answer is: quite strongly. This strongly in fact: given that actual amounts of energy cause neural states to have specific responses, and structure neurons in quite specific ways, these energy affected neural states will carry, quite literally, a mark of how much energy there was of a particular kind interacting with a particular sense organ at a particular time. Imagine it being like the way a beach bears the markings of high tide. Since we accept, most likely, that neural states are causally affected by the transfer of energy from the environment via the surface of the organism, and that this energy is transduced to the brain's neural

[41] This rod against which both measurements are taken is assumed to not have changed in length over the course of the two measurements, so that the length of the measured rod is identical during each measurement.

structures, and that energy and its transfer occurs in amounts (of energy), it is reasonable to hold that the most basic representational neural structures will be best understood in terms of amounts like m/s, kg, etc.. If this is the case, then more complex representational structures of more robust properties, like DOG, will be constructed from these more basic units. Such units would also enable our naturalistic theories of representational content to be quite consistent with the units postulated by our other physical theories, like the natural sciences[42]. The upshot is that the most basic representational states will be recordings of amounts (of energy), which would make them, quite literally, measurements. So, more complex representations will be constructed from measurements, and as such, these more complex representations will be interpretable as conglomerate measurements

The remainder of this chapter will be concerned with the detailing of and the making a case for the reasonableness of this position, especially with regard to the disjunction problem, and will rely upon material presented in Eliasmith (2005).

3.3 The Application of Measurement to The Disjunction Problem

There is a measurement focused solution to the disjunction problem (misrepresentation) offered by Eliasmith (2005) involving the use of elementary notions in measurement theory, emphasizing the idea that it is unhelpful to assume that the application of a concept to an entity must be either right, or wrong, absolutely, as content identity theorists do. Because even if we were to accept the absolute rightness or wrongness of the application of a concept, such a view may still be a non-starter, since

[42] See Eliasmith 2005 pages 1041-1042 for a more detailed discussion of the interpretation of basic neural representations as m/s, kg, etc. and this being consistent with the natural sciences.

such stringent conditions on representation would make matters such that it would not be completely clear that we ever successfully represent at all.

Why might this be the case? Misrepresentation, is, as discussed in the previous chapter, to "say of something that does not have a given property, that it has that property; for example, saying that a black dog is brown. In other words, misrepresentation is representing one thing (a black dog) as another (a brown dog)" (Eliasmith 2005, 1050)[43]. Consider the following case of what might be called misrepresentation. Suppose one is seated in a sparsely furnished and well-lit room, under standard conditions of observation, and a black dog is placed before one for a duration of three minutes. At the end of this three-minute interval, the dog is removed from one's view. Following the canine's removal one is asked a series of friendly, but specific, questions: "What color was the dog? Answer: Black. Dark or light black? Answer: Dark Black. This color (showing a color swatch)? Or this color (another swatch)?" (Eliasmith 2005, 1050) and so on.

It is a virtual certainly that one would eventually answer incorrectly—that is, ascribe to the dog a shade of black that it is not, and so, misrepresent the dog with regard to its colour, according to the standard notion of misrepresentation, as being a shade of black that the dog is not. While, technically, in the content identity, all or nothing interpretation of misrepresentation, this would be a misrepresentation of the dog. But is this the appropriate approach to take with regard to misrepresentation, given that viewing this case of the black dog as a misrepresentation would almost certainly mean that successful representation would be a very rare phenomenon?

[43] Notice how pervasive understanding mental representation as being language-like actually is, even those theorists, like Eliasmith 2005, who believe that thinking about mental representation linguistically is a mistake, still use terminology consistent with it.

Perhaps then we should reconsider what we take to constitute successful cases of representation versus cases of misrepresentation. Eliasmith (2005) asks us to consider the following basic notions in measurement theory while in search of a solution to this problem.

> Measurements are said to be *accurate* if they are near the right value. If I measure darkness and get the right answer, I have made an accurate measurement. Measurements are said to be *precise* if they are reproducible. If I measure the darkness of a color swatch over and over again, and get the same answer every time, I am making precise measurements of darkness. If my measurements are precise and accurate, they are said to be *exact*. Notably, precision is a property of a set of measurements, while accuracy is a property of a single measurement. But we can define the accuracy of a *set* of measurements as the *average* nearness to the right value. In statistical terms, precision is measured by the variance of a set of measurements, while accuracy is the difference between the average measurement and the correct answer. (Eliasmith 2005, 1051).

The "right value" here is the actual length, or amount of whatever it is that is being measured in the world. Notice how this is consistent with our discussion of measurement in the previous section having to do with rods R1, R2, and R3, and with the case of multiple measurements of the same item with the same resources might result in different answers to the question of "How long is item-x?"

The interrogation considered above can be interpreted as an uncovering of the precision of the subject's representation of the dog's colour. "Although the representer may be perfectly accurate [we should probably say "highly"] at one degree of precision

(black versus white) the representer may be inaccurate at another (one color swatch versus another" (Eliasmith 2005, 1051). The basic point to take home is that the notion that "representations are best characterized as 'better or worse,' not as 'right or wrong'" (Eliasmith 2005, 1051), where 'better' indicates a high degree of accuracy and 'worse' indicates a low degree of accuracy. Further, the more often a representation ascribes a highly accurate property to an object (or set of objects) being represented, the better the representation will be. This is because the ascription will not only be accurate on a particular occasion, but accurate often, and so will be useful for a broader range of cognitive operations than if it were the case that the representation were only accurate once.

If one would still like to hold on to some absolute criterion of conceptual content and applicability, then one could attempt to 'peg' a degree of accuracy at which point one has correctly, or rightly, represented something, and where a level of accuracy below this point would constitute getting it wrong, or misrepresenting the item in question. However, it is not clear how this would proceed in a principled manner such that sorites arguments like the following could be avoided. If .5 degree of accuracy (between 0 and 1), say, were to count as a correct representation and .49999… would be insufficient for representing correctly, why would .49999… not be sufficient, since the difference in content (and value) between .5 and .49999… is negligible, and so likely cannot constitute a significant difference in amount of content, and so on all the way down to the number 'above' Zero.[44] So, to follow Eliasmith (2005) a little bit further:

[44] Holistic theories of content face a similar problem. See the discussion in Chapter 2 above.

> Using the term "misrepresentation" to divide representations into two groups obscures important subtleties of representation. We will have a more general understanding of misrepresentation (i.e. one that does not depend on choosing particular standards and can account for degrees of deviation from any given standard), if we accept that representations come in degrees, i.e., that they lie on a continuum from good to bad. (1051.)

Of course, in the matter of representation and misrepresentation, our standards from which representational contents vary by degrees (i.e. are gradable) is going to have to be the actual things in the world to which we ascribe properties via the application of a concept.

Further, it is likely that the standard interpretation of successful representation and misrepresentation that a representation applies to an object or it does not[45] will be insufficient, as it may simply not be possible to get certain parts of our world that we do perceive, and represent, absolutely correct, given that there are physical limits to what sort of information/data from the environment can be represented by a system at all times, or at all. Lessons learned from psychophysics tell us that there are specific ranges in brightness that our retinal cells can encode. "Outside of those ranges, differences in intensity, are indistinguishable; that is a fact about physiology. If the vehicles cannot carry such differences in intensity, then those differences cannot be used by the [cognitive] system to react to the environment; consequently, those vehicles cannot carry content about those differences" (Eliasmith 2005, 1049). We cannot represent certain things—differences in certain levels of brightness in the preceding example, and very

[45] These are examples of theorists who hold this view: Fodor 1987; Fodor 1998; Dretske 1981; Millikan 1989; Sterelny 1991.

likely we cannot completely accurately represent features of our environment. So, there are always going to be limits to the representational accuracy of a cognitive system's representational states, and as such, it ought to prompt us to question the plausibility of content identity, and the interpretation of misrepresentation that it forces upon us[46], if we accept it. If we accept this view we will very often, if not always, be misrepresenting the brightness of our surroundings, rather than representing said brightness to a particular degree of accuracy. This deficiency of our representational machinery makes itself apparent also with regard to how much information neurons can encode—there does seem to be an upper limit to how much information about a subject neural structures can carry, given that "only about three bits of information are transmitted per [neural] spike" (Eliasmith 2005, 1051).

Robert Cummins (1996) makes much the same point, making it clear that incorrect misrepresentation is likely to abound with regard to certain elements of the world, because in many instances we do not possess the correct physical equipment to represent those things we are misrepresenting. Cummins calls these sorts of errors, "Forced Errors," because they are in a sense forced, or happen because the nature of what is being represented is such that it cannot be captured due to the impossibility of it being captured by the system in question. The system doing the representing simply lacks the requisite expressive power to handle certain potential contents (Cummins 1996, 23).

The approach to understanding misrepresentation promoted in this section provides us with what is very likely a more plausible view of the nature of misrepresentation. This approach to misrepresentation also provides us with a way to

[46] Discussed above.

solving the problems raised by misrepresentation and the disjunction problem when we emphasize the idea that concepts have graded content. It is the project of the next section to see how this understanding of misrepresentation and the postulation of graded conceptual content solve the difficulties misrepresentation raises.

3.4 Gradability

Eliasmith (2005), unradically, understands the content of a concepts as being, most basically, and generally, categories (1037), that is, when a concept is applied to an object, this procedure places the object in a particular category. Another way of conceptualizing this point is that the application of a concept to some entity is to ascribe a property, or set of properties, to the entity; "the meaning of a neural representation is the set of properties ascribed to something by that representation" (1039). This is how I am understanding the content of concepts. This is also the way Fodor (1998) understands concepts: "To say that concepts are categories is to say that they apply to things in the world; things in the world 'fall under them'" (24). So, I am not working on ground completely uncommon to the theorists whose views I am arguing against.

We all also take the referents of our tokened representational states, under content instantiating conditions, to be causes, in that their reflection, or emission, of some form of energy interacting with our nervous/cognitive systems results in the tokening of concepts with a particular content (almost always a property the referent has). Though, I think that in order to solve the disjunction problem, a graded notion of content needs to be maintained, and a graded theory of content necessitates a distinction between content and referent, because on a graded view of content, content and referent cannot be identical.

Fodor, on the other hand, as we have seen above, maintains that content and referent must be identical, and we have seen what sort of difficulties this leads to with regard to accounting for misrepresentation.

So, here I will briefly adopt the theory of content introduced by Eliasmith (2005) in order to show, in some detail, how understanding representational content as being graded, i.e. as something that can be more or less accurate than another representation, and so distinct from its referent, can plausibly solve the disjunction problem. As mentioned above, this theory will involve causation, but mere causation is not enough to account for representation, because if it were, sunburn would represent the sun, which it does not. So, Eliasmith (2005) adds the further element of statistical dependence to flesh out the representation relation.

The responses of neural populations over all stimulus conditions, and the regularities in the response of neural populations to stimuli over all stimulus conditions will determine the referent of a concept. The regularities in the properties/property attributed to an object, or set of objects, by the neural structures that are activated, across all stimulus conditions. The information, or data, extracted by the decoder[47] that predicts, or picks out, a property consistently in the stimulating object(s) over all stimulation conditions, is the content of the concept. But what about the individual referent that the concept is applied to? And what is produced when presented with a particular referent under particular stimulus conditions, as it relates to the content of the concept?

[47] Eliasmith 2005 provides a detailed account of how the representational content of neural populations is to be extracted from their firing patterns, which involves decoding the information that the neurons encode, in order to determine the content of the content of the neural states. See pages 1040-1043.

Well, this way of understanding matters results in two kinds of content: "Occurrent" content and "Conceptual" content. Occurrent content being the information encoded (and decoded) in a neural population with respect to a specific referent (object) at a specific time, under specific stimulus conditions, and conceptual content is the determination of the property ascribed (the content) to all objects with a particular property over all stimulus conditions. This results in the following fact about occurrent content. "The occurrent reference of a vehicle is the set of causes that has the highest statistical dependence with the neural responses under the stimulus conditions in which it occurs." (1048). This has the consequence that the determination of content (the application of the decoder) stays the same for both occurrent and conceptual content, but the determining of a referent will vary between occurrent and conceptual content.

Since occurrent cases are dealing with specific referents under specific stimulus conditions, this indicates a dissociation of content and referent, or the content of a representational state/concept and the object property it represents. And hopefully it is clear that the statistical dependence that holds between a neural state and all of its stimulus conditions can be quite different from the statistical dependence that holds between a neural state and a specific subset of the set of all stimulus conditions. Thus, this theory tells us that differences between statistical dependencies in particular instances of particular stimulus conditions indicate a difference in, or uniqueness of, referent. However, we need to keep in mind that the properties ascribed under a particular set of stimulus conditions (occurrent content) is still going to "determined by property ascriptions under all stimulus conditions (conceptual content)" (1048).

But how does this all come contribute to solving the disjunction problem, and accounting for misrepresentation? And how is it the case that this understanding of content is graded?

3.5 How Gradability Solves The Disjunction Problem

Recall the above discussions of accuracy and graded content (in Chapter 2 and the previous section) in which it is suggested that representations are things that are graded, and so are more or less accurate rather than possessing a nature such that a representation is either right or wrong, correct or incorrect, will explain the phenomena of misrepresentation and solve the disjunction problem by allowing that we can ascribe properties to objects, or entities with varying degrees of accuracy without likely ever getting any attribution of a representational content to some object in the world 100% correct. How does the theory of conceptual content that we have just examined allow us to account for misrepresentation in this way?

Let us take once more our previously considered case of misrepresentation—the cat that tokens DOG. In low light, or otherwise non-optimal conditions, something with four legs moves in the distance. My DOG representation is tokened in response to the critter that I have perceived. However, it is a cat, not a dog. Here we have a case, yet again, of my misrepresenting a cat as a dog; my CAT should have been tokened, because what I beheld was a cat, but my DOG was tokened instead. I have attributed the property of being a dog to the object, rather than the property of being a cat, which would have been correct given that what I saw was, in fact, a cat. So how, on a causal story, do we explain how my DOG means *dog* and not *dog or cat in low light or in the fog* and so on?

We can employ the notion of statistical dependence to explain misrepresentation, and account for the disjunction problem. This cat has a high statistical dependency, the highest statistical dependence in fact, with my DOG under these specific stimulus conditions. This is to say that, this cat, under these particular stimulus conditions, has a high statistical dependence with my concept/representational state that under all stimulus condition has the highest statistical dependence with dogs, which is my concept/representational state DOG.

Thus the occurrent content of the representation of the object in my perceptual field (the cat) is DOG, while the conceptual content of the representation is DOG as well, because under all stimulus conditions it is a neural state that has the highest statistical dependence with dogs (and this is what determines content). The cat, under these stimulus conditions, has a higher statistical dependence with my DOG concept than with my CAT concept. However, under all stimulus conditions the cat will have the highest statistical dependence with CAT. So, we have an explanation of misrepresentation: the cat, which would under other stimulus conditions (indeed most) have the highest statistical dependence with CAT, has, for whatever reasons, a higher statistical dependence in this instance with DOG. DOG has its highest statistical dependence with dogs, but on this occasion, has its highest statistical dependence with what happens to be a cat. While the DOG would, under most stimulus conditions, have the highest statistical dependence with dogs, it does not in this particular case. Thus we have a case of misrepresentation. And we have an explanation of misrepresentation. We have the reasons for why the content of the state tokened is DOG, and we have an explanation of why, on this occasion, a cat caused the tokening of DOG.

However, we and Eliasmith might ask,

> does this really solve the *disjunction* problem? Will it not be the case that the highest statistical dependency holds between dog-or-this-cat under all stimulus conditions? In fact, no. This cat under *all* stimulus conditions will not have a high statistical dependency with my "dog" vehicle; it will have the highest dependency with my "cat" vehicle. It is only under *this* stimulus condition that it has a high statistical dependency with my "dog" vehicle. In other words, because there is another vehicle (the "cat" vehicle) that has a higher dependency with this referent under all stimulus conditions, it cannot be this vehicle (the "dog" vehicle) that has this cat as its referent. *Note also that this solution is possible because the notion of representation/misrepresentation is a graded one*. (1051-1052, emphasis added.)

But how is it that content is graded, and how does this enable us to solve the disjunction problem?

So, how is it that statistical dependence is an instance of graded content? The answer to this question is as follows. The DOG states of two distinct cognitive systems, say my DOG and your DOG, will likely not have the same statistical dependence with dogs under all stimulus conditions. This is because we have different brains, etc. and as such our states might vary with regard to how high the statistical dependency of our DOG states is with dogs. Your DOG could have a higher statistical dependence with dogs than mine, and so we would have different amounts of statistical dependence with regard to our dog-DOG relationships. However, my DOG and your DOG would still have the highest statistical dependence with dogs among our own mental states, and as such will

each pick out dogs appropriately, but statistical dependence of our DOGs will differ from each other to some degree. And this plays out in the same way that two measurements of the length of the same item (or two rods compared against another rod with regard to establishing their congruence) will differ to some degree with regard to their length, despite having been taken from, or derived from, the same item, which has the same length/value when each measurement is taken. The dogs will still be dogs, regardless of how much difference there is between how high my DOG's statistical dependence is with dogs and how high your DOG's statistical dependence is with dogs. And dogs will still be dogs regardless of how high the statistical dependence is between my DOG and dogs and how high the statistical dependence of your DOG and dogs.

Further, the DOGs of different neural/cognitive systems will have differing degrees of statistical dependence with dogs. Given that each neural/cognitive system is distinct, and is "hooked up" to the world in its own way, it is fairly likely that a dog, or a set of dogs, will have a higher statistical dependence with one person's (or cognitive system's) DOG, than it might with another person's (or cognitive system's) DOG. Your dog-DOG statistical dependence over all stimulus conditions might well be higher than mine. However, both of our DOG states would still have the highest statistical dependence within our own cognitive systems with dogs. As a consequence, your representation would be "more doggish" than mine; that is, have content that was more consistent and better correlated with dogs than mine would be. Your DOG would likely be tokened with a higher degree of accuracy, and with greater precision than mine. Also, some dogs are going to have a higher statistical dependence with one's DOG state than other dogs will. So, some dogs will have the property of dogness assigned to them with

more accuracy, and greater precision, than other dogs. This should not be a surprise, since the highest statistical dependency over all stimulus conditions between a set of objects with a particular property and the mental state that assigns that property to those objects is going to be something like an average of the highest statistical dependencies between said mental state and said objects over every individual set of stimulus conditions. So, there are going to be entities in the set of objects with a particular property (that the highest statistical dependence is determined over), dogs say, that are going to have a lower statistical dependency with the mental state, DOG, than other objects in the set, other dogs.[48]

Since the statistical dependency of dogs and DOG will vary (depending on the dog), even within a single cognitive system, it is going to be the case that dogs, even when they have their highest statistical dependency with DOG, are going to have lesser degrees, or amounts, of statistical dependence with other concepts, such as CAT, FOX, TREE STUMP, etc. (I would venture a guess that dogs will have some degree of statistical dependence with, at least, everything that a cognitive system could misrepresent a dog as). And it will be the case that dogs (as well as other entities) will have varying degrees of statistical dependence with those concepts with which they still have some statistical dependence (CAT, MAT, RAT, etc.), but that are not the concept with which it has the highest statistical dependence over all stimulus conditions (DOG).

Given these considerations, it seems then that all entities and properties that a neural/cognitive system is going to be capable of representing are going to have this be a

[48] However, this member of the set (a dog) with a lower statistical dependence than another member of the set (a second dog) is going to still have a statistical dependence with the mental state (DOG) that is higher than that possessed any other object (cat, chair, etc.) with a different property--the property not assigned by the mental state in question (DOG).

fact about them. Just about every, if not every, concept is going to have some statistical dependence with nearly every property in the world that it is capable of representing. However, it will be only the highest statistical dependencies between a property possessed by an object in the world and the property assigned to this object by a particular mental state that will determine the content of said mental state. This content, as mentioned above, will be the property that the mental state assigns to an object (that has that property, or something very close to it) when tokened in response to the object with that property.

This will all even be the case with cats. So, we can now return to our favoured example of a cat being (mis)represented as a dog for the purposes of wrapping up this discussion. In light of what has just been discussed, a cat, under certain stimulus conditions (those that would make it more doggish, whatever those would be), could quite clearly cause the tokening of a DOG. And further, given the variation in degrees to which statistical dependence can differ with respect to the relation between dog and DOG[49], and even cat and DOG, it would not be surprising that some cat under certain stimulus conditions would have a higher statistical dependence with DOG than with CAT. Under this particular set of stimulus conditions that results in a cat causing DOG, there would still be the dog-DOG highest statistical dependence over all stimulus conditions relation, but on this occasion[50] it could be the case that the cat-DOG statistical dependence relation would be higher than the cat-CAT relation. And this would be

[49] i.e. That there are degrees of statistical dependence between dogs and DOG, and so that statistical dependence is graded, and since statistical dependence determines content, content is graded.
[50] Under these stimulus conditions.

because the content of concepts is graded, i.e. that conceptual content comes in degrees of statistical dependency.

Due to the fact that content is graded in this way, we can see how graded content allows us to solve the disjunction problem, and even adhere to the framework Fodor provides with his theory of Asymmetric Dependence. The cat-DOG relationship, an instance of high statistical dependency between cat and DOG under these specific stimulus conditions is going to be dependent upon the dog-DOG relationship, which is a high statistical dependence between dogs and DOGs over all stimulus conditions. And if it were not the case that the dog-DOG relationship existed, providing DOG with its content in the particular way that it does: via the highest statistical dependence between dog and DOG under all stimulus conditions, it would not be possible for cat to token DOG in this one instance, because DOG would not exist[51].

However, all is not as neatly put away as we might like. A theory of conceptual content like this is going to invoke, or rely upon, a notion of content similarity (as this account has done, and so is going to face the objection, voiced by Jerry Fodor (1998),[52] that any notion of content similarity is going to presuppose a robust notion of content identity, and so a similarity of content theory of meaning like we have above, is a non-starter. In the next section (or two) we will discuss both Fodor's argument, and why it is not as effective, or certain, as Fodor would probably like.

[51] Please note the fact that despite this account fitting the formal structure of Fodor's Asymmetric Dependence, it is not the case that Fodor, or any other content identity theorists, can accept this solution. This is because it requires a distinction between content and referent, as well as employing a notion of content similarity among concepts of the same thing in different cognitive systems, rather than identity.
[52] Also in Fodor and Lepore 1992, 1999.

3.6 The Similarity of Content Presupposes The Identity of Content

As mentioned above, Jerry Fodor (1998) offers a strongly stated argument for the position that any theory of content that proposes that the content among our "shared" concepts (my DOG, your DOG, Bill's DOG. . .) is similar, rather than identical, presupposes a robust content identity, and so does not, and indeed cannot, get off the ground, and we are left with only one possibility—that of content identity. Much of the motivation regarding Fodor's arguments for why it is that any robust notion of content similarity presupposes a robust notion of content identity has to do with the idea that concepts must be public. Concepts must be things that more than one individual can possess at a given time, and that it is the same concept that is shared by every person with a given concept, say WATER. So, you and I have the same concept of WATER, and we must, because if we did not, then we would not mean the same thing by the word 'water', since the word 'water' gets its content from the concept WATER. And only one of us would be able to have WATER thoughts, as everyone else's WATER thoughts would not be WATER thoughts, but something else thoughts. Fodor labels the notion that individuals have their own different, non-identical, concepts of DOG, WATER, etc. "Conceptual Relativism," and he claims, conceptual relativism leads to some very serious problems if you want to take intentional generalizations and explanations of behaviour seriously. So, we are not able, and in principle cannot, employ intentional explanations of the behaviour of individuals generally. And so Intentional ascriptions are not *ceteris paribus* laws, and so there is no hope for an intentional psychology (Fodor 1998, 29).[53]

[53] This point was discussed in Chapter 1 with regard to propositional content, out of which the accounts of most theories we are considering are built.

Similarity of content is invoked in order to allow for people to think about the same things, and talk to each other about the same things, but without their concepts that are about the same things/properties having to have content that is identical. This preserves, or, at least, its proponents would like it to preserve, the generalizability of intentional psychological explanations. We have the "same concept," or at least, concepts that are similar enough to each other such that they can be generalized over for explanatory purposes; the content of my representation of x and your representation of x are close enough alike to allow us to be thinking about, and talking about the same thing without having the exact same content.

Fodor tells us that notions of content similarity are usually not very well explicated and tend not to provide much by way of clear criteria for what this similarity between our concepts will consist in. Fodor also tells us that this is so because it is extremely difficult, nigh impossible, to develop a theory of content similarity that is not dependent upon an assumed notion of content identity. How, then, might this be so?

The basic argument runs as follows. Those who claim a similarity analysis of my concept of *x* and your concept of *x*, always provide the analysis in terms of there being a mapping between our respective concepts of *x*, and while my concept of *x* is different from your concept of *x* in certain ways (this is what motivates the idea that concepts are things that are similar to each other, rather than identical), there are many parts, or aspects, that will be the same. And if they are the same, literally the same, i.e. identical, then anyone peddling similarity of concepts is required to peddle the identity of concepts first. So, similarity does not, and cannot, do any real work, because it requires identity of content, which is the very notion that it was brought in to replace, which is a problem.

For an example of how this argument works in practice let us consider the following. Someone in the business of holistic theories of meaning, who believes that the contents of concepts and beliefs are determined by the other concepts and beliefs that one needs in order to have the first mentioned concepts and beliefs, because the second set are partly, if not wholly constitutive of the first, would say of my concept of George Washington, and your concept of George Washington, that they are similar insofar as a number of our other concepts and beliefs constitutively related to our respective concepts are the same. For example: you and I both believe that *he cut down the Cherry Tree*, that *he was the first president of the United States*, and so on, which will require us to share concepts like CHERRY, TREE, PRESIDENT, etc.[54] Though, of course, there will be differences on a similarity account of our similar concepts, otherwise why posit similarities? You may believe that *George Washington wore wooden dentures*, while I may not. But so long as there are a substantial number of constitutive concepts and beliefs that we share, we have for all intents and purposes, similar enough concepts to not suffer completely incommensurable minds. Thus intentional psychological generalizability is saved. However, such an explanation of content similarity presupposes a robust notion of concept identity. Because it is the same (i.e. identical) concepts, the same (i.e. identical) beliefs, that constitute our similar concepts and beliefs. So, why bother with similarity, and not just accept identity as the basic "sameness" among our concepts?

[54] You will notice that this runs us into similarity all the way down, why would our CHERRY, TREE, and PRESIDENT concepts be identical. Would they not also be similar and not identical on such a view, and then we must go through and catalogue the differences and identities in these constituent concepts, that are in turn constituted by other concepts and so on to the point where our minds are incommensurable. Suffice it to say that holistic theories of content are not without their serious difficulties. But for the argument being considered in the main text we do not need to go into this.

Similar problems also arise for other versions of concept similarity, particularly those that focus on features. On such a theory we might understand the similarity of conceptual contents in terms of overlapping features among the sets of features that make up each of our DOG (say) concepts. My DOG concept might contain the feature tail (among many others, such as furry, quadrupedal, etc.), as would your concept of DOG, but my concept might also contain the feature wet nose, while yours may not. So, our concepts, on such an account would be similar, but not identical. However, according to Fodor the very same problem holism has with regard to similarity rears its head again. The tail feature that is a member of the set of features that makes up my DOG and the tail feature that is member of the set of features that makes up your DOG will be identical (and so will be the case for any common feature our two sets have) in order that they overlap and can be called similar (Fodor 1998, 33).

The most important thing to draw from this is that the conclusion is always that the notion of content similarity necessarily presupposes content identity.
By identity, Fodor means *identity*, in the sense that the contents of our DOG representations are numerically identical. Fodor invokes Leibniz's law to make just this point.

> It looks as though a robust notion of content similarity *can't but* presuppose a correspondingly robust notion of content identity. Notice that this situation is not symmetrical; the notion of content identity doesn't require a prior notion of content similarity. Leibniz's Law tells us what it is for the contents of concepts to be identical; Leibniz's Law tells us what it is for *anything* to be identical. (Fodor 1998, 32).

Well, Leibniz's Law, also known as the "Identity of Indiscernibles" and the "indiscernibility of identicals," provides the conditions for what it is for an individual to be identical to itself. The law does this by way of a biconditional, which I will split into its two conditional components. The first part, the "Identity of Indiscernibles": if two individuals have all of their properties in common, then the two individuals are the same individual. The second part, the "Indiscernibility of Identicals": If two individuals are the same individual, then the two individuals will have all of their properties in common. Being an individual subject to this biconditional is also known as being "numerically identical," that is being the same individual as oneself. Thus, identity is a relationship between an individual and itself, and only an individual and itself. And Fodor holds that your concept of DOG, and my concept of DOG have identical contents, that the contents of our concepts of DOG are one and the same, being the set of all dogs, or the property of dogness. My concept DOG and your concept DOG are tokens, individual instantiations of "Literally the same concept type" (Fodor 1998, 28).

I ask that the reader keep in mind Fodor's insistence upon content identity, because it will be the primary focus of the next section. In the next section the case will be made that insisting upon content identity results in some very serious difficulties for any theory of content identity. Further, it will be argued that it is not so obvious that similarity strictly depends on identity in some way such that identity escapes being dependent upon similarity.

3.7 Arguments in Defense of The Similarity of Content

In this section we will examine two related problems that arise for theories of content that want to postulate content identity as basic, and that content similarity as being dependent upon content identity, where content identity is the idea that the content of a concept is identical to its referent, and it is therefore the case that concepts of the same things in various people have identical content—the referent. These problems with content identity regard its dragging into matters Platonic Forms will be discussed first—followed by an argument questioning the correctness of the assumption that similarity presupposes, and depends upon identity.

3.7.1 The Identity of Content brings with it The Platonic Forms

Fodor (1998) really means identity, that is, numerical identity. He even invokes Leibniz' Law to make the point. This means that peoples' concepts are *literally* the same, not the same in the manner of 'you have the same car I do.' This means that when we are discussing a property that a particular concept represents, say dogness, in the case of DOG, that the content of DOG is dogness. So, the content of every correct instance of a tokening of DOG is an identification of the property of dogness in the object that causes the tokening, otherwise known as the referent. This means that the referent possesses the property of dogness. This, then, has the consequence that every dog possesses the identical property of dogness, and that this property is the numerically identical property of dogness. The reader may be able to see where this argument is heading. It is heading in the direction of Platonic Forms. Here is why: If every dog is in possession of the same property of dogness, as this must be the case in order for all of our DOG concepts

to have identical content, then each dog has the numerically identical property of dogness. This then means that there is some property, dogness, in which every item that is a dog shares, or "participates," and this is, in broad outline, the Platonic theory of Forms. As the reader is no doubt aware, the Platonic theory of forms is not without its difficulties, particularly with regard to explicating how particulars (tokens) share in, or participate in, universals (types).

Further, one cannot remove oneself from the grip of this problem by claiming that what we have is an instance of individual tokens of dogness (individual entities that possess dogness) of the numerically identical (and one and only) property type, dogness. This will not work, because the theory of Forms is an attempted solution to the problem of exactly how it is that we have types and tokens, or universals and particulars, in the first place. Generally speaking, invoking the phenomena that the theory you wish to reject is attempting to explain as a way to avoid the problems that result from the rest of your theoretical endeavors is not an effective strategy. Recall that Fodor (1998, 28) quoted above invokes the type/token distinction, but he does not hazard a guess as to the nature of the type/token relationship. It would not surprise me if this were why.

It should be noted that explanations of issue of the relationship between the universal and the particular have been offered in a more Aristotelian vein. On such a view universals, or types, or general terms, are abstractions created by the minds of people. These abstractions take the form of groups generated from a number of particulars that have seemed to the mind appropriate to group together, because some aspect of each of their natures has been judged to be common enough to be a suitable way to categorize the individuals as belonging to this mind-created grouping. Though,

this would seem to proceed on the basis of the similarities among the various features of the individuals grouped together, and could not be accepted by an identity theorist. This approach, unfortunately, is not without its problems. The argument that I am attempting to undermine reemerges. What makes it the case that these judged-to-exist-similarities are not dependent upon real identities of features between the individuals?

3.7.2 The Identity of Content Presupposes The Similarity of Content

Does it not seem then that similarity is always going to be dependent upon some notion of identity that underwrites the similarity relation? Are we not always faced with the same difficulty? As Fodor puts it, "It looks as though a robust notion of content similarity *can't but* presuppose a correspondingly robust notion of content identity. Notice that this situation is not symmetrical; the notion of content identity doesn't require a prior notion of content similarity." (1998, 32).

This problematic, I believe, needs to be turned around in order to be properly dealt with. What is it that is possibly identical between two distinct, but similar entities, such that what is identical could underwrite similarity? If we are to understand identity as numerical identity (as Fodor clearly does), then nothing can do the job, because numerical identity is a relation between an entity, or element, or whatever, and itself. As soon as one introduces two distinct entities, one gives up numerical identity. So, what sort of identity are we to rely upon for the determination of similarity? And further, what is going to be identical between two entities, or representations, or anything? No two entities are going to have much that will be identical, whatever that is, between them. What would be identical? Relative positions? Line segments? Points? Points of colour?

If we have to whittle down what is identical between entities, and mental states, to such small and fragmented components, then what value could identity have for a theory of content?—Even with regard to something like dogness, what is dogness? What is any such property? Some set of further properties that can be identical to the properties of other things that are dogs that then constitute doghood? Do these further properties continue to break down as well? Again, we are left, it would seem, with small fragments of identical elements, and it is unclear how this would allow for the construction of a robust, or useful, theory of content.[55]

But how is it that these distinct elements, or entities are identical anyway? I would like to label this problem "the problem of identifying identity relations between two entities." It seems to me that any judgment of identity that is not based upon numerical identity, is going to have to be dependent upon some criteria that establishes the identity of distinct entities. And the further this seems to go, the more and more it begins to seem that any notion of identity between two entities that are not numerically identical is going to run aground on the question of: so what makes these distinct entities identical? The answer does not seem to be forthcoming,[56] and it looks as though, any notion of inter-entity identity is going to depend on something very much like Platonic Forms (and we are familiar with those difficulties already). So, it seems that any judgment, or account, of inter entity identity is going to have to rely upon some basic

[55] A content identity theorist might object by claiming that a property is an abstract entity. But serious difficulties still arise. How would a neurological/cognitive system have a naturalistic relationship with an abstract entity? If the answer is: via the particular instances of the property in particular individuals, then we still need some account of how the abstract property is instantiated in each individual that possesses it, and why it would be the case that every particular instance of the property is identical to every other individual instance of it. The problems discussed in the previous section return.

[56] See Deutsch 2002 and Noonan 2006.

notion of similarity, or at least some judgment of similarity. For example: suppose that we want to say, or believe that item x is identical to item y, with respect to property z, because the properties z_x and z_y, are very much alike, i.e. similar, and we have no conditions for identity between z_x and z_y, because they are two distinct things. The thrust of this is that it is not unreasonable to believe that similarity, and identity between entities, are interdependent.

3.8 Conclusion

It would seem that we have reached an end. The linguistic approach to the representational content of concepts, and the notions of content identity it inspires, run into what appear to be, at the moment, near insurmountable difficulties. This is especially so with regard to the matter of misrepresentation, since it makes misrepresentation close to impossible to account for. This is a serious problem for any theory of mental representation, because misrepresentation is obviously ubiquitous in cognitive systems, and so any theory of mental representation that is going to be anywhere near acceptable is going to have to account for misrepresentation. Further, in order to account for misrepresentation, and solve the "disjunction problem" that comes along with it, a causal theory of content seems to require a graded notion of content. And this leads, I have argued, to a need to model mental representation on the workings of measurement, as opposed to focusing upon the workings of human natural languages, which aside from inspiring faulty intuitions with regard to the nature of mental representation, like content identity, also tends to result in ignoring the constraints placed upon accounts of mental representation by the vehicles of those representations, namely,

neural structures. Further, the notion of graded content that seems to allow us to account for misrepresentation, and solves the disjunction problem was dependent upon content similarity, as opposed to content identity. Even further, it has also been argued that a notion of content similarity need not presuppose a robust notion of content identity asymmetrically, as it seems to be the case that identity with regard to content (as well as other things) is dependent upon similarity. Finally, the focus upon language and content identity in attempting to account for conceptual content requires Platonic Forms, and so seems to drag along with it a great deal of unwanted metaphysical baggage. Personally, I would rather not have to carry it.

References

Block, Ned. 1986. "Advertisement for a Semantics of Psychology." *Midwest Studies in Philosophy*. 10. 615-678.

Chomsky, Noam. 2006. *Language and Mind*. 3rd ed. Cambridge: Cambridge.

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

Cummins, Robert. 1996. *Representations, Targets, and Attitudes*. Cambridge, MA: MIT.

Davidson, Donald. 2001(a). "Thought and Talk." *Inquiries Into Truth and Interpretation*. 2nd ed. New York: Oxford, 155-170.

Davidson, Donald. 2001(b). "Rational Animals." *Subjective, Intersubjective, Objective*. New York: Oxford. 95-105.

Davidson, Donald. 2001(c). "The Emergence of Thought." *Subjective, Intersubjective, Objective*. New York: Oxford. 123-134.

Dennett, Daniel C. 1987. "Beyond Belief." *The Intentional Stance*. Cambridge MA: MIT. 117-202.

Deutsch, Harry. 2002. "Relative Identity." *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/identity-relative/

Devitt, Michael, and Kim Sterelny. 1999. *Language and Reality: An Introduction to the Philosophy of Language*. 2nd ed. Cambridge, MA: MIT.

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

Dretske, Fred I. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge MA: MIT.

Dretske, Fred I. 1999. “Machines, Plants and Animals: The Origins of Agency.” *Erkenntnis*. 51. 19-31.

Eliasmith, Chris. 2005. “Neurosemantics and Categories.” In *Handbook of Categorization in Cognitive Science*. Eds Henri Cohen and Claire Lefebvre. London: Elsevier. 1035-1054.

Evans, Gareth. 1982. *The Varieties of Reference*. Ed. John McDowell. New York: Oxford.

Field, Hartry. 1978. “Mental Representations.” *Erkenntnis*. 13. 9-61.

Fodor, Jerry. 1975. *The Language of Thought*. New York: Thomas Y. Crowell.

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

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

Fodor, Jerry. 1998. *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford.

Fodor, Jerry and Ernest Lepore. 1992 *Holism: A Shopper’s Guide*. Oxford: Blackwell.

Fodor, Jerry and Ernie Lepore. 1999. “All at Sea in Semantic Space: Churchland on Meaning Similarity.” *Journal of Philosophy*. 96:8. 381-403

Fodor, Jerry and Ernest Lepore. 2002. *The Compositionality Papers*. New York: Oxford.

Harman, Gilbert. 1982. “Conceptual Role Semantics.” *Notre Dame Journal of Formal Logic*. 23:2. 242-256.

Krantz, David H., R. Duncan Luce, Patrick Suppes, and Amos Tversky. (1971). *Foundations of Measurement: Volume 1: Additive and Polynomial Representations*. New York: Academic Press.

Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Oxford: Blackwell.

Lepore, Ernest. 1997. "Conditions on Understanding Language." *Proceedings of the Aristotelian Society*. 98. 41-60.

Loar, Brian. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

Lycan, William. 1984. *Logical Form in Natural Language*. Cambridge, MA: MIT.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT.

Millikan, Ruth Garrett. 1989. "Biosemantics." *Journal of Philosophy*. 86:6. 281-297.

Murphy, Gregory L. 2002. *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Putnam, Hillary. 1975. "The Meaning of 'Meaning.'" *Mind, Language, and Reality:Philosophical PapersVolume 2*. Cambridge: Cambridge. 215-271.

Ryder, Dan. 2004. "SINBAD Neurosemantics: A Theory of Mental Representation." *Mind and Language*. 19:2. 211-240.

Noonan, Harold. 2006. "Identity." *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/identity/

Russell, Bertrand. 1905/2005. "On Denoting." *Mind*. 114:456. 873-887.

Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge.

Speaks, Jeff. 2006. "Is Mental Content Prior to Linguistic Meaning?" *Nous*. 40:3. 428-467.

Stainton, Robert J. 1996. *Philosophical Perspectives on Language*. Peterborough: Broadview Press.

Stalnaker, Robert. 1999. "Propositional Attitudes." *The MIT Encyclopedia of Cognitive Science*. Eds. Robert A. Wilson, and Frank C. Keil. Cambridge MA, MIT Press. 678-679.

Stampe, Dennis. 1979. "Toward a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy 2*. 81-102

Sterelny, Kim. 1990. *The Representational Theory of Mind: An Introduction*. Oxford: Blackwell.