Naturalized Theories of Representation

by

Andrew Buzzell

B.A. Carleton, 2006

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

Carleton University

Ottawa, Ontario

Your file Votre référence
ISBN: 978-0-494-68678-2
Our file Notre référence
ISBN: 978-0-494-68678-2

Carleton University

Faculty of Graduate Studies and Research

THESIS SUPERVISOR APPROVAL FORM

I have read the Thesis of:

Andrew Buzzell

and agree that it is ready to be examined.

_____

Thesis Supervisor

_____

Date

Abstract

This thesis discusses the philosophical literature concerning the problem of developing naturalized theories of mental representation. The motivation for attempting to provide these theories is described, and analyzed in terms of the extent to which such theories are required to solve problems relating to the role that the representational theory of mind plays in the foundations of cognitive science. A set of standard problems that such theories face is described, and the ability of candidate theories, especially teleosemantics, to solve them is discussed. The homunculus problem and the indeterminacy problem are identified as particularly important. Proposals to solve these problems are discussed and a solution based on Dennett's decompositional strategy and Brook's self-representation theory is endorsed. The solution requires that a distinction between representational states at the psychological level and non-representational states at sub-psychological level be maintained, a consequence that is explained and defended. The applicability of this distinction to neurosemantic theories is discussed and recommended.

Acknowledgements

Acknowledgments

I would like to thank Vincent Bergeron, Gabriele Contessa, and Jay Drydyk for agreeing to be on my committee and providing very insightful comments and suggestions. I would also like to thank all those at Carleton, especially Sandra Kirkpatrick, who have made this department so excellent, and helped me to be a part of it.

It has been a great privilege to work with Andrew Brook as my thesis advisor and also as his student for many years. His guidance, advice, and encouragement have helped me immeasurably, for which I am very grateful.

My partner Kerry-Lee Powell has taught me a lot, and her love and companionship means the world to me. I could not have completed this without her patience, inspiration, and support.

I would like to thank my parents, Gary and Gail Buzzell, for their extraordinary and unshakeable support. It would be impossible to exaggerate the extent of their encouragement and assistance in all aspects of my life, or my gratitude for this.

Table of Contents

Chapter One

The concept of representation is fundamental to mainstream cognitive science and philosophical work on the mind, cognition, and consciousness. We can characterize what has been called cognitivist explanation as fundamentally computational, in that it explains and describes cognition as computation over mental representations. In light of the considerable success of the cognitivist paradigm, it is both surprising and unsettling to observe that the very concept of representation has been the subject of an extraordinary amount of philosophical controversy. Here we have a hugely successful research program that uses representation as a basic element, yet there is overwhelming disagreement and lack of clarity about what representations are and how they work. In this chapter I'll provide an overview of some of the philosophical problems surrounding representation and then focus on a few particular problems that I argue are central, which I will analyze in greater details in later chapters.

Computation (in the standard sense, connectionist systems are special cases) involves the manipulation of symbols, and those symbols, both at the lower levels of input and computational activity and at the higher levels of output, are deemed to 'stand for', to be about, other things. We can associate this with what as come to be called the Representational Theory of Mind (**RTM**): Cognition is computation over mental representations.

1.1 Intentionality and Physicalism

Intentionality itself has long been thought to pose some dilemmas for physicalism in that it is difficult to see how the relational properties of representation can be naturalized. *How* can some material thing be about some other thing? At least since Brentano claimed that intentionality is the mark of the mental, explaining how things can be about other things has been a major philosophical concern. Mental representations are states of mind that have this relationship of aboutness to something else.

Physicalism is the working hypothesis that underwrites nearly all work in cognitive science and philosophy of mind – the idea that all mental states are, ultimately, physical states. There is a large body of literature concerning the extent to which physicalism requires or entails the reduction of mental properties to physical ones. For my purposes here, the issue is simply whether we can find a way to speak of the intentionality of a physical system in terms that are not intentional. I take no stand on the question of the reducibility of mental states generally to physical states. The goal of the kind of naturalism I am concerned with can be described as the Naturalization Thesis (**NT**): mental representations supervene on physical states that can be individuated and their causal powers discerned based on their physical, non-semantic properties. Supervenience in this sense requires that changes in representational content must entail physical changes in whatever material state subvenes the representation. If this supervenience did not hold, then the explanation fails, since NT demands that the semantic properties of the representation cannot be independent of the physical properties.

To satisfy NT, theorists have sought to develop Naturalized Theories of Representation (**NTR**). In this sense of the term, naturalized means that the theory must explain the semantic features of a representation in terms of the physical, functional or causal properties of the matter that instantiates it. There is a distinct question as to whether or not an NTR is required to underwrite claims about reducibility of mental events to physical events, a question that falls outside of the scope of the argument I will be developing here.

One of the attractions of NT is that it offers a way to resolve the apparent conflict within RTM. If we are going to use representation to explain how cognition works, then it seems we must have an explanation for representation that does not require minds. "A definition of representation in semantic terms would simply be circular" (Cummins, 1996, p. 3). Yet, as we shall see, it is difficult, if not impossible, to provide mind-independent accounts of representation. There are two ways in which theories of representation are liable to be circular, which I'll refer to as **CP**, the circularity problem. First, representation can't be explained in terms that are themselves semantic. Secondly, representation can't be explained in terms that require a representing agent, a problem that has been called the "Homunculus problem" and which I will discuss at length later in this chapter, an again in chapter two. To resolve the conflict within RTM, we require a theory of representation that somehow provides a purely physical and syntactic account of representation without resort to the intentions of the wielder of the representation. If we want RTM, we need to solve both of these problems. Some form of NTR appears to be a way to do this. Although there these two kinds of circularity problem I'll refer to them collectively as CP unless I'm analyzing one form of it in particular.

It is important to note that CP only arises when we are interested in the foundations of RTM and functional accounts of representation, as there are alternative accounts of cognitive science, such as dynamicist approaches (Clark, 1997) (Wheeler, 2001), that might avoid this issue. This is outside the scope of my project here, which focuses on the philosophical problem of representation as it arises in the mainstream conception of cognitive science, which is still very much committed to RTM.

Representations can be about things that do not, or could not, exist, but that they are representations at all requires a relation between the representation and something outside the representation, or a representation of this relationship. Similarly, representations can fail, and when they represent inaccurately that which fills the content role is not the thing the representation was actually supposed to be about. It doesn't seem to make sense to suggest they somehow stop being representations when they err, nor does it make sense to suggest that the representation still represents X if its content is actually Y. That something might be the kind of thing it is by virtue of a corrigible relationship to another thing is the sticking point. Merely 'being a representation' requires, as Brentano puts it, "...reference to a content, direction toward an object..." (Brentano, 1874/1973 p. 88) whether or not the object exists, or is the right one. If this is the case, it is daunting to explain how one bit of the physical world, on account of its particular physical properties, is a representation of some other part of the world, when the other part of the relation doesn't event need to exist.

At first glance, representation would seem to involve a two-part relationship, between the representation itself (the vehicle), and what it is about, the content of the

representation. In perceiving a tennis ball, we might characterize whatever neural states on which the representation supervenes as the representational vehicle for which the content is the tennis ball. If I reach for the tennis ball, only to discover I've confused it with the round lid of the tennis ball can, it becomes obvious that I have misrepresented. I saw the lid, but judged it to represent the ball. We can introduce the notion of a representational target (Cummins, 1996) to account for what a representation is *supposed* to have as content. I have a tennis ball representational vehicle, the target is the tennis ball, and the content is the (mistaken) round lid. Even if there is a straightforward way to account for the relation between representational vehicles and their content, it is the target that resists reductive explanation most stubbornly. Consider a formulation such as 'neural state N correlates to the presence of physical state P'. Indeed, the project of describing what have been called the "Neural Correlates of Consciousness" (NCC) proceeds in just this manner. In terms of the NCC project, it is supposed that some mental states can be conscious via an identifiable neural instantiation. The problem for both approaches is determining how we demarcate and individuate these states. Do we individuate mental states by their content? If so, how do we do this when they misrepresent? If we individuate by target, we seem to have a problem, in that we need to know something about the state already before we can know its target. How can we provide a physical account of the 'directedness', the sense in which the intentional target requires 'intending'?

The content/target distinction complicates the task of describing what kind of physical state can count as a representation. A central part of my argument is the extent to which the representational target resists naturalization, yet is essential to the very idea of

representation. The target is often indeterminate and involves appeal to intention and interpretation. There is no straightforward mapping of vehicles and content to targets. Supervenience almost certainly cannot always be true of targets, in that you can take a representation of one thing and redeploy it as a representation of another thing without changing anything about the representation itself. This will be a major theme of the second chapter. For now, I just want to highlight the representational target as an essential part of the concept of representation, and as resistant to naturalization.

With this preliminary sketch of the territory at hand, I can briefly describe the overall position that I will be developing throughout this thesis. The problem that is of chief importance to my argument here is that representational targets are inherently indeterminate, and the resolution of this indeterminacy occurs at the level of the agent. This threatens any attempt to reduce representation to something that does not have cognitive properties. Among the consequences of this is that a successful NTR looks to be impossible. However, the failure of the NTR project need not undermine RTM. Instead, it is my contention that there are problems with the formulation of NT and CP, and the use of the concept of representation within RTM. Resolution of these problems invalidates NT and CP, and obviates the necessity of NTR. We can't have an NTR, but, that's OK, we didn't need it anyways. As a result, I argue that sub-psychological theories need not take into account issues relating to NT, and that psychological theories are not threatened by CP.

1.2 Problems facing NTR

Before discussing some candidate NTR's, I'd like to describe what I think is the standard problem set for these theories that has emerged in the literature. By way of illustration, consider the most basic form of NTR, the causal theory. Causal theories of representation attempt to provide an NTR by explaining representation in terms of causal relationships between the content and the vehicle. A simple form of this is exemplified by what Fodor called the 'Crude Causal Theory', "...symbol tokenings denote their causes, and the symbol types express the property whose instantiations reliably cause their tokenings" (Fodor, 1987a, p. 225). The content of a representation is simply what caused it to exist. This kind of theory meets the requirement for being an NTR in that something can be a representation under conditions that do not involve agents and semantics.

Here are some standard problems that can be raised for most, if not all, candidate NTRs.

A. Generality Problem

How do we constrain the set of representation-conferring causal relations from the much larger domain of all the causal relations that a cognitive system is involved in? Absent this, all causal chains the mind participates in might necessarily trigger representation. Whatever physical relations are used to describe the relationship between representations and their objects, there is the risk that some relations of that type can be found in contexts where we would not want to say there is representation. If representation is conferred by relation R, there had better not be situations in which relation R obtains, thus conferring representation, in places where it does not make sense to actually ascribe representation. For instance, the function of the heart is to pump blood,

but a functional theory of representation ought not ascribe the semantic property of meaning 'pumps blood' to hearts. Fodor calls this problem "Pansemanticism" (Fodor, 1990, p. 118).

B. Abstract Concepts and Unacquainted Content.

Much of our representation is quite complex, for instance, when we have mental states that are about things like 'democracy'. Theories that posit mechanisms between things in the world and representations in the brain will have difficulty with such concepts because there are no obvious candidate things in the world that we can link them to. As well, they may resist decomposition into things that are. This is particularly true of representational objects that might be fictional, or that we have never interacted with causally, what has been called "unacquainted content" (Scott 2002). There aren't going to be any mechanisms between things I've never seen or that don't exist and my representations of them. Compositionality will not work for all such cases: while concepts like "unicorn" may be subject to decomposition, concepts like "match point" or "justice" probably are not.

C. Proximal/Distal Stimulus and Epistemic Conditions

It would seem that it is the activity in the retina that is most directly implicated in the causal chain from object to representation, and, indeed, change in or damage of the retina would certainly have a direct effect on the representation. But, surely we want to say that it is the cause of the retinal stimulus that is represented. It must be possible for an NTR to pick out the proper object of a representation distinct from the more immediate

physical stimuli that may in fact have higher correlation with the tokening of the representation. Particular patterns of activation of my sensory system will have higher correlations with some representations than their distal cause which creates a problem for theories of representation that link causes too closely to content. This is also true of epistemic conditions, because objects in favorable conditions will correlate more closely with their representations in poor conditions. If this is so, there would seem to be reason to claim the conditions are part of the representational content, which has unintuitive consequences. Of course, a representation *could* be about epistemic conditions, so there must be some principled way of ruling this out that is sensitive to cases where epistemic conditions are permissible content.

D. Misrepresentation and Disjunction

In the case of misrepresentation described above, the tennis ball representation has 'top of tennis ball container' as content. We need a basis to claim that the representation is erroneous. Does my representation of the ball have the content <balls or lids>, or was the lid an error? In the absence of a robust criterion for error, all cases of misrepresentation can be re-described as disjuncts. For any representation with the target X, when that representation is mistakenly tokened by Y, it is possible to describe the content of the representation as the disjunct <X or Y> instead of ascribing error. Note that some representations may very well represent tennis balls and round yellow plastic discs. Some method must be provided to distinguish errors, legitimate disjuncts, and illegitimate disjuncts. The disjunction problem demands a solution that will fix the content of representations in the face of normal, acceptable, and common

misrepresentation by agents. For causal theories, there is a need for a 'filter' to sort the representational causal relations from the non-representational, while preserving the target/content distinction by allowing for error.

E. Indeterminacy

For many representations it is not obvious that we can specify the precise content and target at the level of the individual representation without appeal to the user. Whether the content of a representation is a legitimate disjunct or an error may require appeal to the intent of the agent using the representation. The same goes for proximal/distal stimulus and epistemic conditions. As we will see, especially in chapter two where indeterminacy will be discussed in detail, there will be cases where any number of characterizations of the content and target will be equally viable. Since at least some representation has very precise objects, it is a problem if an NTR entails that all content is indeterminate.

F. Homunculus Problem

It is widely held that representation is a relational property, not an intrinsic one (Dennett, 1978) (Searle, 1992). As such, a user of some sort is required, something in relation to which the representation represents. Dennett calls this "Hume's Problem", that, "... nothing is intrinsically a representation of anything; something is only representation for or to someone" (Dennett, 1978, p. 122). A distinction can be made between the personal level of analysis, which is the level at which we can refer to psychological states available, or available in principle, to the agent, and the sub-

personal, which involves states that are not. The subpersonal level of analysis involves non-psychological (or cognitive psychological) terms describing states of the brain, such as neural activations. At the subpersonal level there is a problem accounting for the representational roles of information without deriving the intentionality from the higher levels. Dennett suggests that classical (vs. connectionist) AI shows us a way out, that the interpretive work of the homunculus can be delegated to progressively simpler homunculi until a level of processing is reached that is simply binary. If we hold a representational theory of mind, then we must have an account of subpersonal representation that is not derived from higher level psychological intentionality.

The very idea of representation requires a user for whom the representation is meaningful. The problem for an NTR is to reduce the sophistication required in the user such that semantic competence is not presumed. This relates to the distinction between original intentionality and derived intentionality. The representational nature of a sentence, picture, or gesture is derived from the fact the agents use them to stand for things. The meaningfulness of the lines on a tennis court, or the word 'ball', is wholly dependent on the existence of agents who employ these things in meaningful ways. The intentionality in the agent's mind, on the other hand, is thought to be, or at least ought to be, original and un-derived, or we are left with the need to explain this intentionality in terms of additional users.

This is related to CP, in that we can't require consumption by an agent to be a necessary condition for there to be representation if this consumption itself requires representation. If we need a fully-functioning cognitive agent for there to be

representation at all, we are in trouble if we want to explain cognition by appealing to representation.

The extent of the homunculus problem becomes clear as we examine the notion of the representational target more closely. The notion of the target allows us to distinguish between correct and incorrect representations. When we claim that a representational vehicle V has content C, in addition to a story about how it came to acquire C, we additionally need to explain that it was *supposed* to have C, that the representation is being used *as* C. Otherwise we haven't yet shown the phenomena in question to be a bona fide representation.

With these problems in mind, I'll turn my attention to some candidate NTR's, specifically, asymmetric dependence and telefunctionalism.

1.3 Asymmetric Dependence

Fodor has developed a causal theory of content called asymmetric dependence that is designed to solve some of these problems. As Fodor describes it:

> "X" means X if:
>
> 1. 'X's cause "X"s is a law.
>
> 2. Some "X"s are actually caused by Xs

3. For all Y not = X, if Y's qua Ys actually cause "X"s, then Ys causing "X"s is asymmetrically dependent on Xs causing "X"s (Fodor 1990 pp 121).

I'll adopt the notation /REP/ to denote a representation. So a cat causes a /CAT/ because it is a law that cats cause /CAT/s. If a dog causes a /CAT/, this is dependent on the fact that it is usually cats that cause /CAT/. Were it not the case that cats normally cause /CAT/, the dog would not have triggered that particular representation.

For our purposes, we can read premise one as an appeal to some version of the kind of causal theory discussed in section 1.2. Some kind of psychophysical law links representations to their causes. We need premise 2 for there to be any internal representations in the first place, unless we were to posit that all representations are innate. Exposure to at least some X's that token their correct representation is required. Premise 3 is where the hard work is done. Fodor solves the disjunction problem by linking erroneous representation to veridical representation. The /CAT/ representation is not about <cats or dogs mistaken for cats> because the only reason the dog tokened the /CAT/ representation was because cats normally do. The tokening has taken place as a result of the /CAT/ psychophysical law, not a /DOG/ one. Thus the /CAT/ representation acquiring dog content does not render it disjunctive, because we have a basis to consider this an error.

Philosophers have found several problems for asymmetric dependence, but I'll only focus a few that are most relevant to my interests here. The generality problem is still an issue for Fodor. There is lots of asymmetric dependence that is not representation-conferring. Adams and Aizawa have argued (Adams and Aizawa 1992, 1994) that we can

find cases where we have asymmetric dependence where we would not want to say there is also meaning, and that, as a result, Fodor's theory overextends semantic properties into domains where we wouldn't normally want them.

To limit the domain of genuinely representation-conferring asymmetric dependence, Fodor introduces a fourth premise in a later statement of the theory, that the dependence must be synchronic. Whatever keeps X and "X" in phase must be operant when the representation is tokened. This blocks situations in which something like X, but other than X, starts to token X's. A counterfeit dollar tokens /DOLLAR/ because dollars do, but it doesn't mean dollar because the mechanism that connected the counterfeit to the genuine is not synchronic. (Fodor, 1994, p. 19). This sounds promising if Fodor can specify the kinds of mechanisms that could play this role, or better yet, that do play this role. Then there would be grounds for blocking non-synchronic mechanisms that could produce the wrong meaning, or meaning where we don't want any. However, as Brook and Stainton argue, Fodor in fact doesn't have a good account of these mechanisms, and his arguments to the effect that, in fact, he doesn't need them after all, are unsuccessful (Brook and Stainton, 1997). As a result, the synchronic requirement can't help defend the theory.

There is also a worry that the only way to cash out the notion of synchronic mechanisms that Fodor needs is going to be circular, in that, short of appeal to what the representation is really supposed to mean, it does not appear that there is a principled way to single out just those mechanisms that really do cohere with what the representation ought to mean.

Fodor has a bigger problem distinguishing proximal from distal causes. Why does X mean "X" and not "projection of X on the retina"? Fodor argues that the letter does not satisfy the requirement that the link be law-like, "... there is no reason at all to suppose that the causal dependence of the perceptual states on distal objects is asymmetrically dependent on the causal dependence on specific arrays of proximal stimuli ..." (Fodor, 1990, p.108). Fodor's argument is that because it is not always the very same proximal states that correlate to /CAT/, there is no reliable connection between those states and /CAT/ and therefore there is no worry about the content becoming the proximal stimulus instead of actual cats. However, there is still a problem here, which we can see when we switch to things that are not natural kinds, such as tables. There is just as much variability amidst all the objects in the world that might, in some situation, token /TABLE/ as there is in all the various proximal stimuli associated with perception of these tables. Absent appeals to natural kinds, there is no way to tie these all together as a type that particular representations can token. As Adams and Aizawa (1994) argue, the situation is even worse when we consider hallucinations and thoughts, in which the immediate cause *is* the proximal stimulus. It might seem as though it should be possible for Fodor to just stipulate that states of the sensory system don't count, but this creates a huge problem for the content of thoughts. Also, sometimes we might very well *want* to represent proximal stimuli. What Fodor needs is something that is not allowed within the theoretical framework he wants to defend, namely, an appeal to what matters to the agent.

I mentioned in passing that Fodor's position with regard to the generality problem may well be circular. At a fundamental level, there is reason to believe that the whole project is, in fact, inherently circular.

> Fodor thinks that intentional laws are not basic laws. So if Y-"X" syntactic laws synchronically and asymmetrically depend upon X-"X" laws, there must be some underlying mechanism to explain this. Further, if the Fodorian account is to reduce meaning to non-meaningful concepts, the explanation cannot involve anything semantic. But what could the explanation of the synchronic, asymmetric dependency be, if not something semantic? (Adams and Aizawa, 1994, p. 227)

They argue that Fodor in fact resorts to this kind of circularity in a number of places, and cite a few passages where this is evident. Indeed, this seems unavoidable, in that there doesn't seem to be anything other than semantic facts that could maintain the asymmetric relationship between Y's and X's. "Fodor thought that syntactically construed, causal asymmetries generate meaning. In fact, it seems to us that things are close to being just the other way around." (ibid.). That there is this asymmetric dependence is *because* of the meaning, which is an interesting result, but does not buy Fodor what he wants, which is an NTR. The project falters on Brentano's problem; it is just very difficult to see how any explanation whatsoever can be provided to explain how a bit of matter can be *about* another bit of matter independently of the acts and intentions of a representing agent. Part of the attraction of causal theories is, I think, because they alone seem to offer any obvious way to attack this problem. But they just don't work.

1.4 Telefunctionalism

In telefunctionalist theories of representation, for R to be a representation of C is for R to have the function of being a surrogative means of interacting with C for some agent A. Telefunctionalist theories of representation introduce a functional constraint on bare causal theories. Teleofunction involves an appeal to the 'proper function' of the representational vehicle. A state represents X and not Y if its proper function is to represent X and not Y. Teleosemantic theories describe the functional relationship in biological terms, whereas information theories do so in terms of the information-conveying functions of the representation. Telefunctionalism has the advantage of straightforward contiguity with other established scientific theorizing, in particular, evolutionary explanation. The sophistication of the representational capacity we want to explain would seem to require increasingly sophisticated agents. That the target of my representation in the example in section 1.2 is 'tennis ball' even though the content happens to be the lid is, on a simplistic teleological model, because the representation is *supposed* to represent tennis balls, not lids. A satisfying teleofunctional causal account of representation will have to find some way to account for the agent for whom the target is 'supposed to' represent, without appealing to intentions and agency, if it is to be a candidate NTR.

A canonical example of representation found in the literature is that of a frog's representation of flies (Lettvin, Maturana, Mcculloch, & Pitts, 1959). A frog perceives, tracks, and strikes at flies. If we want to explain how the frog is able to do this, we study the frog's brain and model its cognitive processes. Such an explanation typically posits that the frog represents the fly. You can set up a situation in which the frog encounters small black objects, say, BB's, and strikes at them as though they were really flies. If a

frog strikes at a BB, mistaking it for a fly, we might say that it has misrepresented, because the function of the representation is to represent the target 'fly, not BB (or the disjunction fly or BB), because it is the ability to represent flies that explains the existence of the discriminatory ability. It is because it is useful for the frog to represent flies that it has the mechanism that led it to strike the BB. Striking at the BB, which is not the object for which the mechanism was intended to detect, is thus an error.

Among teleofunctional theories, we can distinguish two ways of understanding function. Biologically-based etiological approaches ("teleosemantics" hereafter), most closely associated with the work of Ruth Millikan (Millikan, 1989), define proper function in terms of fitness for consumers of representations. The proper function of a representational state is X if X explains the presence of the representational capacity, either directly or derived, via the selection for fitness with relation to the consumer. (Millikan, 2000) The frog's content is fly and not BB because detecting flies is what contributed to the selective success of the frog's ancestors. Information approaches, most closely associated with the work of Fred Dretske, explain proper function in terms of the reliable and deliberate conveyance of information. "[A] system, S, represents a property, F, if and only if S has the function of indicating (providing information about) the F of a certain domain of objects"(Dretske, 1995, p. 2). Information semantics understands representation in terms of functions that respond to specific stimuli. The function of the representation is just to convey information about some particular state of affairs, and there is a learning period in which all tokenings represent just that stimulus and nothing else. Contrast this with the ambiguity that consumer-based teleosemantics must accept when accounting for the frog's representation of the fly, in which it can be said to

represent 'fly' or 'small moving black dot' or 'nutrients' (Agar, 1993), to name a few candidates. For an information theory, it is flies that are represented because it is the fly qua stimulus that the representation is supposed to represent, not the fly qua 'biologically useful food'. Information theories correlate information about objects with the representations of the objects, whereas teleosemantic theories are looser in the specification of the content of a representation.

As a result, teleosemantics has a problem with indeterminacy, since there are any number of ways we can characterize the content and target of a representation with respect to the function it performs for the user. Worse, it would seem that anything at all *could* be a representational vehicle for a user, so it would seem that the theory of representation will have to extend far past the vehicle to explain content, which would to threaten the viability of teleosemantics as an NTR.

While information semantics is designed to be less exposed to the indeterminacy problem, it still has difficulties distinguishing proximal and distal content. Dretske raises this worry with regard to magnetosomes, microaerophilic bacteria which move away from oxygen-rich water via the ability to discriminate magnetic north (Dretske, 1993). If fooled with a magnet or a move to the southern hemisphere, they will move into more highly oxygenated waters and die. It is not obvious if we should best understand the mechanism's function as representing low-oxygen water, or, magnetic north. Dretske's solution is to set a threshold of epistemic complexity requiring multiple methods to obtain information, below which misrepresentation is not possible. If the representer is complex enough to be able to find more than one way of representing an object, then it

can be argued that it isn't representing a set of its proximal features, since any number of such sets could be generated with each representation. Below this threshold, we would not say that the system is representing at all, since it would just be responding in a binary way to the proximal stimulus.

Information theories are concerned with the functions that representations have *on their own* to convey information. We can distinguish these from the sense in which teleosemantics is consumer-based, in that it is the function of the representation for the consumer (not the producer) that determines content. Pietroski (Pietroski, 1992) raises an objection based on a form of indeterminacy for consumer-based theories, describing a situation in which a representational capacity accidentally and indirectly covaries with selection value. The fictional 'Kimu' accidentally evolve the capacity to perceive and enjoy redness. As a result, they climb a nearby hill every evening to watch the sunset, which has the consequence of causing them to avoid a dusk predator, thus conferring survival value to the ability to perceive red. The continued presence of the representational capacity is explicable in terms of a benefit that has no causal connection with the capacity.[1] This isn't a problem for information approaches, as it remains the case that the actual stimulus is what fixes the representational content, but it seems to be for teleosemantic approaches.

Rountree, (1997) mounts an interesting defense, broadly along the lines of Dretske's (1993) suggestion regarding informational complexity, in which he argues that the teleosemantic approach can be spared the indeterminacy in magnetosome-cases by

[1] This is not unlike a Gettier case (Gettier 1963).

distinguishing between representation and belief, and claiming that it is at the level of belief that the putative equivocation would occur, but that bacteria are not complex enough to have beliefs. He argues that the same is true of the Kimu. Pietroski's claim that the Kimu believe that there is redness must amount to belief that there are no predators is undermined if it can be shown that Kimu are not complex enough as stipulated to hold beliefs. Rountree argues that we can distinguish between 'representers' and 'believers' relative to the complexity of stimulus input and behavioral output.

The argument is that either the Kimu have enough complexity to have beliefs, but also enough to undermine the stipulated conditions that lead to indeterminacy, or they do not have beliefs, in which case there is no content indeterminacy. The concept of belief has its own set of philosophical problems, discussion of which goes well beyond the scope of my interests here. On some accounts, there is good reason to doubt that having beliefs will be a guarantee against indeterminacy. This could undermine Rountree's argument. However, as I will argue shortly in my discussion of Millikan, there are other ways to avoid this problem. If it is not individual representations that are explained by appeal to teleofunction, but the capacity to represent itself, then it is not the case that we must derive the content of individual representations from teleofunction, which avoids Pietroski's problem completely.

It is harder to use either version of telefunctionalism to explain abstract representation. Terms that do not refer to concepts that could contribute to fitness pose a significant problem. One strategy is to appeal to compositionality, arguing that we can decompose these concepts into more basic ones that teleosemantics can explain.

However, this approach can founder when faced with terms that do not refer, terms that are not biologically useful, and terms that are not easily decomposed (such as "democracy"). This is a very difficult problem for theories that rely on the history and function of representations.

Millikan's approach offers a different way to solve this problem, via an indirect appeal to compositionality that is better positioned to avoid this kind of objection. "The content of representations that are not used or not used productively is determined by the way other representations in the same system have been successfully used in the past." (Millikan, 2006, p. 106). As I will discuss in the next chapter, this relies on metarepresentation, in that the content of the representation is determined with reference to the other representations we have *about* this representation. It is as part of a truth-preserving system that representational content can be described, and it is the truth-preserving nature of the system as a whole that is teleofunctionally based, not that of individual representations. Referring to Paris does not directly meet a biological need, but being able to do this does.

Note the extent to which this claim undermines the original formulation of teleosemantics, and its potential as an NTR. Instead of deriving the content of representations from a functional or historical description of the vehicle, which at least has some hope of underwriting an NTR, we now have to refer to the agent and other representations that agent has. We were looking for a theory that could explain representation that is independent of the agent. Teleosemantics was supposed to do this by specifying conditions by which some things would be representational by virtue of the

functions they have. Being a representation would be a local property, not involving a relation to a user, and CP would thus be solved. But, on Millikan's account, we lose all of this.

We can see two different roles that teleofunction can play in semantic theories. One can claim that individual representations acquire content via teleofunctional mechanisms, in the way that information theories do, and teleosemantics does without the argument from Millikan above. If we can specify processes by which representations come to be associated with content in a way that doesn't refer to the agent, but instead a sub-psychological cognitive process, then this kind of theory will work well to underwrite an NTR. This is what motivated the teleosemantic project in the first place.

Alternately, one can ground the representational nature of the entire system in teleofunction, as we see in Millikan's argument. How particular representations come to have content involves how the agent uses them, but that the agent has representations at all, and perhaps that they are generally truth-preserving, is to be explained by teleofunction. It explains the phenomena of representation in terms of the intended function of particular representations within the system, involving intention at the psychological level and reference to the semantic properties of other representations. This weakens considerably any claim that such a theory can underwrite an NTR. Invoked in this way, we can think of teleofunction as an answer to the kind of skepticism about representation and indirect realism raised by 17th century philosophers such as Berkeley and Arnauld. If all we know is our ideas, how can we be sure we know anything about the world? The teleofunctional answer is that because if our ideas weren't about the

world, they wouldn't give us good enough information to keep us alive. This is what licenses the claim that our mental representations yield information about external states. Doing so confers survival value; representing the world is what Dennett calls an 'evolutionary good trick' (Clapin, 2002, p. 112). Of course, there remains a philosophical question regarding the resemblance, if any, between the external world and the representations by which we know it, but this is well outside my interests here.

"To the shell that is "teleosemantics" one must add a description of what actual representing is like. When the bare teleosemantic theory has been spent, the central task for a theory has not yet begun (Millikan, 2004, p. 66) Teleosemantics can explain representation as a capacity, but can't be invoked for representational particulars, which are therefore not individually naturalizable in a teleosemantic theory. The functional relationships involved in individual representations are set up by users, not nature. Teleosemantics can be used to explain the nature of representational capacity, but, it doesn't account for the way a lot of human representation actually works. This will be the focus of the next chapter.

One problem teleosemantics does seem to avoid is the generality problem. For one thing, even though the heart has the function of pumping blood, it does not represent, because representational capacity was not, in consumer terms, a contributor to fitness. The approach I've just described from Millikan is immune for the simple reason that nothing gets to be a representation 'on its own'. There's lot of signals, indication, natural signs and the like that reliably covary with their sources. Often these will play an important role in representation. However, it is open for an agent to treat anything as a

representation, setting up the functional relationship that confers representational status and against which error conditions arise. We don't need to worry about generality because there are no 'automatic' representation-conferring mechanisms.

The reason we don't have these is because things don't become representations because of features they possess independently, but because of a relation to a user. As I discussed earlier, this is an important part of the concept of representation, and one that causal theories can't capture. It is for this reason that teleosemantics offers us a much better picture of the phenomena of representation, albeit at the expense of indeterminacy and naturalizability. Representations are representations because agents use them to do things, not because of their physical features or the way they are hooked up to the agent and world. They misrepresent when they don't do what they were supposed to.

1.5 Summary - The Naturalizing Project

There are three things in particular from the foregoing discussion that I'd like to highlight as being particularly relevant to the argument developed in the next chapters.

First, I hope to have motivated some sympathy for teleosemantics as the most viable account of representation. With Millikan's caveat that it is the representational system, not individual representations, that is explained biologically and historically, we have the basis of an account of representation that seems to reflect its most important features, inasmuch as it is sensitive to the relational nature of representation. Of course, this picture is a markedly poor candidate for an NTR, in that it is committed to referring

to the agent and the semantics of other representations in explaining the representational target, which is exactly what we didn't want to do.

Secondly, the problem of content determination has emerged as perhaps the chief problem faced by all NTRs with regard to accurately describing how representation works. This is because the need for error conditions complicates alleged causal links between candidate contents and representational targets. Allow too much room for accurate representation and you can't realistically categorize errors (you get disjuncts). Allow too little and you require unrealistic discrimination on the part of the agent. The difference between flies and food pellets presented as flies is not a real difference to a frog. Any successful theory of representation is going to have to account for this indeterminacy.

Finally, I've framed the problem of naturalizing representation in terms of RTM and the circularity problem. A naturalized theory of representation that solves CP will have to show how representation can underwrite mentality without requiring it. I think this is a fairly uncontroversial assessment of the motivation for the literature on NTR, but I think dealing with CP is a very tough requirement for any NTR. The notion of the representational target seems to lead us straight into both kinds of circularity. Target specification occurs relative to the agent, but we wanted to explain the agent in terms of representation. Target specification requires the ability to represent the target as distinct from other possible content, but this leads to the other CP, because we need to explain representation in terms that are themselves non-semantic.

Taken together, these three problems look daunting. We can't have an NTR because teleosemantics is the best picture of representation, but doesn't naturalize. As well, teleosemantics still appears to have problems with indeterminacy. These issues will be the focus of the next chapter.

Chapter Two

The analysis of NTR's in chapter one treated content/target determination as an important part of a successful naturalized theory of representation. As we saw, the charge of indeterminacy is a standard objection to a candidate NTR. However, while it is commonly taken for granted that indeterminacy is a liability for a theory of content, Dennett (Dennett, 1987), (Dennett, 1995), (Clapin, 2002) has argued to the contrary, that it is unavoidable, and that we wouldn't want to avoid it if we could. For instance, if our representational capacities must necessarily have determinate content, it becomes harder to explain how new discriminatory powers emerge. Allowing for indeterminism also makes room for the differences that exist between individuals – my cat concept will be somewhat different from yours, depending on any number of factors (such as whether or not one has ever seen, or bathed, a cat). On the other hand, associating targets too closely with representational content can lead to a view of meaning that threatens to make communication impossible – our concepts become too personal and idiosyncratic to ground shared understanding of a proposition. Target specification must be sensitive to the cognitive and perceptual capacities of the agent. Content that can be individuated in one way from some external perspective may not be similarly differentiable from the perceiver's perspective, as we've seen with the fly/black dot/nutrient example for the frog. We can generate a number of different accounts of the frog's target, but some of them might be based on distinctions that, while we can make them, probably can't be made by the frog.

In this chapter I will describe three problems that face teleosemantic theories of representation, and develop solutions to them inspired by some of Dennett's work on intentionality.

The indeterminacy problem is a problem with specifying the content of a representation. Teleosemantic theories are thought to be unable to disambiguate between competing potential and plausible interpretations of representational content. Drawing on Dennett's arguments with regard to content essentialism and original intentionality, I'll argue that there is no indeterminism problem, because representation itself is inherently indeterminate.

The precision problem emerges if we accept the essential indeterminacy of representational content. As I will argue, human representation is often remarkably precise and frequently has very determinate targets; a fact that seems at odds with an indeterminist theory of representation. This appears to conflict with the claim that representation is inherently indeterminate. I will argue that there is a way to account for this by referring to the metarepresentation of beliefs and intentions in the production and consumption of representations.

The first two problems are general problems for teleosemantics, whereas the homunculus problem only arises if we want to underwrite the representational theory of mind (RTM) with it. Teleosemantics requires that there be users for there to be representations at all, and it is with reference to a functional description of the user's application of the representation that we can specify the content and target. This seems fine if we want to explain why something like a map or a sentence is a representation, but

not if we want to talk about what is going on in our brains. To recapitulate the earlier discussion of the Circularity Problem (CP); RTM explains the mind by positing representation, therefore, an account of representation is required that does not presuppose minds that use the representations.

Most of the problems raised in objection to naturalized accounts of representation involve the need to resolve content indeterminacy and specify error conditions by which a representation can be said to have the right or wrong content. Without some account of misrepresentation, a theory of representation lacks the ability to track the sense in which representations stand for other states of affairs, and thus corrigible in some way. There are two primary candidates we might appeal to for normativity in this regard. One is 'things themselves'; misrepresentation occurs when a representation gives us wrong information about a thing, or information about the wrong thing. This is associated with causal approaches to mental content, and strong realism about the objects of intentional content. The other is function – representation is successful when it does what it was supposed to do for the system that treated the representation as a stand-in for something external. This approach is more closely associated with teleosemantic theories of content and a more relaxed approach to the ontological commitments we might make to intentional objects.

As we saw in chapter one, a standard example of representation in the literature involves the representation of a fly when a frog visually tracks and strikes at it. In positing representation, we are required to provide, in addition, an account of misrepresentation. When the frog makes a mistake, for instance by striking at a BB, in

the absence of a theory of error, the representational content becomes disjunctive, and represents <flies or BB's>. A theory of error must distinguish errors from disjuncts.

On a causal theory of content, it is expected that a viable naturalized theory of representation will yield "fly" as the content of the representation, and "BB" as the content of an erroneous representation. Teleosemantics might yield something quite a lot less specific, possibly even 'nutrient'(Agar, 1993). Imagine we trick the frog with food pellets that look like flies. Now what is the content, and is there an error?

As we saw in chapter one, causal theories of representation have a difficult time dealing with error, since causal interaction with an object tokens a representation, and, without reference to purposes, it is difficult to distinguish veridical tokenings from erroneous ones. Teleological theories might claim misrepresentation if a frog strikes at a BB, mistaking it for a fly, because the function of the representation is to represent flies, not BB's. We might suppose that really, the function was there to represent 'small dark object moving across the visual field', which is a skill that coincides with the frog's need to eat. No misrepresentation has occurred, the frog has just been tricked. It is only from our external point of view that we can call this an error. Teleosemantics has been criticized for getting content wrong by being too broad. What kind of facts might we appeal to in claiming that a theory of content must yield 'fly' as the content of the frog's representation, and not 'fly and black dots, or food, or 'not dying')? It might appear that we have all the facts at hand, we know it was supposed to be a fly, and we know it was really a BB! But, what, for the frog, does the fly/BB distinction amount to?

2. 1 Dennett on Indeterminacy

Dennett's work on the problem of representation has been somewhat overlooked, and his influence has been largely at the periphery of the NTR debate, despite making several arguments that directly relate to central issues. Dennett's arguments have many affinities with the teleosemantic project. In "*Consciousness Explained*" he declares his position most clearly, "I am a sort of 'teleofunctionalist', of course, perhaps the original teleofunctionalist" (Dennett, 1991. p. 460). Dennett has marshaled some compelling arguments and intuitions in defense of the claim that there is in fact nothing determinate about representational content. If they hold up, then there simply is no indeterminism problem for teleosemantics, as there's no determinate content.

Dennett distinguishes between three kinds of explanatory strategy, or, as he calls them, 'stances'. We might explain the behavior of a thermostat from the physical stance by describing the physical and conductive properties of the mechanism within it. From the design stance we could describe how the thermostat was made and the role it plays in the heating system of a building. Or, we might invoke the intentional stance, and say that the thermostat turns on the furnace because it believes the temperature has fallen below 20. For some time Dennett argued that we could choose between stances, selecting the one that is most convenient to explain some particular phenomena, but, later in "Real Patterns" (1998) he argues some facts are only visible from the intentional stance. I'm not going to take a stand on this issue here, for now I am just introducing the terminology.

We can trace the development of Dennett's views on this point to the substantial body of literature that emerged in the 1980's concerning the viability of artificial

intelligence as a means to shed light on, and even explain, human cognition. In part, what is at stake is the idea that the brain is some kind of computer. In objecting to this, Searle argues that only human cognition generates real meaning, and that other things, such as computers producing artificial intelligence, only have meaning that is ascribed by, and thus derived from, human intentionality (Searle, 1992). Searle argues that symbol manipulation and syntax cannot support semantics in the absence of agents, and the meaning of the symbols is derived from the agent. While Dennett agrees in part, he argues that the distinction is unfounded, that the idea of original intentionality is mistaken. There is only one kind of intentionality, and it is all derived from the interactions between representers and the basic teleology of natural selection.

Dennett's paper "Evolution, Error, and Intentionality" (Dennett, 1987) provides a definitive statement of his approach to intentionality, one that is re-iterated in "The Myth of Underived Intentionality" (1990), and again in Chapter 14 of "*Darwin's Dangerous Idea*" (Dennett, 1995). For Dennett, nothing commits us to a particular account of the representational content of an artifact. A word can be used to refer to something new, the same vector of charges in a memory chip might be about tetris or taxes. Dennett asks that we imagine a vending machine, designed to detect American quarters and dispense soda upon receipt of a genuine quarter (he calls the machine a 'two-bitser'). Describing a mechanism that detects quarters and declines counterfeits is part of how we explain what the machine does. As it turns out, Panamanian quarter-balboas are struck from the same stock as American quarters. They are indistinguishable by the mechanism. Suppose the machine is sent to Panama; the quarter detector is now a quarter-balboa detector. Although nothing internal about the machine has changed, the detector has different

representational content, a different target, and different error conditions. This illustrates what Dennett calls the fallacy of content essentialism, the idea that representations inherently possess content distinct from that attributed to them in use. The representational content is derived from the use to which the representational vehicle is put, which constitutes the target. This is the sense in which Searle and others claim that computer intentionality (indeed, all artifact intentionality) is derived intentionality.

Dennett notes that an important aspect of intentionality is failure of substitution of reference of co-referring representations. The detector refers to quarters or quarter-balboas, but not both, even it if happens that they are in some sense 'the same'. He observes that there is something very similar in nature. A biological process does not demand the particular enzyme that it turns out to use, just that there be something there to perform its function. The underlying teleology of survival drives the need for biological engineering problems to be solved, but the particular solutions are not mandated in themselves. "Just as George IV wondered whether Scott was the author of Waverly without wondering whether Scott was Scott, so natural selection "desired" that isoleucine be the intermediate without desiring that isoleucine be isoleucine" (Dennett, 1987, p. 316). Anything could have performed the function isoleucine performs, anything could be used to represent quarters, all that matters is that the right functional relationship can be created between the user and the representation. Teleology grounds functional descriptions, which are notoriously indeterminate in the same way Dennett claims that representation is. The upshot is that all intentionality is just as indeterminate as that of the computer chip. The functional properties of isoleucine are derived from the functional

role it plays. The representational properties of a representation are similarly derived, in that they don't get them 'on their own'. As such, there is no underived intentionality.

When we mean things, the content relates to how we think the representation works, and how we tried to use it. This is how we can explain the content of a particular representation. But that we mean things at all, this we derive from a deeper teleology, which, if we keep peeling back the layers, lands us back at the underlying teleology of nature. The ability to represent, and, more importantly, to knowingly represent, has turned out to be enormously advantageous (Dennett, 2000). There is an appealing symmetry between teleological theories of representation and the fact that we have the rich representational capacity that we do for similar, teleological, reasons.

Dennett worries (Clapin, 2002) that the idea of written language prejudices our understanding of representation. While we expect precision in the case of frog and the fly, no such precision exists for many of our linguistic concepts, such as 'table', or 'atom'. You can't get twin earth cases off the ground if instead of water, or horses (which we naturally take to have fairly strict extension), we use tables. Fodor (Fodor, 1990) makes a similar point, that natural kind constraints are smuggled into twin earth arguments. I'm writing this outside, its windy, I'm using my mug as a paperweight, it is not clear that there is going to be a way to demarcate 'original paperweights' from 'derived paperweights' in the absence of some rule for the stipulation of things that are inherently paperweights.

Suppose we accept these lines of argument that representation is inherently indeterminate. Considering the two-bitser, we might note that in fact it is not a quarter

detector *per se*, it detects just enough quarter-features to avoid getting too many counterfeits. It doesn't actually detect quarters. We posit that because it was built to have certain physical capabilities that bestowed competencies we value. In the same way, a "hawk detector" in a mouse is really just a shadow detector if we want to explain it as engineers. But, when we ask, why does the mouse have a shadow detector, the answer will have to include hawks, 'shadow detectors turned out to be an easy way to avoid being eaten by hawks'. Mice that were good at detecting shadows were good at detecting hawks <or owls, or eagles, or ....> Since they didn't get eaten, on lives the shadow/hawk detector.

It would appear we have a straightforward defense for teleosemantics. If representation is indeterminate, then teleosemantics doesn't have an indeterminacy problem. The burden now falls on the objector to show that representation really is not indeterminate. Alas, this is all too easy. Precision abounds in the human use of representation.

2.2 The Precision Problem

When we think about cases such as the fly and the black dot, two-bitsers, and shadow detection, we might easily go along with Dennett; representational content is inherently indeterminate. On a teleosemantic construal of the two-bitser, if I trick it with a slug that is a valuable gold coin, it hasn't misrepresented, since it represented <value that will pay for operating costs+profit>. Depending on how we specify 'fitness' we can generate a big space for disjunctive content.

However, there is a problem here; much of what we might want to call representation is not like this at all. In fact, it is incredibly precise. I can describe to you where I buried the treasure in precise detail, and either you will find the treasure, or not. If you find some different treasure because the directions were wrong, this doesn't make the directions disjunctive. If someone else buried their treasure in the same spot, and you grab theirs instead of mine, you grabbed the wrong treasure, my directions don't become disjunctive.

Note that this isn't the indeterminacy problem resurrected. At the very least, we've seen very good reasons to acknowledge that an awful lot of representation really is indeterminate. We've seen that teleosemantics can explain that kind of representation. The problem now is whether it also explains these precise representations. A theory of representation must account for both.

Suppose I'm telling you about a particularly good place to catch trout, off a series of small islands on the eastern shore of a particular lake in Maine. I show you a photo, point to the islands, gesture to my canoe on the shore and the place where I camped. I explain that the road to town ends, '... just beyond the hill over there on the left".

This looks like very precise representation. I didn't mean, 'this, or any number of similar lakes'. I was telling you exactly where to catch the fish. Does it matter if the photo was actually made at a very similar looking lake in Quebec? Somehow I'd mixed up my photos, and, well, the lakes really are quite similar from that angle. We are powerfully compelled to complain that my story was importantly false. That was *not* where I caught the fish; those were *not* the rocks where I left my canoe. I might claim

that it doesn't matter, that for all intents and purposes the lakes are the same; so really, the photo represented the lake in Maine. One is again drawn to insist that I am fudging things. Sure, I can pretend the photo is of Maine, but I have to admit it is really about Quebec.

In the same way, when we misunderstand each other, we do not find it natural to revise our own account of what we meant. If you misunderstand me, it is the case that you didn't get the actual meaning I wished to convey (even if I was vague). I may have potentially meant any number of things, but, usually, there was something in particular that was intended to be communicated.

From a teleosemantic perspective, the issue turns on whether or not it is the history of the artifact, or the history of the representational *use* of the artifact, that matters. In chapter one I distinguished between two ways of understanding teleofunction, in terms of the information-yielding function of individual representations, or, in terms of the function the user ascribes to a representational vehicle. Teleosemantics is generally associated with the former, in which it is the history of the representation that determines its function, but, as I argued in my discussion of Millikan's position in the latter part of chapter one, this is not the only way to go, and probably not the right one. The alternative invokes teleofunction only to ground general claims about the representational nature of the mind, not to determine content for particular representations.

2.3 Artifact History and Representational Content

I want to distinguish between two properties the photograph has in the above story. First, there is the artifact history, the actual casual chain of events that explain the existence of this particular bit of matter. Secondly, and distinctly, there is whatever representational content it might have. The intuition behind the idea of original intentionality can be understood as a conflation of these two distinct properties. Of course, there is something about the photograph that makes it a better representation of one thing than another. But this doesn't limit what it could be used to represent. That it can be a representation at all is determined by the use it is put to. The cause of the artifact does not constrain the kind of representation it can be used as, even though this may be relevant to its usefulness relative to some targets.

As an artifact, the photograph originates in exposing my film to the light reflected by the lake in Maine. This does not imbue the representation with content, this does not even alone confer representational status on the photograph. The strength of the intuition of original intentionality stems in part because for most of our use of representation, content happens to coincide with artifact history. Most of the representing that we encounter in the world has history and conventions that we already know about, and against which we can make judgments about particular instances of representation. We know that photos have causal connections to particular events, and usually that's also what we use them to represent. But this is accidental. I usually know what I mean when I

say something[2]. I have only limited access to the extent to which you really know what I mean when you nod your head, but there are circumstances that will tip me off if it turns out you didn't get it. We are accustomed to knowing why a representation is being used. Typically, photographs are about their causes, and maps are about why they were drawn. As a result, there tends to not be a lot of situations where we can accidentally generate disjunction. "You mean, that was really a map of Houston? I've been getting around Baghdad with it just fine!" But, a sufficiently grainy photograph of me on my bicycle can just as easily be used to represent any number of people on their bikes, just as pretty much any statement, in the right situation, can mean just about anything. The frequent co-referentialilty between the artifact history and the content the artifact is used to represent is entirely accidental, even though it is systematic. As long as the photo is used to tell you about where I caught the fish in Maine, that's what it is about. The fact that the history of the vehicle happens to diverge from the representation itself is wholly beside the point.

Turning back to Searle and original intentionality, we can observe that isomorphism is often invoked to try explain representation (Cummins, 1996), as it has been by functionalists to explain multiple realizability. It is tempting to posit that it is a kind of isomorphism that ties representations to their contents. However, Searle notes that isomorphism is cheap, "... the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecular movements that is isomorphic with the formal structure of Wordstar" (Searle, 1992 pp. 208). In that case, why isn't the wall running Wordstar, granted that exact isomorphism between the

[2] Or, at least, we think we do. Often we are wrong. Substantial revision of our account of what we think we mean can easily be motivated by reflection and discussion.

program on the computer and pretty much anything else can be identified? And why is Searle's computer running it? Searle argues that this is because running Wordstar is derived from our original intentionality to treat something as running Wordstar, and we can't easily do that for the wall.

If we agree with Dennett, we can refine Searle's explanation; it's not simply that we have chosen not to treat the wall as implementing Wordstar, but that we have no reason to. Things are not only representations because users make them into representations, but, also, because there are features of the world that we can only perceive and understand if we adopt the intentional stance and treat things as representations There is nothing about the wall that I can't understand without positing word processing, but there is in my computer.

Dennett imagines two computers, linked by a cable, that flash a light when a button is pressed (Dennett, 1995), which a group of scientists are studying to try and understand why the light flashes. Examination of the computers reveals that the signal is different each time, and (painstaking) physical explanation of any particular instance of the behavior is possible after the fact. However, it remains impossible to predict how the system will behave. As it turns out, the computers are exchanging ASCII encoded natural language sentences, and each contain an independently generated database of truths. If the producing machine's statement is true for the consuming machine, the light flashes. According to Dennett, the brute microcausal physical facts of the system are not enough to explain the behavior of the artifacts, ascribing representation is required to perceive and explain the regularities. By contrast, in the Wordstar example, we don't need to

resort to the intentional idiom to understand what the wall is doing, because it isn't doing anything. Searle wants something more than syntax to ground real meaning, Dennett thinks there is no real meaning. Searle thinks there must be real (particular, principled) reasons to explain why the wall isn't representing something, (and, why something else is), whereas Dennett thinks invocation of the intentional idiom is up for grabs. It may be possible to explain things without it, but this would be needlessly complex.

In later versions of this argument Dennett defends a stronger position (Dennett, 1998), that in fact sometimes we must adopt the intentional idiom to describe some phenomena. An important issue turns on this distinction, but it is some ways outside the scope of my argument here. Briefly, the issue is whether or not using the intentional idiom is a matter of contingent or absolute necessity. By contingent, I mean that the only reason a physical explanation can't work is because the problem is not tractable given the computing resources available in, well, the whole universe. The reasons for resorting to the intentional level are then a matter of practicality; however wildly impractical the alternatives, there is no other principled reason they could not work. Or, are there certain kinds of facts about the world that can only be expressed, observed, and understood within the intentional idiom? Perhaps we can imagine substituting "my c-fibres are firing" for "my head hurts". But, consider, "He stepped back because he wanted to see how the painting looked compared to the others". Are there explanatorily essential facts that can't be recast in the other idiom? It is worth mentioning this issue in passing since it is very central to Dennett's work on intentionality, but it is not central to those aspects of it that are important here, so a full discussion is not needed. Whether it is a matter of necessity or just in-principle intractability that motivates use of the intentional stance, we

still have to use it. For my purposes here, I don't think I need to adopt a position with regard to this metaphysical question. There is a job description for anything we might call a representation, and I am concerned with whatever fits that description. If the state has the kind of aboutness and fitness for surrogative reasoning that representation as I am using the term does, then it will count as a representation for the purposes of my argument.

2.4 Precision and Metarepresentation

On this account, one reason why we are able to use representational vehicles that can have ambiguous content in ways that are actually precise is because the producer represents and remembers the reason why they used the representation, and the consumer has similarly cognized these reasons. Error conditions arise from these intentions and misrepresentation occurs when communication fails. We represent our actual intentions when we use precise representations, we know why the vehicles, such as an utterances, photos, maps, gestures, etc. exist, which in turn informs our interpretation of them. It is because they were made to mean something to us that they exist, and knowing this can lead us to more disambiguating them. Precision can be explained by appealing to the constitutive circumstances in which the representation is a representation. The systems involved between the frog's eye and the frog's brain are impressive, but, they are just not set up to make really precise discriminations of the kind we want to make. There is indeterminacy in part because of the granularity of the frog's ontology.

More importantly, frogs don't know that they are representing. Dennett has called the representations we have about our representations 'metarepresentations'. We have beliefs about our beliefs and communicative intentions, about what is semantically exploitable and how. Only human beings can represent their own representations and have attitudes and beliefs about what we represent. (Dennett, 2000) My claim is that precise representation is to be explained by metarepresentation of the history of our representational behavior. In setting up a function for the representation to perform, we introduce constraints on the indeterminacy of possible content. Sure, the photo could be of any number of things, in fact, it is not about anything until I use it. When I do, I know why I am using it, and how, and these are the kinds of metarepresentation that lend precision to the representation itself. Both the producer and the consumer of the representation rely on the metarepresentational context in order to interpret precise representation.

This seems dangerously close to amounting to a regress, by deriving the precision of the representation from a representation of our precise intentions. However, the precision does not emerge from some particular, individual, intention. Rather, the capacity for precision is supported by the entire network of representation and metarepresentation required for any particular precise representation. An enormous amount of infrastructure is required for precise representation, the most basic and most powerful of which is language. The fact that we all generally cognize our environment similarly, that we all perceive colours and shapes roughly the same way, and have shared knowledge and culture, provides an enormous amount of shared background that we can exploit in order to be precise. We assume teleology and rationality, that people really

mean things, and try to figure out what they mean by examining the artifacts and utterances they use. We can't and don't know everything about what others mean, but we know an awful lot of it by virtue of what we know about ourselves. By appealing to metarepresentation, I'm appealing to all of the infrastructure needed to make sense of the particular, precise representation.

This distinction between artifact history and representational content yields an interesting perspective on an old problem. The intimate, yet contingent, relationship between history and meaning can be used to respond to Chomsky's argument (Chomsky 1969) against Quine's indeterminacy of translation thesis. (Quine 1960). Quine argues that we cannot translate with certainty expressions in an unknown language based on observations of behavior, and that the meaning of the terms will always be inherently indeterminate. This is because we need to make ontological assumptions about the speaker that we cannot confirm empirically, and, thus, there is no way to verify that our meaning ascriptions are accurate. Comparisons of competing ontological assumptions have to be made in the original language; as a result we can't confront the ontology of the other language on its own terms. Chomsky argues that Quine exaggerates the extent of indeterminacy of meaning, that it is, in fact, no more serious in the semantic domain than in any other, and, thus, can't ground the radical indeterminacy thesis.

Chomsky argues that is a matter of empirical fact, for instance, whether our ontology works one way or another, and a matter of empirical fact whether or not human languages can differ in particular ways. Therefore, we can determine which ontology a human language user will be using. Allen (2010) distinguishes Quine's "Argument from

above", based on Duhem's claims for the underdetermination of theories by evidence, from the "Argument from below", that the methodological barriers to radical translation entail indeterminacy. Quine's reply (Quine 1969) rests most heavily on the "argument from above", he claims that while there are certain facts that can arbitrate between competing natural theories, there are no such facts to appeal to in the problem of radical translation. Chomsky's attack is more focused on the argument from below, minimizing the methodological barriers and thus claiming that the indeterminacy is unduly inflated.

There is, I think, some confusion here between two ways of making claims about semantic facts. Chomsky's argument is appealing in that it is a matter of simple observation that such indeterminacy just isn't rampant in language usage. Consider the way Dennett illustrates the radical translation problem. He creates a crossword puzzle that has two compatible solution sets; there is, then, no correct answer (Dennett 2000). He notes how difficult it was to construct, which he calls 'the cryptographer's constraint'. Consider how highly artificial the puzzle is; it is hardly a convincing illustration of an allegedly universal phenomena. Dennett argues, "The reason we don't have indeterminacy of radical translation is not because, as a matter of metaphysical fact, there are real meanings in there (Quine's museum myth) but because the cryptographer's constraint makes it a vanishingly small worry" (Dennett, 2000, p. 346). The very requirements for there to be meaning at all undermine the real world possibility of radical indeterminacy. The nature of crossword puzzles make it hard to build Dennett's 'Quinean crossword', and the nature of representation and metarepresentation minimizes indeterminacy in the wild. We need the whole network of beliefs and capacities to have

meaning at all, and the precision of that meaning, the extent to which we are able to individuate particulate meanings, emerges from the whole.

Chomsky insists that there are empirical matters of fact about how human language works, and how human ontology works, that can potentially settle questions of interpretation. Thus, the special indeterminacy of language is reduced to the garden-variety inherent in theory construction. But these are not the kinds of facts that can settle *specific* problems of interpretation; rather, they are facts that can inform more generalized attempts to understand semantic competencies. While it is the case that human beings invariably adopt object/property ontology, and, thus, we can rely on this when trying to understand a language, it is not the case that we *must* use language in that way. So long as it is the case that we can't rule out the possibility that, using Quine's example, "gavagai" could mean "undetached rabbit parts" instead of "rabbit", the indeterminacy remains. For Chomsky's argument to succeed, it would have to be the case that it *must* mean rabbit (or must not mean undetached rabbit parts), which is far too strong a claim. The only facts we can appeal to are speaker intentions, which we must be able to describe, but which we would be forced to describe in our own language. In normal communicative situations, we almost always know these things, and thus, are almost never exposed to this problem. But this knowledge does not amount to knowledge about the semantics of the words themselves (they don't have any!).

Chomsky's argument relies on general facts about language and meaning that do not amount to laws that govern all specific instances, and it because of just this that Quine's position can be defended. The lack of generalized laws is the consequence of the

fact that content essentialism is a fallacy. If, on the other hand, certain kinds of representation must have certain kinds of content, then we would be able construe these generalizations in a stronger way, as being lawlike, which would buy Chomsky what he needs. Then he could argue that because humans *necessarily* use linguistic form X to mean Y, we are in fact able to do radical translation. But, as I have argued, we can't make this kind of claim about representational content, so at best we have generalizations, which aren't enough to support Chomsky's argument. Therefore Chomsky's epistemological argument, which is probably accurate, does not undermine the radical indeterminacy thesis. In fact, the epistemic claim Chomsky defends bolsters Dennett's position, in that it helps explain how it is that we avoid indeterminacy by revealing some of the shared infrastructure we all bring to bear on our representational and interpretative activity. There can be indeterminacy about meaning without it undermining most of our representing because meaning doesn't inhere in individual representations, but in the network of representations, intentions, and shared understandings; which is in part what Chomsky is emphasizing. Our semantic competence is in spite of indeterminacy, not evidence against it.

2.5 The Homunculus Problem

"... [P]sychology without homunculi is impossible. But psychology with homunculi is doomed to circularity or infinite regress, so psychology is impossible." (Dennett, 1978, p. 122)

Suppose that we can articulate a teleosemantic theory that is not vulnerable to any of the standard objections, such as indeterminacy, generality, and disjunction, and,

additionally, this new objection regarding precision. Can a teleosemantic theory of representation construed in the way I've been arguing be used to support RTM? As I argued earlier, to do this requires avoidance of *both* circularity problems. The theory must explain representation without requiring a mind to act as the user of the representation, and the primitives used to explain representation cannot themselves be representational. The homunculus problem is the problem of the user. If representations are only representations for or to someone, then for there to be a representation at all would require a user. The challenge is to explain how we build users out of representations without taking representation out of the user.

Dennett is well aware of this problem, and the importance of it, although his work on it is scattered across just a few pages and footnotes. His most explicit sketch of a potential solution can be found in "Artificial Intelligence as Philosophy and as Psychology" (Dennett 1978), and he gestures towards the same solution in "*Consciousness Explained*" (Dennett, 1991). Dennett observes that Hume's associationism faced this problem, and that Hume's attempt to blunt the requirement of a central user of internal representations is ultimately unsuccessful. Dennett argues that what is needed is a kind of representation that, "... can be said, in the requisite sense, to understand themselves" (Dennett, 1978, p. 123). He argues that the kinds of data structures used in AI (and in software design generally) seem to exhibit exactly this property. When control is passed from one module to another, when specialized processing of input is offloaded to another part of the system, the various subsystems understand each other in their own right, without reference to the end user. The decomposition of complex tasks into sub-tasks, which can be performed by specialized

modules, involves information processing that can preserve and respect the semantic properties of the information, even when the semantic content is not actually available to the module. The sub-systems interact with the structure of the information in a way that can be said to understand it, for their own purposes, without understanding it at the level of the agent.

Dennett distinguishes between the personal and the sub-personal level of explanation, between phenomena in the province of person's aims, beliefs, thoughts, and interactions, and sub-psychological or neural states (Dennett, 1969). Sub-personal representation can be attributed without direct reference to the ends of the person. "What the frog's eye tells the frog's brain is not what the frog's eye tells the frog" (Dennett, 1978, p. 101). Between a psychological act, such as understanding a sentence and a low-level sub-psychological process such as detecting phonemes, we might find several computationally distinct processes that interact with the input in a way that can be understood representationally without referring at all to the meaning of the sentence to the user. If we can, then intra-modular representation consumption would warrant the claim that true representation is occurring. The modules themselves become the homunculi for the representations they consume. We then don't need to refer to the psychological agent as the consumer of the lower-level representations, and can discharge the lower level homunculi by positing consumption of the representations by the lower level modules. Dennett argues that the same move works within the module, that the functioning of these can be explained by positing, "... smaller, more stupid homunculi" (Dennett, 1978, p. 124) and so on, until we get to homunculi so basic (such as a binary state) that they are discharged.

I'd like to distinguish two different ways we could implement the decompositional strategy. The first, and this is how I read Dennett's suggestion, might bear the slogan "Representations all the way down". As we descend levels of computational explanation, we find simpler and simpler representations, with correspondingly less intelligent consumers. There is still representation at the lowest level, but it is just very simple. The second approach instead claims that positing representation gradually loses explanatory traction as we work our way down, and eventually the attribution of intentionality becomes unwarranted. As we work our way down to increasingly more basic sub-psychological levels, there is less and less reason to posit representation at all, and eventually no reason, thus eliminating the homunculus. We attenuate our endorsement of representational status as we examine simpler levels of the system. At higher levels, positing representation is the only way to understand cognitive activity, but, perhaps, at lower levels, it becomes either false or un-illuminating to claim that there is representation. Because representation is not inherent in any state or process, but, instead, attributed to it, we have some basis for evaluating whether or not such an attribution is warranted.

Dennett's account seems to commit us to an 'all or nothing' view of representation. We do not adopt a 'semi-intentional stance', in which we invoke proto-representation, there is either reason to claim there is representation, or not. But, when I consider the final levels of the decompositional strategy, I see no good reason at all to call this representation. What we do find is indication, co-variance, transduction; all things that philosophers, including Dennett, have gone to great pains to show us are not bona fide representations! There does not appear to be a plausible way to describe a

theory of *mis*representation at these low-level systems that does not make direct reference to the personal level, which is exactly what we don't want to do. Errors at this level are best understood in terms of the noise that normal online embodied cognition must face all the time. Only when the errors become systematic, and, thus, errors relative to higher levels, do we have room for misrepresentation. Even then, it seems like we would be better warranted to call this error malfunction instead of misrepresentation, since nothing would be represented at all (vs. the wrong thing being represented). The first strategy just doesn't work.

For the decompositional strategy to work, we need a way to speak of degrees of representation, and to distinguish sub-representational information-bearing states from bona-fide representational states. In the course of our decomposition, we need to work our way from representation to information to something like direct causal indication. We retain the requirement that we must be able to explain what is going on at the lower level without referring to the goals and intentions at the psychological level. There will be edge cases and gray areas where we can go either way on calling something a representation, no clear line is needed. I'll call this the Modified Decompositional Strategy (MDS). It treats representation as a property only of psychological states, and treats all subvening states as non-representational. They can be information-bearing in rich ways, but are not representational. Because there is no representation at low levels, there is no circularity in the sense of there being representation in the mechanism that is proposed to explain how it is that we have representation. This also removes the homunculus at lower levels, absent representation there is no need for a consumer. However, as we will see shortly,

there remains a big problem with the homunculus at the higher level. What is the user of the psychological representation at the higher levels?

Ramsey (2007), in the course of arguing that cognitive science isn't really committed to representation, claims that much of what is deemed to be representational, turns out under analysis to be something else, to fall short of any acceptable definition of representation. A large part of Ramsey's argument involves a critique of claims that reception, transduction, indication, and tacit representation are truly representational. Borrowing some of Ramsey's arguments and nomenclature, we can describe a taxonomy of sub-representational states for MDS to use to identify, from neurons up, how representation can emerge in cognition, and can play central role in cognition, without being basic to it. Mental states can be said to indicate and even simulate without entailing that they represent.

Ramsey considers three strategies to overcome the homunculus problem. The first involves an appeal to tacit representation that is outside my purposes here. The second is Dennett's strategy, the decomposition of the complex, person-level homunculi into increasingly simpler sub-personal homunculi, until the function is mechanically simple enough to explain easily. The third it what Ramsey calls the "mindless" strategy, which demands an account of representation that can explain 'use' without 'users' – representation without presupposing representing minds. This distinction between the decompositional strategy and the mindless strategy is made in terms of a distinction between Input-Output (IO)-Representations, and Simulation (S)-Representation. IO representations are, "... a sub-system's own inputs and outputs that are internal to the

larger super-system's explanatory framework" (Ramsey, 2007 pp. 73). They are treated as representations by the sub-systems they interact with.

Simulation representations are data structures that can be used to model the world via isomorphism. Whereas IO representations represent by indicating or entailing content, S-representations represent by standing in for their content, so that inferences about the content can be made via the representation. We can use S-representations to learn about the world instead of consulting the world directly. The kind of surrogative reasoning that S-representation enables presumes a level of mental competence that is hard to decompose, which makes the mindless strategy seem initially implausible. He borrows an example from Cummins of a mechanical car that must navigate a track. It does this by coupling its steering mechanism to a card with a groove that corresponds to a scaled-down version of the track. The steering is controlled via interaction with the groove in the card, which thus safely brings the car through the track. Ramsey argues that this example shows us non-mental, automated usage of isomorphic structures as representations. The card is interacted with not in a brute causal way, but *as* a representation bearing information about the track. As such, he raises this as an example of a representation without a mind, and thus an example of how the mindless strategy can yield representation without users.

For Ramsey's argument to work it must be shown that the groove is inherently a representation. We've seen many arguments to the effect that this is not possible. We can call it a representation only if we have adopted the intentional stance to it (totally unnecessarily in this case), and elected to call the car a 'user' with the 'goal' to navigate

the track. In essence, we have to put the mind back into the mindless strategy if we want to call the groove a representation.

Here is what I think is going on here. Decomposition of IO-representations, which Ramsey thinks can work, can only work if we believe that individual IO-representations have content (and are *really* representations) at the sub-personal level, which I've argued in various ways that we can't. Only when we relate them to the personal-level user can they be representations, and, in this sense, they are actually S- representations, since they are being used as stand-ins. When they are S- representations, they can indeed be decomposed, via the Modified Decomposition Strategy, which demands that we can't call the IO representations 'representations'. The mindless strategy doesn't work for S-representations if it purports to work below the level of the agent, because the card doesn't have content, and the car isn't *doing* anything, at sub-personal levels of description. We don't need to appeal to representation to explain how it works, although we can. Things go wrong as soon as we posit content without users. Thus MDS works for S- representations, by explaining them in terms of subvening informational states, such as IO-representations and tacit representations, which are themselves explained by less sophisticated states, and so on, in which representation proper plays no role in the underlying explanation.

Aside from the main problem for RTM that I am interested in, it is worth noting that another important philosophical problems relates to the NTR project, namely, the general question of the reducibility of mental states to physical states, and the specific issue of mental causation. This is a big issue, with a large literature that I can't do justice

to here. If all mental states can be naturalized and described in physical terms, then questions arise concerning the status of explanations that use intentional language. Eliminative materialists conclude that there would in fact be no legitimate explanatory role whatsoever for such explanation, that all of our psychological explanations are false. Fodor says this result would be "...the greatest intellectual catastrophe in the history of our species" (Fodor, 1987a, xii), since it means that all of our explanations about why we do things are wrong. Ramsey discusses the argument that eliminative materialism would be entailed should a non-representational theory of mind prove to be correct, and that we would be forced to deny that propositional attitudes are real causes. This does not necessarily follow. Whatever causal powers a psychological state has (*qua* psychological) are presumably (or hopefully) identical with those of whatever set of neural states instantiate it. However, it may be that only in that particular organization do the powers exist, so it is only via a level of description that involves psychological states that such a state could be demarcated (that is, that there is no non-psychological method to individuate the state at all). That the primitives of a theory of mind are not themselves representational is not fatal to a theory that holds that only in terms of representations can what the mind is doing be understood. This is certainly the position I am defending. I raise this briefly here to argue that MDS does not necessarily commit us to eliminativism.

2.6 Homunculus problem and self-representation

We can distinguish between two ways we can face the homunculus problem, at the upper level of psychological states, and at lower levels of sub-psychological

explanation. The upper problem is "what is the consumer of the psychological-level representations"? The lower problem is, "what is the consumer of lower-level representations". The upper problem is wholly distinct from the circularity problem, which is concerned with removing representation from the lower levels of any theory that purports to explain representation. But the lower problem is basically the same thing as CP. A solution to CP such as MDS removes representation from the lower levels, therefore, the need for homunculi goes with it.

However, it is precisely because of this that it leaves untouched the upper problem. We need to explain the homunculus somehow; there remains the problem of accounting for homunculus for the higher level psychological representations. It is curious that so little attention has been paid to this problem. While there is some discussion of the lower problem, there is very little discussion of the upper one. Dennett posits a solution that is aimed at the upper one, but which can only be consistently articulated as MDS, which only solves the lower. Given that representation requires use, what confers representational status to the psychological representations of the agent?

The most obvious candidate for a consumer of these representations is the mind itself. However, this has some obvious problems. If we want to avoid a regress in which we have an undischarged homunculi that is the constitutive consumer of the mind's representation, then we will need to explain how the mind works, how it is able to consume these representations, without using representation in the explanation. Another candidate could be "consciousness"; we might claim that consciousness consumes the sub-representational states described by MDS as representations. This too has serious

drawbacks. It requires that we explain consciousness without reference to representation. It also entails that there are only conscious representations, that no non-conscious states are representational. Both of these consequences are problematic.

If we want to stay within the framework of RTM, and if we do not have the kind of strong NTR that we might have hoped would eliminate the need to posit a sophisticated representational consumer, it would appear that the only way out is going to be to claim that somehow it is representation that consumes its own representations. If we are going to do that, we need to derive the mind from representation. One way[3] to do this without undermining RTM is to adopt an approach developed by Andrew Brook (1994), treating the mind as just a self-representing representation. While the argument is developed within the context of a larger discussion of the Kant's philosophy of mind, it targets this problem specifically. "If various mental functions are performed by different systems in the brain or different simple-minded homunculi, what could the process be like that synthesizes these various activities into the single, simultaneously introspectible patterns of representation, belief, and behavior so central to beings like us?" (Brook 1992, pp 30). Even if we can eliminate the smaller lower level homunculi, we still need to deal with the big one.

I will sketch the elements of the argument that are most directly relevant to the problem at hand, at the expense of glossing over issues relating to Kant and consciousness that are also part of Brook's overall argument. Kant uses the term "general

[3] Quite possibly the only way as there do not appear to be alternative approaches to this specific problem.

experience" to denote the sense in which all of our individual representations appear to us in a way that is unified. We have a single unified experience of multiple representations and objects. Brook describes the general experience as "global representation", which has a "single global object" (Brook, 1992, p. 33). The single global object "... represents a number of intentional objects and/or the representations that represent them, such that to be aware of any of these objects and/or their representations is also to be aware of other objects and/or representations that make it up ..." (ibid.). What has been left out of the theory I have been defending is the thing in relationship to which all our representations represent *to* something. This is what Dennett argues is missing for Hume, but as Brook argues, it is still missing for Dennett and other theorists (Brook, 1992, p. 209).

What is missing is an account of how it can be that representations are representation to me, and that being representations *to me* is common to all the representations in my mind. They all have this quality. There is not just a need for a user, a homunculus, but it must be the subject who is the user. The homunculus has to be part of the mind if the subject is always part of the representation. "If representation and subject are one, however, the mind could itself be a representation" (ibid). The concept of the global representation shows a way that we can account for the sense in which all our representations are not just to a user, but to a particular user, the subject. Brook argues that there are grounds to support a reading of Kant in which "... the mind itself, the thing that has global representations, itself is these global representations ..." (ibid.). The mind is a self-representing global representation with a single global object. Of particular relevance here is the self-representational claim. Can we discharge the homunculus by

arguing that the global representation is its own homunculus, acting as the consumer of it self via self-representation?

Positing self-representation of a global representation consolidates two different problems here. Instead of a question about the homunculus for each of many psychological representations, we have instead just one question about the homunculus for the global representation. Secondly, the approach to representation that I've defended entails a particularly complex homunculus. It requires a lot of capacity and complexity, including metarepresentation. I've argued that the very nature of the phenomena of representation requires this. The global representation itself, as the whole mind, is a much better candidate for this homunculus than any collection of smaller, simpler entities. In addition to consolidating these two problems regarding the homunculus, Brook argues that another problem is simplified by this approach "... instead of two mysteries, the mystery of representation, and the mystery of the subject, we would have only one, the mystery of representation" (Brook, 1992 pp. 232). The circularity problem would still require an explanation, but at least the homunculus problem at the upper level would be solved.

A full exposition of Brook's argument would be out of place here. I hope only to establish that his strategy could indeed discharge the homunculus, and that it is compatible with the approach to representation I have been defending. I think it is clear that the argument would certainly discharge the homunculus, and that in fact it is appears to be the only thing that could. If the self-representational claim can be defended, then it would suffice to fill the role the homunculus must fill. As to the question of compatibility

with the approach of representation I've been defending, I've already noted a couple of ways in which it is. However, one characteristic of the view of representation I've been developing is the extent to which it is resistant to positing representation in sub-psychological states. If the mind is the global representation, and if my arguments against sub-psychological representation are sound, then it probably has to be the case that the single global object represents multiple objects, but doesn't contain multiple representations.

2.7 Summary

I have argued that teleosemantics is not vulnerable to the indeterminism objection, for the reason that representation is inherently indeterminate. Faced with the fact that most human representation is very precise, I argue that metarepresentation of the reasons, nature of human language, cognitive infrastructure, and intentions involved can explain precise representation without committing us to anything inconsistent with teleosemantics. If we want to explain what representation is, in aid of supporting the representational theory of mind, teleosemantics must be explained in a way that solves the homunculus problem. I argue that this is possible. Because representation is attributed, not inherent, Dennett's decompositional strategy can be modified in a way that discharges representation. I've argued that Brook's argument shows us a way that the higher level psychological homunculus can also be discharged.

A theory of representation that can explain intentionality in a way that can secure an account of enough cognitive complexity to have 'agents' seems like a natural

objective. MDS aims to do this by attempting to draw a principled line between psychological representation and subvening cognitive information-bearing states. The next chapter will examine the consequences of endorsing MDS.

Chapter Three

In the last chapter, I proposed a modification to Dennett's decompositional strategy (hereafter MDS) to solve the circularity problem, and argued that we can draw on Brook's argument for self-representing global representations as a solution to the homunculus problem that is otherwise left untouched by MDS. With MDS, I argued that as we move from complex states to simpler ones we can stop attributing representation to them. Otherwise, the decomposition fails, and we still have representation (and thus the need for consuming homunculi) at lower levels. This solution to CP draws a line between bona-fide representation and other sub-psychological states which, while we might be tempted to call them representational, we ought not. Cognition produces representational states, but it is not constituted by them. If we can maintain this distinction between representational states and the non-representational states that produce them, we have grounds to claim that we can build a theory of representation that does itself not depend on it.

By now, I hope to have motivated support for the following:

1. That we do need a naturalized theory of representation (NTR) to solve the CP for RTM.
2. That teleosemantics can be understood as a good theory of representation, and can be construed as a candidate for an NTR.

3. That we can safely accept the indeterminacy inherent in teleosemantic theories.
4. That there is a solution available to the homunculus problem in the argument for self-representing global representations developed by Brook.
5. That the circularity problem can be solved with MDS.

To help keep my terminology straight, I need a blanket term for the cognitive states typically dubbed representational, but which I wish to re-classify; all those states which I argued in chapter two are not properly representational. There are already some candidates in the literature, but many are burdened with theoretical baggage I do not want to take on. There is, though, a need for a label to describe these states; as Dennett notes, "There is no standard term for an event in the brain that carries information or content on some topic" (Dennett, 2005, p.131). I'd like to bundle all the sub-representational states under the term 'code', so that I can easily contrast it with 'representation'. The term code is useful because it encompasses things like transduction and indication, encoding of information, neural and cognitive structures, and things that can be thought of computationally. The code for a chess-playing computer doesn't represent chess until one hooks it up to input and output devices and plays chess with it, in the same way I argued previously that the code underwriting a frog's ability to track flies can not be said to properly *represent* flies, design, history and intent to the contrary. Central to my argument will be a claim that code can have representational content, but it can't have representational targets. Without targets, of course, you don't have representation.

MDS entails a representation/code distinction at the boundary between the psychological and the sub-psychological/computational. Psychological states are available to the agent as reasons for actions and beliefs, sub-psychological states are not. Whole cognitive agents represent, sub-psychological processes do not. Indeed, there seems to be a consensus of sorts that a line needs to be drawn between different levels of representational sophistication: Rountree (1997) and Dretske's (1995) distinction between representation and belief, Cummin's notion of intenders (1996), Haugeland's 'resilient commitment' (Clapin, 2009 pp. 129-130), and Dennett's 'florid representations' (Dennett 2000), these all turn on some distinction between something like a computational action, and something like an agent's action (in a world with other agents). Here I am drawing that line between representation and code.

Ramsey (2007) argues that most attributions of cognitive representation are, in fact, applied to states that are not truly representational. Whereas Ramsey invites the conclusion that since so much cognition is not representational, therefore RTM is wrong (in that it doesn't describe what cognitive science is really doing), I'm taking the other tack, arguing that since so much cognition is not representational, therefore RTM is not circular.

But, in doing so, have I misconstrued or possibly hamstrung RTM? One might complain I've just replaced it with the "Code Theory of Mind", since I've just redefined all the interesting stuff that's doing computation and cognition in the brain, or, worse, removed the power that representation lends to the theory. Representation links what's going on in our heads with what's going on in the world and explains how cognitive

processes track content and preserve meaning. It's not enough that code has information about things, psychological states are *about* things. How can we explain this? I need to show that re-describing cognitive states as sub-representational code has not undermined this.

I'll quickly note one line of defense that can be found towards the end of chapter two on page 56 where I discuss eliminativism. Just because the states that subvene representation are not themselves representational, that does not entail that the representational states are not explanatorily necessary or causally efficacious. Argument for those conclusions can't be made on the basis of the lack of sub-psychological representational states alone.

The code/representation distinction can be defended against these charges by complimentary lines of argument. There are properties that representation must have that code can't, and, vice versa, there are properties of code that representation can't have. One of the critical distinctions between code and representation is that while representation requires a target, and has meaning, code does not. It may correlate with things, possess information about them, but they don't *mean* anything. The crucial question is whether or not this claim threatens the power of cognitive explanation. As I will argue, there are good reasons to believe that there are ways that representation and intentionality can be produced by states that lack these properties, but, more importantly, that states lacking these properties can nonetheless preserve and track intentional content.

Philosophers have argued that there is a major barrier to the proposal I am considering, that computation over code cannot be sensitive to global and context-

sensitive properties of representation, and that as a result it is impossible in-principle to construct a computational theory of cognition that can explain these aspects of psychological states. I will now describe and discuss both problems in turn. If either claim is correct, then it is not possible to maintain the code/representation distinction that I want.

3.1 Representation and Globality

Fodor (1983) (2000) argues that global properties of mental representation subvert computational accounts of cognition, since the syntax of the representation alone can't be relied upon to yield the data necessary for computation to track meaning. This is because Fodor thinks that this is all the mind can detect, and all the computational processes in the mind could access. For Fodor, syntax is a local property of whatever instantiates a representation, and if global properties are required to perform a cognitive task there is a problem. If the semantics outrun the syntax, then computation can't take it into account. Fodor argues that we can't avoid this by supposing that computation can take place on only the local properties of the representation, since the issue of globality is precisely that holistic properties not local to the individual representation are essential to the cognitive role of the representation.

The globality problem is closely related to the frame problem; in Zenon Pylyshyn's words, "Hamlet's problem: when to stop thinking. The frame problem is just Hamlet's problem viewed from an engineer's perspective" (Pylyshyn 1987 pp. 140). How can a seemingly unbounded domain of potentially relevant information be quickly and

accurately navigated without checking? Two kinds of frame problem can be discerned. One is the problem of planning, the question of how we determine all the relevant information required to implement a plan to pursue a particular goal. Secondly, there is the problem of belief updating: how do we determine which beliefs are updated as information is received and which are not? Fodor illustrates this (Fodor 1983) by imagining a robot dialing a phone. It must know which beliefs (out of *all* its beliefs) need to be updated and when, as well as which are invariant (under some set of conditions). Absent this, a robot dialing a phone doesn't know whether placing the call changes the number. "Since nothing in its belief complex averts the possibility of this kind of causal relationship [between placing the call and changing the number], the robot rechecks the number, begins the call, rechecks the number, and so on, effectively disabled by a real-time, frame-problem-based infinite loop" (Crockett ,1994, p. 74).

It is not always clear which frame problem Fodor has in mind, as Sperber and Wilson observe, "Fodor's [formulation of the frame problem] is the loosest and grandest reinterpretation of all" (Sperber and Wilson, 1996, p. 530). I think it encompasses aspects of both kinds I distinguished above. For instance, Fodor asks, "... what is a non-arbitrary strategy for delimiting the evidence that should be searched in rational belief fixation?" (Fodor, 1987, p.140). This more general formulation is closest to the instance of the problem I am concerned with here. If we need to know something about the representational properties of mental states to determine global properties, then code, lacking representational content, can't subvene representation. Representational targets are global in this way, as we saw in chapter 2; determining what a representation is about requires reference to metarepresentation at the psychological level. This account of

cognition becomes computationally intractable, because there is no way that code alone can be sensitive to context and yield enough information to support computational processes that can determine which global properties must be respected to preserve the representational content. If there is a large domain of things that could be relevant to running the code, and no way to limit that domain without reference to representational properties, the approach fails. As Fodor sees things, globality is a deep-running problem for classical computational theories of mind.

Fodor uses the term 'isotropic' to describe processes for which there is "no way to delimit the sorts of informational resources which [they] may affect, or be affected [by]" (Fodor, 1983, p. 112). Fodor distinguishes between input systems and central systems, characterizing input systems as modular and non-isotropic while central systems as fundamentally isotropic, quinean, and global. By 'quinean', Fodor means that the processes have "... certain vital epistemic properties defined over the system as a whole [...] the system as a whole [is] evaluated for consistency and coherence." (Bermúdez, 2009. p. 219). The rationality of central processes can't be determined locally, but only with reference to the whole system. Thus, the adequacy of local processes would seem to be determined only with reference to the whole system. The results of local processes can require adjustments to processes elsewhere in the system to maintain global rationality, but we can't know locally *which* processes these are. It is global systems that suffer (and somehow solve) the frame problem. Part of the job of specifying how input modules work is to ensure they do not require interaction with central systems. Input systems are modular, encapsulated, local and unproblematically algorithmic. Central systems are isotropic, global, computationally expensive, and liable to the frame problem.

I'd like to explore the idea that the code/representation distinction might run alongside the local/global one for things like meaning and mental content. Roughly, code doesn't need to be sensitive to representational content, and in fact can't be, you need the representational target to determine this, which code doesn't have. If there are independent reasons to claim that computation can indeed proceed via local properties without attending to the global ones, then, for at least the domain I'm worried about here, we have an answer to Fodor's problem.

3.2 Physical and formal syntax

Fodor's argument seems to block this strategy by definition. Computation over code can only operate with the syntax of representations, not the semantics, but syntax is a local property, and as a result it is not possible to explain how computational processes over code can cope with the global nature of representation. Fodor argues that computation in the brain can only operate with syntactic properties, and can't process semantic processes directly.

The problem is that Fodor is operating with an unnecessarily austere definition of the term 'syntax', an argument made by Brook (2009). Brook discusses arguments from Dennett and Fodor to the effect that if there are computational processes that produce semantics, they must rely on syntactic features of representations exclusively. "[T]he semantic content must be built out of the 'syntactic' (i.e., the non-semantic) in some way" (Brook, 2009 p. 267). It is not just that computation in the brain can't process semantics, but that it must do so *via* syntax. He notes that there has been a tendency in the literature to conflate a few different meanings of syntax, especially between syntax in the

linguistic sense (I'll call this 'formal-syntax'), and the broader sense that philosophers invoke when discussing this problem about deriving semantics from syntax. In this looser usage, the term means something like 'physical properties' (I'll call this 'physical syntax'), and it is this sense of the term that is used to pose the question of how we get global semantic properties from the physical properties of representations.

Brook argues that local physical properties do not exhaust the means by which computation over mental representation might access global properties.

> There are lots of other physically salient properties – order, shape, and spacing for example – that are not essential to MR's [mental representations], that are context-sensitive, and that computational processes could detect and make use of, including all of the relationships of each MR to other MR's. (Brook 2009 pp. 269).

That there may be strong arguments against formal syntax as a viable substrate for content is not fatal, as Brook shows that there are other physical properties of the representational vehicle that may be able to encode the information thought to be missing from syntax. Contra Fodor, these physically salient properties can indeed be context sensitive.[4] Fodor is worried that syntax alone can't support the global properties that mental representations possess, but Brook argues we have a much richer resource to draw on by way of a broader conception of physical-syntax. It therefore a mistake to claim that

[4] It is worth noting that this line of argument combines well with Brook's solution to the homunculus problem discussed in the end of chapter two. If RTM works best when it posits a single global representation with multiple objects, this broad construal of the representation gives us access to a very large set of these physical properties in the representational vehicle that can be exploited.

it is only local properties of the sort Fodor discusses that could make global information available. An example of a theory that exploits just this kind of physical-syntax to solve a global representational problem can be found in Clark's (2002) argument that internet search engines show us a way to analyze content without accessing meanings and solve the frame problem.

3.3 Physical syntax and the frame problem

Clark argues that algorithms such as Jon Kleinberg's (1999) show us how to interact with data in a way that is sensitive to representational content without actually knowing this content. Indeed, as I read it, it does this without having to access representational properties of the data directly. Clark describes the problem as global abductive inference, "... how to find the most relevant body of beliefs and information in a massive knowledge-base" (Clark, 2002 p. 119). What is important here is not the measure of relevance, which can be seen as a distinct problem, but that the meaning of the information and, thus, context sensitivity and globality, must be taken into account to solve the problem.

Kleinberg's procedure (hereafter KP) exploits the linked nature of information on the internet to produce result-sets based on search queries. Links are created between words and sentences in one document to other documents. These links are made with some kind of intent, and represent judgments of value and relatedness. A body of such linked documents such as the world wide web can be analyzed on the basis of this link structure. In a large enough environment, the problem of searching this body of information becomes intractable, because the number of results for purely pattern-

matching searches is too large to be useful. Kleinberg's solution is to employ an analysis of the structure of links in the result-set to produce a subset of more relevant and useful results than brute pattern-matching alone would provide.

Some documents in a result-set generated from matching patterns in the search query to patterns in the body of documents may have very few links to them from other documents in the web at large. This can be construed as an indication that they are less valuable. If the web contains few links to a document, this is an indication that the document has not been judged to be useful. In other cases, it may be that a result-set contains documents that, while there are many links to them, the documents that contain these links do not themselves match the query-pattern. In these cases we can conclude that they are not relevant to the query being performed, and remove the documents they link to from the result set. This is a search strategy that is sensitive to content without accessing it, and that produces very good results. For instance, it can help detect that a query for 'bike parts' contains a result-set that can be divided into two subsets, one with link structures relating to motorbikes, and another that has link structures relating to bicycles. Thus, a search for "Bianchi bike parts" can be disambiguated algorithmically; because Bianchi manufactures bicycles and not motorbikes the link structure will have a pattern related to bicycles much more strongly than it relates to motorbikes. KP can eliminate from the result-set documents that may mention all three terms, but which have patterns of links that deviate substantially from the majority of documents in the result-set, under the assumption that this deviance in link-patterns indicates irrelevance. Documents with the most links from documents that themselves match the pattern will generally be more relevant.

Clark argues that this model can be plausibly formulated in neuroscientific terms, which, if correct, is good news, but the details of this are beyond my purposes here. Worryingly, there is the appearance that the whole process simply offloads the hard work on the previous judgments made in creating the links in the first place. As Clark puts it, "... perhaps it is only because human brains can already solve the frame problem that we can set up the link structures that allow KP to work, thus rendering KP circular as a solution to the frame problem" (Clark, 2002, p. 17). Kleinberg notes, "... hyperlinks encode a considerable amount of latent human judgment ..." (Kleinberg, 1999, p. 6), thus, if it takes "... frame-problem solving humans to set up the hyperlink patterns ..." (Clark, 2002, p. 29), we may have solved an internet search problem, but we haven't come close to solving the problem of globality for cognitive tasks.

Clark argues that the network of information and links is created anarchically, with different agents creating and linking information for different reasons, and with different degrees of authority. Over time, as other agents interact with the information, and create further links, according to KP, it becomes possible to mine the structure of the links for information that they were never intentionally imbued with. The agents were not working together to link the information in a way that makes it useful for the target search task. This fact, for Clark, is the basis for claiming that the solution is not circular. "The link-structures emerge, in both web and brain, as a result of local decisions made on the basis of purely local interests and knowledge." (Clark, 2002). The argument here is that, while it is true that frame-problem-solving human intelligence was used to make the local decisions that are then used as the basis for KP, because it is not the semantic

content of the links representing these decisions that is salient for KP, that feature of the information is irrelevant.

However, if we used KP to solve the frame problem, and KP requires information created by solving the frame problem, this still appears to be circular. How did we solve the frame problem to create the information KP exploits? By way of analogy, consider the difference between the execution of a complex task in the early stages of learning it, and the unconscious skill that develops with expertise. In the early stages, we must consciously, carefully, and laboriously learn what information and feedback is relevant to performing the task. Once skilled, we can use this information without having to consciously access it, let alone reacquire it. It is the same information we initially had to work very hard to acquire, but represented in some way whereby it can be exploited easily. In developing a web page, an author must carefully and laboriously decide how to link it internally and externally to other pages; a slow, cognitively intensive isotropic process. One creates the page with a specific intent, for instance, to document a meeting, and in the process of doing so creates links to other pages. With KP, the outbound links in the page, and the external links to the page that may be created by others, reveal information about topics other than the meeting that was documented. Associations from linked words to other content are exploited by KP to solve semantic problems unrelated to those solved in creating the page. We used our frame-problem-solving cognitive capacities to document the meeting, but not to anticipate and provide the information that KP can glean from it after it has been created.

When we look at KP this way, I think it is clear that there would be circularity only if you had to solve the very same problem KP solves in exploiting the links when you created them, if each search problem KP solves using the document had to be solved in creating it. KP is circular only if it is used to solve problem X, and it is reliant on information that was created by frame-problem-solving humans *to solve X*. And that is not the case. KP doesn't consult actual semantic properties of information, but rather physically salient properties of the information that correlates with it.

While Fodor dismisses the possibility that computation over local properties of representations can respect the global nature of cognitive processes, I think we might now see a way out. Given Brook's argument about syntax, we have a much larger set of physical properties that we can exploit, and Clark's argument is one example of how we might use this larger domain of syntactic properties in algorithmic processes that can track representational content. The internet search engine can be sensitive to meaning without knowing it. This is an example of how code can subvene representation, and preserve and respect representational properties, without the need to attribute representational properties to the code itself.

3.4 Psychological descriptions of sub-psychological states.

If we are to maintain the distinction between code and representation, which I've argued runs alongside the psychological/sub-psychological distinction, then there would seem to be a problem with describing sub-psychological states in terms of their psychological uses. We should be able to describe computational processes in the brain in

non-representational terms, indeed, this was also part of the result of the argument for MDS. Since it is in fact usually the case that such states *are* described in psychological and representational terms, there is a problem for my argument. However, I think there may be a way to resolve this, based on Bergeron's (2008) arguments about cognitive working zones.

Bergeron argues that the specification of function in cognitive architecture is more accurate when it proceeds in terms of working zones, rather than specific uses. He distinguishes between cognitive working zones as domain-neutral computational capacities, and cognitive modules, as domain-specific competencies. "[C]ognitive modules are typically specified in terms of psychologically available operations—i.e., operations that ensue in psychologically available outputs, such as word/sentence recognition, speech production, face recognition, music perception. On the other hand, cognitive working zones will typically (if not almost always) be specified in terms of psychologically unavailable operations" (Bergeron, 2008, p. 70).

Cognitive modules are typically identified by their psychological uses. For instance, Broca's area is usually described as having the function of speech production, for which it is indeed used. As neuroscientists noticed more and more cognitive processes that seemed to rely at least in part on Broca's area, a more abstract functional specification was proposed, that Broca's area would be better described as a 'hypersequential processor', i.e. a cognitive working zone that performs the "... detection, extraction, and/or representation of regular, rule-based patterns in temporally extended events" (Fiebach and Schubotz, 2006, cited in Bergeron, 2008, p. 71). Bergeron

adduces further support for this approach by reference to research on facial recognition. The fusiform face area (FFA), in psychological terms, is a cognitive module for facial recognition. However, there is a growing body of evidence that, like Broca's area, the FFA plays a role in a wide variety of cognitive tasks. Bergeron argues that, "face recognition is a cognitive activity in which the fusiform face area (FFA) participates, not a cognitive activity that it performs" (Bergeron, 2008, p. 29).

If this is right, there is good reason to think that we can in fact avoid describing sub-personal processes in psychological terms. As I have argued, representations have irreducibly psychological properties that cannot be attributed to the code that subvenes them. If we need to describe the psychological uses of subvening processes, then the distinction between code and representation can't be maintained, because then the sub-psychological processes would in fact have to be described in terms that refer to the higher-level use and representational properties. I take Bergeron's argument as grounds to be optimistic that we can in fact keep these two things separate; for instance, we can claim that representing faces is something that agents do, but not cognitive modules like the FFA.

We saw two arguments that support my claim that cognitive processing can be sensitive to global properties, while taking place over a sub-representational substrate that does not need to attend directly to the global properties of the cognitive task. If, as Clark, argues, computation can be sensitive to content without knowing it, and, as Brook argues, that formal syntax is not the only physical property of the subvening code that might bear information, then we can see how code, minus the properties of representation proper,

might still be powerful enough to explain cognition. I think this casts doubt on the alleged impossibility of local computation respecting the upstream globality of the cognitive role it plays. As I read Bergeron's argument that cognitive modules are best described non-psychologically, this further supports the claim that we ought not invoke representation in explaining sub-psychological processes, as there are general reasons why doing so is unmotivated. I think there is reason to be optimistic that the code subvening representation doesn't have to make reference to global and representational properties. Indeed, we are better off if we don't posit this, which I take to be support both for the distinction itself and its validity.

3.5 A Counterargument from Pragmatics

Recanati's work on pragmatics (Recanati 2004) may offer a different counterargument to the code/representation distinction that I am defending. Recanati argues that the logical structure of a proposition cannot be algorithmically determined because it is not just the meaning of an utterance that is context sensitive but the logical form itself. Specifically, the place and number of indexical variables in a proposition is context sensitive. If he's right, then early stages of linguistic processing are isotropic in that you can't deal with the form without dealing with the content. If so, you can't do semantics distinct from pragmatics, and the position I'm advocating with regard to globality, code, and representation is undermined, because you can't process code without dealing with the representational properties at higher levels.

Before I discuss this, I'll try to clarify some of the terminology from a few different domains that overlaps here. We can distinguish two meanings of 'syntax':

Physical Syntax – All the physically salient properties of a representational vehicle

Formal Syntax – Formal structure in the linguistic sense

In philosophy of mind, syntax is often contrasted with semantics, which is usually construed broadly to refer to "meaning". However, in philosophy of language, we have a threefold distinction which introduces the notion of pragmatics:

Syntax – orthography, grammar (formal syntax)

Semantics – meaning of entities that is constant across utterances that can be used to build propositions.

Pragmatics – context-sensitive meaning of these propositions.

Semantics is supposed to yield meanings that can be reliably re-used across utterances, whereas pragmatic processing deals with utterance-relative context-sensitive properties. Syntax in this context does not mean physical syntax, but formal linguistic syntax.

In light of the arguments in the preceding chapters, we can further analyze semantic properties in terms of semantic content, which is what the representation refers to, and semantic targets, which is what the representation is *supposed* to refer to.

If there is no viable semantic/pragmatic distinction, then we have a frame problem, and communication looks to be very difficult to explain, because all language processing will have to deal with context-sensitive, global processes. Getting targets out of semantic processing would help with this, but this is exactly what Recanati thinks we can't do. Although most of the preceding discussion of syntax was about physical syntax, Recanati's argument, which is described generally as 'contextualist', refers to 'formal syntax' when it uses the term syntax.

Kent Bach coins the term "contextualist platitude" to describe the increasing recognition that "... what a sentence means falls well short of what a speaker means in uttering it" (Bach, 2005, p. 22). The platitude itself is not controversial, it is the question of how we can accommodate it that is difficult to answer. It presents a dilemma; either sub-propositional utterances like (1) "Nina has had enough" are in need of contextual enrichment, in which case we run into the problems Recanati raises, or, such sentences are in fact all that semantics needs to deliver to downstream pragmatic processing.

Cappelen and Lepore (2005) articulate a theory of semantic minimalism which grasps the second horn of the dilemma; arguing that the sentences like (1) are in fact propositions after all. Nina has had enough iff < Nina has had enough>, whatever that

might mean. This approach has the virtue of obviating the need for semantic processing to be sensitive to targets, at the expense of a seemingly untenably loose definition of what it is to be a proposition. However, if their position can nonetheless be defended, then we don't need anything from context to specify how the sentence works, so we can safely do semantics while letting pragmatics work out what the sentence actually means. But can we accept this loose definition of propositions?

Cappelen and Lepore's argument is best illustrated by their reaction to a suggestion offered by Bach that would appear to be very attractive to their theory. Bach denies that the basic unit must be a proposition at all, and defends an approach he calls radical semantic minimalism (RSM), which "... does not imagine that sentences that intuitively seem not to express propositions at least express 'minimal propositions'." ... instead, RSM "... says that the sentences in question are semantically incomplete" (Bach, 2006 p. 437). Instead of calling them minimal propositions, which then requires defense of their status as propositions, Bach calls them 'propositional radicals', propositions that are missing important features. Bach argues that so-called 'Propositionalism', the belief that every sentence expresses a proposition, can be abandoned; some sentences just express radicals. If this approach can work, it would undermine the contextualist project, inasmuch there is no need for primary pragmatic processes to fill in the constituents of incomplete sentences to render them as propositions. It would seem attractive to the semantic minimalist, as they would not need to insist that minimal propositions are really propositions.

It would certainly appear that semantic minimalism is improved by taking on board Bach's theory. However, Cappelen and Lepore protest extensively that they cannot

accept RSM. They insist that any kind of incompleteness must invariably lead to contextualism. As I read their position, the concern is that no level of specificity can be shown to represent the 'final' proposition, that further enrichment is always possible, and that therefore to allow for incompleteness is to leave the door open for contextualism. In the absence of "... a principled distinction, that is not context sensitive, between 'Nina has had enough' and 'Nina has had enough pasta', i.e. a criterion by which the former is incomplete and the latter complete..." (Cappelen and Lepore, 2007, p. 260) we will not be able to tell when a proposition is complete and when it is a radical. It is thus preferable to accept that the minimal proposition is the complete proposition that pragmatic processes then treat as input to determine meaning for.

Cappelen and Lepore claim that the minimal proposition can function as 'shared fallback content' in cases where contextual information is unavailable or misconstrued. At the very least, we know what was said, even if we don't know what it means. The contextualist can object that if the minimal proposition radically underdetermines speaker meaning it can't function as fallback. Cappelen and Lepore reply, "There's a lot of stuff to talk about in the universe. The proposition semantically expressed pares it down considerably". (Cappelen and Lepore, 2005, p. 185).

They argue that the minimal proposition restricts the domain of possible meaning just by reducing it to whatever the terms could possibly mean. A proposition including the term 'red' is at least restricted by whatever bounds the class of red things. We can concede that the job of the minimal proposition is not to generate word meanings, but just needs to restrict the domain of probable reference. Conflation of semantic and pragmatic

processes is exactly what semantic minimalism ought to avoid. If semantics requires access to 'all this stuff in the universe', without anything in the syntax to restrict the domain, communication and language processing look improbably difficult. Metarepresentation at the pragmatic level can narrow it down, but absent that, we ought to accept indeterminacy with regard to the meaning of the minimal proposition.

What is relevant to the overall position I am defending here is the extent to which the minimal proposition is missing the representational target. It is the target, which is generated by speaker intent, that disambiguates the proposition, and semantic minimalism does not admit such disambiguation at the level of semantic processing, at which we just don't know what "Nina is ready" is being used to represent. The minimal proposition, lacking a target, also lacks determinate content, which it doesn't obtain until it is interpreted. This approach keeps the representational and psychological properties of the utterance at the level of pragmatic processing, and thus out of the syntactic and semantic levels. As long as targets aren't involved in syntactic and semantic processing (as they would be if Recanati is right), then they can work computationally and algorithmically.

Minimalism seems implausible only because there is intuitive appeal in denying that the minimal proposition could possibly bear the kind of meaning we expect that the enriched proposition will. But I think this is the result of an unfounded optimism that the enriched proposition, on its own, more able to yield the rich meanings involved in language. Neither tells us very much about what is really meant.

In "The alleged priority of literal interpretation" (Recanati, 1995), Recanati cites data from psycholinguistic experiments that show that processing time does not appear to

vary between literal and non-literal utterances. Recanati argues that, "... this casts doubt on the standard model, according to which the literal meaning of a non-literal utterance is processed first" (Recanati, 1995, p. 207). I think that in fact the opposite would be true. Semantic minimalism necessarily posits that there is no difference between processing literal and non-literal utterances, or between sense and nonsense, or sentences with and without content. All semantics does is produce the minimal proposition, it doesn't know or care if it is literal or not. In fact, if there was a difference, it would suggest that contextualism was right, that semantics was affected by psychological and representational properties. What we ought to expect is that experimentation would show that there is a difference between processing syntactically regular sentences and those that violate syntactic (eg. grammatical) rules, but not between meaningful and non-meaningful syntactically valid sentences.[5]

I've argued that we can be optimistic that semantic minimalism offers a viable way to keep meaning and globality out of semantic processing. Contra Recanati, it is at least possible that we can do semantic processing without having to refer to the content and meaning of the propositions. Semantic processing is at least a candidate for a stage at which algorithmic local processing can do important cognitive work without dealing with meaning and content in ways that are isotropic and global and that refer to psychological and representational properties. While such processing can't yield the kind of meaning we might expect it to, I've given a number of arguments to the effect that yielding meaning is more than we want it to do.

[5] Featherston (2001) has conducted experiments that support this.

Most importantly, one of the characteristics of the minimal proposition is that lacks representational features, especially a representational target. The minimal proposition is a representational vehicle that can code for content, and be used by the mind as a representation, but is not itself representational. This is why we don't need to enrich it to resolve meaning ambiguity, and why Capelin and Lepore are right to resist RSM so strenuously. Individual sentences and propositions are never going to be able to support on their own the kind of meaning that is actually a broader property of psychological and representational states. As I argued in chapter two, this kind of more determinate meaning relies on metarepresentation, which would have to be a pragmatic process, since it does not involve only the proposition expressed.

3.6 Neurosemantics – Ryder's SINBAD theory

One of the claims in the overall argument I've been defending is that the job of an NTR isn't to naturalize particular representations, but the representational capacity as a whole. As I discussed in chapter one, this might appear to weaken the ensuring NTR, but I've argued that in fact it does not, that for reasons such as indeterminacy and precision, stronger NTR's aren't going to give us an adequate picture of the phenomena of representation as we find it in human cognition. This more modest conception of the requirements of an NTR relies on the ability to maintain a distinction between sub-psychological code and psychological representation. I've considered a number of arguments in this chapter to the effect that this is not possible and suggested ways to overcome them.

There are two things in particular that I think are gained by this approach. First, the burden of facing up to problems of representation in the construction of sub-psychological theories is lifted. There is no reason to have to account for representational properties at these lower levels, or to construct theories that aim to support NTRs in the strong sense. Secondly, although I take this point to be somewhat less secure, there is at least reason to reject pessimism that there is some deep foundational problem with psychological explanation. If the arguments I've offered here are sound, then I think psychological explanation can bottom out in representation without having to offer further explanatory linkage to lower levels. However, this is a large problem, and depends on arguments I've been able to consider only briefly here, such as those relating to reducibility and mental causation. So I'll conclude by focusing on the first point. In particular, I will describe and discuss a new approach to naturalized representation that has been developed under the banner of "neurosemantics". I'll argue that this is promising, and that it is bolstered by not attempting to meet all the requirements that philosophers have thought an NTR must, especially those relating to representation at the psychological level.

Neurosemantics introduces a constraint on causal theories of content, that they be specifically accountable to the neurobiology of the brain, and to physiology generally. There are a number of theories that have been developed under the umbrella of neurosemantics, here I will focus on one developed by Dan Ryder (Ryder, 2004), a causal and teleological theory of representation, based on an account of modeling functions in cortical networks (Ryder & Favorov, 2001). I focus on this one because it explicitly targets psychological level representation, and it uses teleosemantics as an essential part

of the theory. The SINBAD model (Set of Interacting and BAckpropagating Dendrites) claims that, "... networks in the cortex have a powerful tendency to structure themselves isomorphically with regularities in their environment" (Ryder, 2004, p. 212). The output of a neuron is determined by that of its dendrites. The dendrites in cortical pyramidal cells adjust their output to maintain equal contribution to the cell's overall output by adjusting sensitivity to the excitatory and inhibitory inputs they receive. As the dendrite receives variable input, it will then adjust its sensitivity as required to maintain equal contribution to the cell's output. In maintaining this equilibrium, the dendrite effectively finds a mathematical function to approach a steady state, with feedback from a backpropagating signal from the cell's output.[6] Dendrites will be able to synchronize if the stimulus the cell receives is regular, if there are correlations between the inputs each dendrite receives – so there must be pre-existing correlation in the inputs, which is claimed to correspond to regularities in the distal stimulus – in effect, they will tune to individuals and natural kinds.

Networks of SINBAD cells are claimed to be able to find functions relative to macro level environmental regularities, to become dynamically isomorphic to aspects of the organism's environment. Ryder notes that isomorphism alone cannot amount to representation, since a structure may be isomorphic to any number of states of affairs. He invokes teleofunction to underwrite the claim that SINBAD networks are representational via their function, it is not just *that* they are isomorphic, but that they have the *function* to be isomorphic. "An etiological account of teleofunction cashed out in terms of

[6] This may be a good example of an account of isomorphism that does not refer to users and interests.

evolutionary history delivers us normative isomorphism, and thus representation, in cortical SINBAD networks" (Ryder, 2004, p. 225) The presence of SINBAD networks in the brain is to be explained in terms of their evolutionary value as a 'predictive trick' – as such, to be useful, the isomorphism must be to real environmental regularities. They represent the explanatory source of the regularities they model, and they misrepresent when they respond to a source other than that they were selected to respond to. The explanatory sources of the regularities that are modeled are things in the world. Ryder argues that selection pressure would thus direct the isomorphism to real kinds.

In staking a claim for SINBAD network models as representational via appeal to teleofunction, Ryder argues that we can call a detection of a dime by a penny detector an error, "Because it is pennies, and not dimes, that explain the cell's achievement of (imperfect) equilibrium and its 'predictive' abilities, its job is to correspond to pennies, not dimes. Therefore it represents pennies and not dimes." (Ryder, 2004, p. 231). This is, of course, reminiscent of Dennett's 'two-bitser' example discussed in chapter two. That it was pennies and not dimes that made valuable a particular model doesn't mean it was pennies qua pennies that warrant this. Being a penny is accidental to the penny-features that were of use to the system. It might be that in the course of reliably modeling penny-features pennies are modeled, but if a dime has saliently similar features, then there is no error, except from some external perspective in which is it useful to distinguish pennies from dimes. Such a distinction would be based on features of pennies and dimes that are not of interest (possibly not perceptible, or relevant) to the system in question. Ryder argues that SINBAD is not exposed to this indeterminacy problem, because, "... it relies on evolution only to narrow down the contents of mental representations to sources of

correlation; their specific contents is derived, in a very different way, from learning history" (Ryder, 2004, p. 236).

Since isomorphism is not sufficient for representation, teleofunction is used to pinpoint isomorphism *to* sources of correlation as the kind of isomorphism SINBAD networks form. This point relates to one Searle has urged strongly, that there is some level of description that can describe any two structures as isomorphic (Searle, 1992). Because SINBAD networks are described as general learning devices, they are not selected for any particular representational capacity, as a result, a teleosemantic account can't be linked to any particular content-vehicle relationship. SINBAD cells are selected to model sources of correlation, and to develop sensitivity to as many forms of evidence for the presence of the source of correlation as possible. So teleofunction establishes that the job of SINBAD networks is to develop isomorphism to real kinds, but the particular content of a particular network is established via learning.

In the case of the frog fly/food/dark dot disjunctive problems, how might we apply the SINBAD model? Fodor, in discussing Millikan, raises this problem, "... appeals to mechanism of selection won't decide between reliably equivalent content ascriptions" (Fodor, 1990, p. 73). On Ryder's model, because we are modeling sources of input regularity, and because teleofunction does not determine content, we can rule out 'nutrient' as representational content, and related problems in the form of benefit/content indeterminacy. With regard to the black dot, we can't use the standard teleosemantic answer to the disjunction problem here, since SINBAD explicitly does not use teleosemantics to determine content. An argument might be mounted along the lines of

'the black shadow/dot does not represent a proper source of correlation', but I think this is ultimately unpromising, especially when faced with ambiguity between kinds and sources.

Ryder argues that because the evolutionary value of the SINBAD network's representational ability is explained with reference to the value for the organism of predictive modeling, content that is not predictive can't be the proper function of the network. While this might help with the fly/black dot problem, it leads to other worries about how it represents non-predictive content, and what kind of definition of 'predictive' is required here. If at all, SINBAD networks overcome feature/object indeterminacy by stipulation. They represent not correlations but sources of correlations, because the correlations themselves are only useful to an organism as indicators of the source. Thus, they represent flies and not fly features because modeling fly features is only useful by virtue of assisting the organism in interacting with flies. It is not clear how this account can work for other kinds of content, and for processes like 'edge detection'.

3.7 Usher's Critique of Sinbad

Marius Usher, who has developed another neurosemantic theory, has criticized the role teleosemantics plays in the SINBAD model (Usher, 2004). I am interested in the extent to which he critiques the role teleosemantics plays in SINBAD, as well as the sense in which the dispute concerns representation at sub-psychological levels. Usher claims that SINBAD inherits problems faced by teleosemantics generally, and fails to overcome them. Usher discusses Davidson's (1987) "Swampman" thought experiment, claiming that it can be applied to teleosemantics. States that lack the correct causal and

historical antecedents cannot be said to be representations because they lack the properties that such theories claim constitutes representation. Davidson's example involves the accidental spontaneous creation of a molecular duplicate of himself. The worry is that we are unable to say it has mental representations, despite being an exact copy, because it lacks the right causal history. Usher objects that SINBAD faces the same problem, in that it is history that is part of what underwrites the claim that it represents. However, in the sense that having the right history is important to SINBAD, it is to explain why the network responds to natural kinds, not how or that it does. It would still continue to do so in the molecular duplicate, so I think SINBAD can be understood in ways that don't motivate the kind of intuitions Swampman generates.

Usher's primary objection is based on indeterminacy problems for teleosemantics, the kind of problem discussed in chapter 2. However, because SINBAD's content fixation is not via teleosemantics alone, but also via learning, it may in fact not be vulnerable. Teleofunction establishes that there is representation, not particular representational content. Usher claims that this move is unsuccessful, specifically that the partial decoupling of content fixation and teleosemantics is in effect total, which makes "... biofunctionalism totally redundant to the theory" (Usher, 2004 p. 243). Usher's argument for this point is rather compact, so I will attempt to elaborate it.

Suppose SINBAD network N represents natural kind K, and the relationship between K and N is ultimately causal. This, as discussed in earlier sections, means we need to make room between K and N for misrepresentation, and we need to resolve indeterminacies such as those between proximal and distal stimulus. SINBAD does this

by positing a teleological constraint, that biological selection has given SINBAD networks the function of representing real kinds versus feature disjunctions. The particular representational content of N depends on how the system has recruited it, via a learning process. If recruited to represent flies, it will learn to respond to fly-features reliably enough to indicate flies. Usher worries that, "... the same cell could have come to correspond (depending on random pre-learned synaptic connections) to kind-L instead" (Usher, 2004 p. 245). It is not entirely clear what Usher is driving at with this claim. I think the best expression of this worry might be something like, 'there is no reason why this is a capacity to represent K rather than L given that both might yield reliable predictions'. However, the capacity and the value of the capacity are separate issues, and the capacity couldn't be valuable if it wasn't already representing. So, in some sense, whatever biofunctional role we grant to the capacity presumes that it is already representing. As such, and this is Usher's conclusion, why not characterize the representation in terms of the capacity alone, and drop the predictive/biofunctional component. This is what grounds Usher's claim that biofunction is ultimately redundant in the model.

Usher's positive claim is that statistical dependence is a better theory of representational content, and that SINBAD could be formulated in a way that is compatible with it. As described by Usher (Usher, 2001), the theory is a form of information semantics, which uses Shannon Mutual Information, a statistical method to measure the mutual dependence between different variables, to arrive at a probabilistic measure of the degree of information shared between a representational vehicle and content. The representational relationship is understood as statistically optimal (not

necessarily maximal) correlation between the vehicle and the target. As such, it is a causal theory, but, as with SINBAD, causation alone does not determine content. Statistical dependence offers a general theory of representational content, grounded by an account of the responsivity of individual neurons. At higher levels of organization, the content of a representation is that referent (stimulus) that has the most mutual information in common with the vehicle. Thus, a mistaken perception of a cat as a dog is mistaken because, overall, the dog has more mutual information with the vehicle than cats do.

Ryder argues (Ryder, 2006) that statistical theories generally have difficulty distinguishing content from epistemic conditions. The highest correlation will be between 'dogs in good light with no fog' with the dog vehicle, rather than just dogs. There are problems with approaches that seek to exclude such conditions, since in some cases conditions may well be part of the content. Eliasmith claims that the, "... the real-world development of an animal's conceptual categories ..." (Eliasmith, 2005a) would prevent possible but unlikely epistemic conditions becoming part of the content; but his argument is too brief to assess. The proximal/distal stimulus problem arises here as well, since the proximal stimulus would have more (or at least equal) mutual information in common with the vehicle than the distal stimulus.

A related concern has to do with the kind of content and referent attributions that we might make at various levels of granularity. At the level of single neurons, for many representations it would seem there isn't any 'real content' until a fair level of complexity is achieved. The content 'ripe apple', at the level of whatever individual neurons

instantiate it, is going to have content that involves many other representations, making it harder to demarcate which states should be considered part of its representational vehicle.

Notice that the problems here all relate to representational properties of sub-representational code. Neurosemantics aims to explain how cognition in the brain implements representation. Why should the intermediary states be expected to have representational properties too? The question as to whether or not statistical dependence can be integrated with SINBAD is beyond my purposes here, as would be a full exposition of both projects. My interest here is the extent to which the dispute itself may be unmotivated. If we don't have to attribute representation to sub-psychological states, the subject for much of the disagreement vanishes. I discuss neurosemantics as an example of a theory of representation directly aimed at the sub-psychological processes that underwrite psychological representational capacities. In light of the arguments that I have defended, I would suggest that both neurosemantic projects could be articulated without positing sub-psychological representation, and that indeed they would be improved by doing so. There are problems about representation, like the ones I discussed in chapter one, that processes at the code level simply don't need to solve. Claims like those used to defend SINBAD from fly/black dot indeterminacy are therefore unnecessary. And the kind of arguments needed to solve them at the code level introduce constraints that make other problems harder to solve, for instance, tying representation to natural kinds will make it harder to explain other sorts of content. If the approach to representation that I've been defending is correct, then neurosemantics could have the aim of explaining the mechanisms of representation without needing to posit representation within the model. As a result such theories would not need to prove that

sub-psychological states in the model meet representational criteria, and solve problems that are typically raised in objection to theories of representation. Neurosemantics can be a code theory on top of which representational theories can be built.

Recalling the distinction in chapter one between the two ways that teleosemantics can invoke history, we can see that SINBAD equivocates between them. On the one hand, SINBAD networks exist because they are general modeling devices and this has selection value. On the other, it is claimed that individual networks have teleofunction to respond to particular kinds, which I have argued is problematic. There's lots of representation that doesn't have selection value directly. Also, as I argued in chapter two, indeterminacy is an inherent feature of representational content, and in sub-psychological states it is unavoidable. The demand that either approach yield determinate content is thus unwarranted; therefore indeterminacy is not a dangerous objection. We can relegate the job of explaining representation of abstract concepts and unacquainted content, which is otherwise thought to be a problem for teleosemantic theories, to the level of representation. Whether by learning or teleofunction, it is very hard to see how SINBAD networks could have the function of representing this kind of thing.

3.8 Conclusion

In this chapter, I've defended the distinction required by MDS between code and representation. The distinction doesn't undermine RTM, in fact, it bolsters it in a number of ways. Aside from the main matter of interest here in defending RTM from CP, it also provides grounds for avoiding the frame problem, and accommodates recent work such

as Bergeron's that discourages psychological description of cognitive states and activities. I've shown that arguments like Recanati's are potentially problematic for this approach, but that there are good grounds to suppose that his arguments can be overcome.

Finally, I've argued that this approach can be used to help solve problems for theories like neurosemantics that attempt to explain how the brain actually implements our representational capacities. I think there is much to be gained by keeping these two concerns, representation and code, distinct. I hope to have demonstrated that excising representation from sub-psychological explanation is not just warranted because of the problems with NTR and the demands of MDS. It turns out to be independently useful, and consistent with other results. Attributing representation to these states is not just inaccurate, it is a burden.

A few issues have emerged that I think are potentially interesting as topics for further research. First, there are arguments that I have relied upon here, but that I did not properly explore or defend. My discussion of Bergeron's work regarding the re-description of brain functions in non-psychological terms was quite brief. Further argument and research is required to establish that this is viable in all cases. If it should be the case that there are cognitive functions that cannot be described in this way, that would seem to undermine my argument. Similarly, the solution to what I called the 'upper' homunculus problem relies on Brook's argument concerning self-representing global representations. I briefly sketched this argument, but a proper defense would require more elaboration and consideration of objections, for instance those that claim that aspects of the mind that can't be construed representationally. The same holds true

for my discussion of pragmatics. This is a large issue and there are further arguments against semantic minimalism that would have to be considered. At best I've indicated one way these arguments can hang together and be applied to the problem of representation, but more work is required to secure the conclusions I defend.

Neurosemantics, at least in part, targets the problem of how the brain generates meaning and representation. However, the theories on offer still posit sub-psychological representational states, or so I have claimed, and I have argued that banishing sub-psychological representational attribution would bolster this research. This claim requires proof. It would be worthwhile to review in greater detail the literature on the neurosemantic project and to analyze the extent to which commitment to sub-psychological representation can be weakened. If it can be, there is then the question as to whether the theory is improved by doing so. Would this solve problems the theory otherwise faces? I have suggested that this is likely but a great deal more argument and research is required to establish this.

There is a philosophical problem concerning the explanatory role of mental states that is connected to the question of mental representation. The position I have advocated does leave open the possibility that the attribution of representation is instrumental. As I mentioned in Chapter 2, the problem of mental causation and the status of folk psychological explanation arises when we find reasons to doubt that mental states are causally efficacious qua mental states. If the position I have defended has merit, it has some bearing on this problem. For instance, if it is the representational character of these states that is explanatorily essential, then by eliminating representation in subvening

states, some arguments for eliminativism are blocked. You can't get by with lower level explanation because there is no representation there. Eliminativists argue that intentional explanation can be replaced by physical descriptions of the subvening states and processes. My argument introduces at least one constraint on such an approach. If intentional ascription is essential to the explanation, then you can't get that from subvening states alone.

I have had a lot to say about the extent to which it the user that makes something representational. However, there remain questions about the particular usefulness of some kinds of things as representations of their targets. There is something about a map of Houston that makes it better for getting around Houston than the map of Baghdad. I argued that content is not a property of the map itself, but a matter of the metarepresentation involved in the production and use of the map. We know what maps are for, and we know why they are made, and thus know that we can reliably use them for certain purposes. However, there is still something about the one map that makes it more suitable for one purpose than the other. I argued that isomorphism alone, in isolation from use, cannot ground claims about representation. But it still would appear that there is some kind of relationship between the representation and representata that we need to explain. I have focused on the use of representation, but I think further research into the structure of representation would be worthwhile. I argued that we ought not conflate history and vehicle structure with meaning, but, for representational vehicles like maps, it would seem that there is something about its history that is still important. It may be possible to argue that the map represents its own history and that this is partially what makes it usable at all. This problem demands further work.

Finally, and more generally, I think that problem of explaining mental representation relates to some important questions about misrepresentation that merit further thought. We can think of the cognitive system as one that produces representation. It does this automatically and unrelentingly. Sometimes this goes wrong. In the construction and maintenance of representations of our selves and our world, the brain relies on tricks and heuristics that can make mistakes, that can be exploited, and that can be self-defeating. This is, of course, largely a problem in the domain of psychology. However, if we can better understand how the brain implements meaning and representation, we may be able to generate insights into how and why it can go wrong. I argued that teleosemantics can ground representation generally as adaptive and truth preserving, and, indeed, representation generally is, and, as a result, generally allows us to navigate the world successfully. But there is enough evidence that it frequently does not that this becomes an important question. McKay and Dennett (2010) discuss the evolutionary aspect of this problem, but I think there is much more interesting work to be done. There is a lot of misrepresentation in our minds, such as beliefs that are untrue and that can be harmful to us. This can be the basis for an objection to any theory that attempts to ground representation teleologically, so attention to the issue is required. But, more importantly, better understanding of mental representation may be able to shed light on the problem.

References

Agar, N. (1993). What do frogs really believe? *Australasian Journal of Philosophy*, 71:1, 1-12.

Allen, Sophie (2010) Can Theoretical Underdetermination support the Indeterminacy of Translation? Revisiting Quine's 'Real Ground' *Philosophy*, 85, 67-90

Adams, F. & Aizawa, K. (1992), 'X' Means X: Semantics Fodor-Style *Minds and Machines* 2:2, 175-183.

Adams, F. & Aizawa, K. (1994), Fodorian Semantics In Stich, S. and Warfield, T. (eds.) *Mental Representations.* Oxford: Basil Blackwell, 223-242.

Adams, F. & Aizawa, K. (1997), Fodor's Asymmetrical Causal Dependency and Proximal Projections, *Southern Journal of Philosophy,* 35, 433-437.

Bach, Kent (2006) The Excluded Middle: Semantic Minimalism without Minimal Propositions *Mind and Language* 21:5, 626-628.

Bach, Kent (2005) Context ex Machina' in *Semantics versus Pragmatics* London: Oxford University Press

Bermúdez, José Luis (2009) Jerry Fodor: Philosophy of mind and cognitive science in C. Belshaw and G. Kemp (Eds.), *Twelve Analytical Philosophers* Blackwell, 115-133

Bergeron (2008) *Cognitive Architecture and the Brain: Beyond Domain-specific Functional Specification* Vancouver: University of British Columbia

Brentano, F. C. (1874/1973). *Psychology from an Empirical Standpoint.* London. Routledge.

Brook, A. (1994). *Kant and the mind.* New York : Cambridge University Press.

Brook, Andrew (2009) *The Possibility of a Cognitive Architecture* Computation, Cognition, and Pylyshyn, Boston:MIT Press pp. 259-281.

Brook, Andrew and Stainton, R.J (1997) *Fodor's New Theory of Content and Computation* Proceedings of the eighteenth annual conference of the Cognitive Science Society: July 12-15, 1996, University of California, San Diego

Cappelen, H. and Lepore, E. (2005) *Insentive Semantics* Oxford: Blackwell

Cappelen, H and Lepore, E (2006) *Reply to Bach* Mind and Language 21:5 626-628.

Chomsky (1969) *Quine's empirical assumptions* Synthese 19:1 53-68

Clapin, H. (2002). *Philosophy of Mental Representation*. Oxford: Clarendon Press.

Clark, A. (1997). *The dynamical challenge* Cognitive Science: A Multidisciplinary Journal 21:4 461-481

Clark, A. (2002) *Local Associations and Global Reason: Fodor's Frame Problem and Second-Order Search* Cognitive Science Quarterly 2:2 115-140.

Crockett (1994) *The Turing test and the frame problem: AI's mistaken understanding of intelligence* Norwood NJ: Ablex.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge:MIT Press.

Davidson, Donald (1987). *Knowing One's Own Mind* Proceedings and Addresses of the American Philosophical Association, 60, 441-58.

Dennett, D. (1969). *Content and consciousnes*. London: Routledge & Kegan Paul.

Dennett, D. (1978). *Brainstorms*. Montgomery, Vt: Bradford Books.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, USA: MIT Press.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little Brown.

Dennett, D (1995). *Darwin's dangerous idea: evolution and the meanings of life*. New York: Simon & Schuster.

Dennett, D (1995). *Darwin's dangerous idea: evolution and the meanings of life*. New York: Simon & Schuster.

Dennett, D (1998). *Brainchildren*. Cambridge: MIT Press

Dennett, D (2000). *Making Tools for Thinking* Metarepresentation: A Multidisciplinary Perspective. New York: Oxford University Press.

Dretske, F. (1993). *Misrepresentation* Readings in Philosophy and Cognitive Science Cambridge: MIT Press 297-313

Dretske, F. I. (1995). *Naturalizing the Mind* Cambridge: MIT Press.

Eliasmith, C. (2005a). *Neurosemantics and Categories*. Handbook of categorization in cognitive science, 1035 (p. 1054). Amsterdam: Elsevier Science.

Eliasmith, C. (2005b). *A New Perspective on Representational Problems* Journal of Cognitive Science, 6:2, 97-123.

Featherston, Sam (2001) *Empty categories in sentence processing* Benjamin:2001

Fiebach, C. J. & Schubotz, R. I. (2006). *Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains?* Cortex, 42:4, 499-502.

Fodor, Jerry (1983) *The Modularity of Mind* Cambridge: MIT Press.

Fodor, J. A. (1987a). *Psychosemantics* Cambridge: MIT Press.

Fodor, J. A. (1987b). *Modules, frames, fridgeons, sleeping dogs, and the music of the spheres.* Modularity in knowledge representation and natural-language understanding, 139-149.

Fodor, J. A. (1990). *A theory of content and other essays.* Cambridge: MIT Press.

Fodor, J.A. (2000) *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology* Cambridge: MIT Press

Gettier, Edmund L (1963) *Is Justified True Belief Knowledge?* Analysis 23 121-123.

Kleinberg J.M. (1999) *Authoritative sources in a hyperlinked environment* Journal of the ACM (JACM) 46:5 604-632.

Lettvin, J. Y., Maturana, H. R., Mcculloch, W. S., & Pitts, W. H. (1959). *What the Frog's Eye Tells the Frog's Brain.* Proceedings of the IRE, 47:11, 1940-1951.

McKay, R.T. and Dennett, D.C. (2010) *The Evolution of Misbelief* Behavioral and Brain Sciences 32:6, 493-510

Millikan, R. G. (1989). *Biosemantics* Journal of Philosophy, 86:6, 281–297.

Millikan, R. G. (2000). *Naturalizing Intentionality* Proceedings of the World Congress of Philosophy, 9, 83-90.

Neander, K. (2008). *Naturalistic Theories of Reference* The Blackwell Guide to the Philosophy of Language Blackwell Cambridge, USA 374-391.

Pietroski, P. M. (1992). *Intentionality and teleological error* Pacific philosophical quarterly, 73(3), 267-282.

Pylyshyn, Zenon W (1987) *The Robot's dilemma: the frame problem in artificial intelligence.* Norwood NJ: Ablex.

Putnam, H. (1990). *Representation and Reality.* Cambridge: MIT.

Quine (1960) *Word and Object* Cambridge: MIT.

Quine (1969) *Reply to Chomsky* Words and Objections: Essays on the Work of WV Quine 302-311.

Ramsey, W. (2007). *Representation reconsidered.* Cambridge: Cambridge University Press.

Recanati, F. (1995) *The Alleged Priority of Literal Interpretation* Cognitive Science 19:2 207-232.

Recanati, F. (2002) *Unarticulated Constituents* Linguistics and Philosophy 25: 299-345.

Recanati, F. (2004) *Literal Meaning* Cambridge: Cambridge University Press.

Recanati, F. (2005) *It is Raining (Somewhere)* from http://jeannicod.ccsd.cnrs.fr/documents.

Rountree, J. (1997). *The Plausibility of Teleological Content Ascriptions: A Reply to Pietroski.* Pacific Philosophical Quarterly, 78:4, 404-420.

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

Ryder, D. (2006). *On thinking of kinds: a neuroscientific perspective.* Teleosemantics. Oxford: Oxford University Press.

Ryder, D., & Favorov, O. V. (2001). *The New Associationism: A Neural Explanation for the Predictive Powers of Cerebral Cortex.* Brain and Mind, 2:2, 161-194.

Scott, Sam (2002) *The psychological implausibility of unacquainted content* Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society 822-827

Searle, J. R. (1992). *The Rediscovery of the Mind* Cambridge: MIT Press.

Sperber, D. & Wilson, D. (1996), *Fodor's Frame Problem and Relevance Theory*, Behavioral and Brain Sciences, 19:3, pp. 530–532.

Usher, M. (2001). *A Statistical Referential Theory of Content: Using Information Theory to Account for Misrepresentation* Mind & Language, 16:3, 311-334.

Usher, M. (2004). *Comment on Ryder's SINBAD Neurosemantics: Is Teleofunction Isomorphism the Way to Understand Representations?* Mind and Language, 19:2, 241-248.

Wheeler, M. (2001). *Two threats to representation*. Synthese 129:2, 211-231.