Sam Scott. *The Psychological Implausibility of Naturalized Content.* Carleton University Cognitive Science Technical Report 2001-05 (www.carleton.ca/iis/TechReports). November, 2001.

The Psychological Implausibility of Naturalized Content

Sam Scott (sscott@ccs.carleton.ca)
Department of Cognitive Science
Carleton University, Ottawa, ON K1S 5B6

Abstract[1]

Conceptual Atomism (CA) is the view that psychological concepts are atoms (no internal structure, no necessary relations to other concepts). CA is a psychological/semantic theory. I show that CA has become mixed up with a separate, meta-semantic project that seeks to naturalize content, and that this Naturalized Conceptual Atomism has ended up in the self-defeating position of positing non-atomic structures for a large number of concepts. The solution is to separate the two projects again, and allow them to develop on their own.

Introduction

For the last two decades, a number of psychologically-minded philosophers have been pursuing a project aimed at naturalizing mental content (Dretske, 1981; 1986; Fodor, 1987; 1990; Millikan, 1984; 1989; 1993). This is a meta-semantic project that seeks an explanation of how meaningful states can arise from non-meaningful ordinary matter. The leading players in this project are also proponents of Conceptual Atomism (CA), the view that concepts are atoms with no internal structure or necessary relations to other concepts. The combined project – call it Naturalized Conceptual Atomism (NCA) – is still very much a going concern (Fodor, 1998; Laurence and Margolis, 1999; Margolis 1998; Millikan, 1998; Usher 2001).

The meta-semantic project has a big problem with what I will call 'unacquainted content' (defined below). Proposed solutions to this problem either do not work, or lead to a psychological/semantic position that proponents of NCA have explicitly rejected in the past – namely, that a large number of lexical primitives correspond to complex (non-atomic) concepts. Most of this paper will be devoted to defending that claim. I will look at the three main attempts to naturalize mental content and show how they all either fail or lead to a non-atomic structure for large numbers of concepts. The remedy for this situation, as I see it, is to separate the meta-semantic project from the psychological/semantic project, and let each develop, for the time being, independently of the other.

[1] This is an early draft of a paper to be submitted to the 24th Annual Conference of the Cognitive Science Society, August, 2002.

A Few Definitions

Concept

Following Laurence and Margolis (1999), a concept is a sub-propositional mental representation. For the present purposes, I will stick to examples of concepts that are about objects or natural kinds from the universe of actual and possible middle-sized dry goods.

Unacquainted Content

Unacquainted content is the Achilles heel of NCA. It is the kind of content that a concept has if its bearer has had no direct experience with the represented object or kind. For example, anyone who has experience with dogs (i.e. almost everyone reading this) will have a normal DOG concept. But most North Americans who have heard of, but never directly experienced, wombats have a WOMBAT concept with unacquainted content.[2]

The term 'unacquainted content' also covers many kinds of hypothesized, future or fictional content. For instance, Anne's MAX concept, referring to her as yet unconceived child, or anybody's UNICORN concept have unacquainted content because the concept bearers could not possibly have directly experienced the (currently) nonexistent object and kind referred to.

Nonexistent Object

Nonexistent objects are what empty concepts and terms seem to refer to. Maybe nonexistent objects are objects in possible worlds, maybe they have some kind of Meinongian nonexistent being, or maybe they don't exist at all and references to them are vacuous. I don't intend to take an ontological position on this, because the main question of the paper is not whether there are unicorns, but whether there are UNICORNs (atomic representations for unacquainted content).[3]

[2] A word in small caps (e.g. WOMBAT) refers to a concept, while a word in single quotes (e.g. 'wombat') refers to a lexical item.

[3] My hunch is that everyday common sense is pseudo-Meinongian, and therefore my description of a unicorn as a nonexistent object will be perfectly intelligible to all but the most dogmatic readers.

The Problem of Unacquainted Content

The main proponents of a Naturalized Conceptual Atomism (NCA) are Dretske, Millikan and Fodor. All three are engaged in a philosophical project that seeks (a) a naturalized account of (b) external content, and all three tend to assume that (c) concepts are atoms with no internal structure. Their three different brands of NCA differentiate around (d) the special problems posed by accidental misrepresentation (AMR).[4] I will briefly discuss these four points of agreement and then I will discuss the differences between the three proposals, focusing on the special problem posed by unacquainted content.

(a) A Naturalized Account. To naturalize content would be to find a coherent story to tell about how the intentional nature of concepts arises from the non-intentional nature of ordinary matter. In practice this has typically meant grounding the meaning of a symbol in some kind of causal or information-bearing relationship between the symbol and the object it represents.

(b) External Content. Proponents of NCA follow Putnam (1975) in insisting that there has to be an external or broad component to representational content. Meaning is not (only) in the head.

(c) Conceptual Atomism. Dretske, Millikan, and Fodor all make the assumption that concepts and other meaningful mental states must be atomic (i.e. both syntactically unstructured and semantically independent from all other concepts.) A concept refers to an object in the world. If the concept had an internal, syntactic structure of some kind, it would raise the question of what the individual parts of the structure refer to, and it's doubtful that that is even a meaningful question to ask in this context. If, for example, DOG is satisfied by all and only dogs because of a causal relationship between DOGs and dogs, then there is just no internal structure in the equation that needs to be explained.

(d) AMR (Accidental Misrepresentation). If the meaning of DOG is just dog, and if DOG gets its meaning in virtue being caused by dogs, what do we do with the fact that sometimes DOG tokens might be caused by things other than dogs? For example, a cat on a dark night might cause a DOG token. If so, this seems to imply that DOG means the same as 'dog or cat on a dark night', which is intuitively wrong. In fact, this "disjunction problem" is much bigger than that. Pictures of dogs can also cause DOG tokens. So can the word 'dog', thoughts about pets, and so on. So the meaning of DOG, on this account, would actually be an infinite disjunction including things like dogs, cats on dark nights, 'dog' tokens, PET tokens, LEASH tokens, and so on. It is in attempting to solve this problem that the three accounts proposed by Dretske, Millikan and Fodor diverge.

[4] AMR is usually called 'misrepresentation'. I use the former term to exclude the deliberate misrepresentations involved in, for example, thoughts about fictional entities.

Dretske: Multiple Paths

Dretske was the first to formulate a version of NCA built on information theory (Dretske, 1981). According to his Informational view of conceptual content, a concept C represents some X in the world only if C carries information about X. More specifically, if X and only X causes C then C represents X. The formulation is meant to be counter-factual supporting. So if X and only X *would cause* C, then C represents X. Left like this, Dretske's theory suffers from the disjunction problem as badly as it any causal theory possibly could. The condition that only X would ever cause C is far too strong to apply to real cognitive agents in noisy environments.

Dretske's proposed solution (Dretske, 1986) begins by making a distinction between simple and complex organisms. Simple organisms have only one route to a representational state. As an example, he points to marine bacteria that contain magnetic sensors called magnetosomes. These sensors detect the surrounding magnetic field and allow the bacteria to align itself with magnetic north. Since in the northern hemisphere, the lines of the magnetic field are inclined downwards, the bacteria can use the signal from their magnetic sensors to swim upwards or downwards in the water. The bacteria die in the oxygen-rich water close to the surface, so bacteria living in the north are naturally selected to use their sensors to swim towards magnetic north. If they are transplanted to the southern hemisphere where the field lines incline upwards, they will kill themselves by swimming into oxygen-rich water.

Dretske thinks that simple organisms like the magnetosome bacteria cannot accidentally misrepresent, because the information contained in whatever representations they form is ambiguous. In its natural environment, the bacterium's magnetosome representations reliably causally covary with the direction of oxygen-free water. Hence it is tempting to say that when the bacterium is moved to the southern hemisphere, it begins to misrepresent that direction. On the other hand, the magnetosome representations also reliably causally covary with the direction of magnetic north, and this does not change no matter where on earth the bacterium is moved to (all else being equal). So on this latter view, it is not a case of AMR that causes the northern bacteria to kill themselves when moved to the south. The magnetosome mechanism still reliably indicates magnetic north, but something else is going wrong inside the organism that causes it to swim in that direction and kill itself. Dretske concludes from this that where there is only one route to a representation, AMR cannot occur because the informational content of the representation is indeterminate.

In more complex organisms, there can be more than one route to a representation. For instance, one can detect a hamburger by seeing it, smelling it, tasting it, feeling it, and so on. There are multiple sensory routes that end in the same representation, H. If, on the contrary, one could only detect a hamburger by smelling it, H would reliably causally covary with both the hamburger and the odor. So the content of H,

on Dretske's story, would be indeterminate. But since there are at least four sensory routes (in a human) to H, the content can be fixed. A token of H caused by seeing a hamburger does not causally covary with the odor of the burger, so the odor can be ruled out as part of H's content. Now we can see how AMR is possible. Any one of the senses can be tricked into causing a token of H when there is no hamburger present, but since the content of H is fixed by the intersection of multiple causal routes, the resulting token H can sensibly be considered to accidentally misrepresent.

Information-based NCA of this kind suffers from a big problem with unacquainted content. In Dretske's verson, the problem is, in many cases, one of indeterminacy. Take Jay Leno, the host of the tonight show. Like most people with a LENO concept, I have watched him for hours on TV. I know both what he looks like and what he sounds like, so I have two causal routes to my LENO concept. If I ever saw Jay Leno in person, it's reasonable to suppose my LENO concept would be tokened through one or more of these causal routes. So the condition that Leno would cause LENO tokens is satisfied. But the condition that *only* Leno would cause LENO tokens is violated – recordings of Leno also cause LENO tokens. Unfortunately, the multiple causal routes story is no help here because I only have two causal routes to LENO tokens and they would both be engaged whether I saw him live or on TV. It's possible that this problem can be set aside by noting that there is a causal relationship of some sort between the real Leno and the TV Leno, but going down this road will likely produce more problems than it solves. There is a causal relationship between bacteria and pimples, but it should not follow, at least in any Conceptual Atomist story, that any part of the content of my PIMPLE concept is a bacteria.

The problem gets worse when there are no direct sensory causal routes to a token, as is the case for nonexistent objects like the fictional detective, Sherlock Holmes, or the Second Shooter hypothesized in certain theories about the assassination of John F. Kennedy.[5] I do know a lot of facts about what these two nonexistent objects are, having heard the conspiracy theory about the Kennedy assassination and read the stories about Sherlock Holmes. But it does not follow that either of these individuals (should they turn out to exist after all) would cause appropriate tokenings in me if I ever saw them because I have no direct sensory link with them.

Recently, Usher (2001) has proposed a variant of Dretske's Informational NCA that uses more sophisticated concepts from information theory to account for AMR. He uses the Mutual Information statistic to measure the correlation between a concept C and all the possible objects that it might represent. The object O which carries the most mutual information with C is what C represents. Mutual information is a measure of how well tokenings of C covary with the presence of O in the environment. Suppose a cognitive agent is presented with a cat on a dark night, makes its best guess as to what it is and tokens the concept DOG. The concept DOG has been caused by a cat, but as long as the measure of mutual information between tokenings of DOG and real dogs is higher than for any other object, DOG still means dog and can sensibly be said to accidentally misrepresent in this case. Though this is an elegant and well-motivated proposal for dealing with AMR within NCA, it fails the test of unacquainted content even more so than Dretske's theory. Now, my LENO concept means the same as "Leno on TV", and who knows what SECOND_SHOOTER[6] and SHERLOCK_HOLMES might turn out to mean, given that they can't possibly carry any mutual information with respect to the nonexistent objects that we intuitively take them to be about.

Millikan: Evolutionary Design

One way of looking at AMR is to say that it arises when a given representation fails to perform its proper function. For example, if DOG is tokened in response to a cat, we can intuitively say that the mechanism that outputs DOG tokens has failed to do its job properly. The DOG token is only supposed to represent dogs, but it's being tokened in this case accidentally in response to a cat. So all the approaches to explaining AMR within a theory of NCA have in common that they want to find some naturalistic way to describe the proper function of a given representation. Millikan meets this challenge head on by trying to find a teleological solution rooted in the theory of natural selection (Millikan 1984; 1989; 1993).[7]

Consider the human heart. Intuitively, we would like to say that its proper function is to circulate blood, but where do we get the authority to say such a thing? Millikan answers that we can say the heart has the function of circulating blood if we can show that that's what hearts were naturally selected for. Applying this idea to mental representations, Millikan urges us to focus on the system within the organism that consumes the representation (Millikan, 1989). Consider the magnetosome bacteria. The representations produced by the magnetosome mechanism are consumed by some other part of the organism that uses the information to pick the current swimming direction. If we assume that these various mechanisms were selected for their ability to propel the bacteria away from oxygen-rich water, then the proper function of the magnetosome representations must be to represent the direction of such water. So when we transplant the bacteria, it can truly said to be Accidentally Misrepresenting that direction. Millikan's solution has the advantage of allowing us to say what we intuitively want to say about the bacteria – that in normal conditions they represent, and in abnormal conditions they Accidentally Misrepresent.

[5] I have no opinion about these theories. Let's just say for the sake of argument that there was no Second Shooter.

[6] 'Second Shooter' has a compositional feel to it, but within the context of the JFK conspiracy theory, it functions more like a proper name, and hence is likely to correspond to an atomic concept.

[7] In later work, Dretske (1988) also explicitly pursued a similar notion of proper function.

A tempting way of looking at this solution is that it is the same as Dretske's information-based solution, but with the causal covariation occurring on an evolutionary time scale rather than over the lifetime of a single organism. In fact, Dretske (1981: 234) does toy with the idea of innate representational content produced in just such a way – representations that are selected for the informational content they carry. But reflection on the case of the magnetosome bacteria shows the real difference in the two theories. Recall that Dretske (1986) was forced to conclude that the content of the magnetosome mechanism's representations were indeterminate – there were just too many things the representations causally covaried with to judge which was the 'proper' informational content. Exactly the same argument would apply on an evolutionary scale. But by focusing on the selected proper function of the representations, Millikan avoids this indeterminacy.

As appealing as Millikan's solution may seem to some, it has problems associated with it that are at least as bad as those associated with Dretske's approach. As Dretske himself has pointed out (Dretske, 1986), the theory cannot explain representational content for anything that the species either has not encountered during its evolutionary history, or has encountered but had no need or use for. If no member of the species or the species that came before ever encountered a particular type of object, then no part of the organisms that comprise the species could possibly have been selected for the purpose of representing that content. This denies representational content to almost any representation of a nonexistent object, and many representations real things such as works of art, new pieces of technology, or anything that is recent enough to have played no role in the evolutionary history of the species. Millikan has a problem with unacquainted content on an evolutionary scale.

Fodor I: Pure Asymmetric Causal Dependence

For the last 15 years or so, Fodor (1987; 1990; 1998) has been pushing his Asymmetric Causal Dependence (ACD) theory to explain how an information-based semantics could deal with, among other things, AMR. In his essay, "A theory of content II", he toys with both a "pure" and "mixed" version of ACD. The pure version combines Dretskian informational semantics (a concept C means X if it's a law that X's cause C's) with an asymmetric dependence condition (Y's that cause C's only do so because X's cause C's and not vice versa). This takes care of AMRs such as horses on dark nights causing COW tokens (this state of affairs is dependent on cows causing COW tokens but not the other way around), and it is also extendible to explain various kinds of robust tokenings (non-X-caused C tokenings that are nevertheless not error cases – for instance, COW tokens that are caused by pictures of cows or thoughts about farms.)

The problem of unacquainted content for pure ACD is immediately apparent, particularly for nonexistent objects. For example, how can non-unicorn-caused tokenings of UNICORN be asymmetrically dependent on unicorn-caused tokenings when there are no existing unicorns? Fodor thinks that this objection can be answered, by reminding us that, like Dretske, he is telling a nomic story, requiring only counterfactual pairings:

> It can be true that the property of being a unicorn is nomologically linked with the property of being a cause of UNICORNs even if there aren' t any unicorns… There wouldn't be non-unicorn-caused UNICORN tokens but that unicorns would cause UNICORN tokens if there were any unicorns. (Fodor, 1990, p101, italics removed and single quotes changed to small caps for consistency).

Fodor has been attacked on the unicorn front before. For instance Baker (1991) constructed a detailed argument based on unicorns and shunicorns (a creature of her own design) that requires us to speculate about which of various possible worlds containing unicorns and/or shunicorns is "closer" to our own. If your mind boggles at this kind of talk, I offer what I hope is a slightly simpler explanation below for why unicorns are a big thorn in the side of the pure version of ACD.

In this unicorn-free world, all valid UNICORN tokenings must be robust tokenings – they are caused by things other than unicorns. The acquisition of the concept UNICORN in the absence of unicorns comes from exposure to representations (visual or verbal) of unicorns. Having learned about unicorns from books and stories, if a unicorn suddenly popped into existence in front of you, it would likely cause a UNICORN token. So we have two valid causal routes to UNICORN tokens: one from representations of unicorns, and one from possible real unicorns that you might encounter in the future (if unicorns begin to exist). To apply ACD, we have to know what would happen if we broke either of these two causal links. Would breaking the causal link between future unicorns and UNICORN tokens break the link between representations of unicorns and UNICORN tokens? My intuition is that this scenario doesn't even make sense, but suppose for the sake of argument that breaking the unicorn/UNICORN link would break the representation/UNICORN link. Then UNICORN tokens are causally dependent on (future) unicorns.

But what would happen if we broke the causal link between representations of unicorns and UNICORN tokens? According to ACD, if UNICORN is to mean unicorn, then this should not affect the causal link between future unicorns and UNICORN tokens. But it obviously does. In a world without unicorns, if you don't learn about them from representations of them then you don't learn about them at all. This means that if a unicorn suddenly popped into existence in front of you, you wouldn't know what it was. Maybe it would cause tokens of HORSE, HORN or whatever, but it wouldn't cause a UNICORN token because you wouldn't have one for it to cause. So in the best case, causal dependence runs both ways and ACD doesn't apply. In the worst case (where you don't buy the story about breaking the link between future unicorns and UNICORN tokens) you have ACD running in the wrong direction and implying that UNICORN has representations of unicorns as its content, which is pretty clearly false – UNICORN has unicorns as it's

content. Notice that you can run exactly the same argument for any type of unacquainted content, such as my LENO concept. Tokenings of LENO in the presence of Leno are causally dependent on tokenings of LENO in response to representations of Leno.

Fodor II: Radical Nativism

But there is a way out of this trap for a radical nativist. Fodor (e.g. 1998) suggests the possibility that we are born with a stock of atomic concepts that get triggered when we encounter objects in the world that strike us as the right sort of thing to fall under the extension of a given concept. For instance what DOORKNOB refers to, is just whatever strikes minds like ours as being a doorknob. Applying this to unacquainted content, if we all have built-in UNICORN tokens that just need to be triggered somehow, then maybe our first encounter with a unicorn would cause a UNICORN token after all. Of course we wouldn't have a word for this token, but that is irrelevant. So ACD would be satisfied by assuming that we are born with a lifetime supply of tokens that already have their nomic triggering conditions fixed.

Radical nativism has not been a popular aspect of Fodor's overall program. Though Fodor correctly points out that whether or not his nativism is true is an empirical question, it seems very unlikely to most researchers that the empirical facts will bear the theory out. Furthermore, if the project is to naturalize content, then all radical nativism does is open up new questions. We are now owed a naturalistic account of how it can be the case that an individual is born with a large stock of mental states that already have the appropriate nomic connections. Given the problems with both Dretske and Millikan's evolutionary accounts, it seems unlikely that such a story is forthcoming. Without the story, all we have reduces to the statement that UNICORN means unicorn because it has a set of properties that causes it to mean unicorn.

Fortunately, ACD does not have to imply radical nativism. Margolis (1998) suggests that even within NCA, there is a plausible story to be told about concept acquisition – one in which new, previously unused tokens are recruited on the fly to represent newly-encountered content. From that point of view, what makes something a UNICORN token is (in part) whether it has a history of being used as such for a particular organism, and not whether it is innately pre-specified to be used only for unicorns.

Fodor III: Mixed ACD

Fodor's "pure" theory does not work as well for unacquainted content as he claims it does unless we are willing to assume radical concept nativism. But the "mixed" version of the theory is even worse. Fodor proposes the mixed theory to get around the problem with verificationism that arises for twin-earth cases in which, for instance, it is difficult to tell whether WATER means H_2O or XYZ.[8] For pure ACD to get to the desired result that in this world WATER tokens mean H_2O and not H_2O-or-XYZ, H_2O and XYZ have to be distinguishable in principle. The way out of this verificationist trap, according to Fodor, is to add an "actual history" condition to the pure theory of ACD. But this would make it a condition that, for example, at least one UNICORN token had actually been caused by a unicorn in order for UNICORN to mean unicorn. (Fodor, 1990:121)

So we're right back where we started with content being denied to all types of unacquainted objects.

Meet the New Boss…

Fodor is not clearly committed to either the mixed or the pure version of ACD, but both versions seem to fail for unacquainted content. However there is still a way out that is consistent with both Conceptual Atomism and the language of thought. This solution, proposed by Fodor (1990: 124) and Dretske (1981: 222, 230) is to allow some concepts to be non-atomic, structured entities built out of atomic components.[9] So UNICORN, LENO, and so on actually unpack into definitional entities assembled out of primitive atoms. Fodor fails to provide any serious defense of the position, except to state that he thinks the situation in which a complex concept would be required is "*very, very rare*" (1990:124, his italics). Dretske proposes the same solution, but, like Fodor, balks at defending it: "I hope [the compositional solution] is sufficiently plausible not to *need* argument" (1981:222, also his italics).

Dretske may believe what he's saying, but Fodor surely does not. The search for such definitional meta-languages is now pretty thoroughly discredited in both philosophy and psychology (Laurence and Margolis, 1999), thanks in part to Fodor himself (e.g. Fodor, Fodor and Garrett, 1975; Fodor, Garrett and Walker, 1980; Fodor and Lepore, 1999). Not only has years of pontification failed to reveal to everyone's satisfaction what any definition of any word might actually be (see Laurence and Margolis, 1999), attempts to find psychological data that might reveal a definitional structure for simple lexical items have also failed (e.g. Fodor, Fodor and Garrett, 1975; or Kintsch, 1974).

A Temporary Solution

Dretske, Millikan, and Fodor have no solution to the problem of unacquainted content, unless we take one of two rather unpalatable options: a) accept a radical concept nativism in which tokens like UNICORN are an innate part of our psychological make-up; or b) accept that many concepts, including UNICORN, WOMBAT, LENO, and so on must have a definitional structure. Nobody (except maybe Fodor) really wants to go for option a, and it begs the question anyhow, so we're left with option b, which not only has no empirical support, but contradicts the whole

[8] See Putnam (1975) if you are not familiar with this thought experiment.

[9] Fodor proposes this (somewhat apologetically) only for cases of nonexistent objects, but it is easily extendible to any unacquainted content, once we rule out radical nativism.

spirit of the Conceptual Atomist enterprise. What do we do now?

There are at least two projects here: the meta-semantic project of naturalizing content, and the psychological/semantic project of determining conceptual structure. The first project is stalled by the problem of unacquainted content, and in attempting to save itself, has wreaked havoc on the second project. My suggestion is that we do not accept this conclusion, and that we separate the projects from now on. Let those interested in the meta-semantic problem try to solve it on its own terms, and leave Conceptual Atomism (CA) to develop on its own. That way CA can be consistent with itself in claiming that UNICORN and WOMBAT are atomic, just like DOG and COW.

I suspect that there will be some skepticism as to whether CA can survive without its accompanying meta-semantic theory. Therefore, I will end with two reasons why I think that that it can.

1. No competing theory is tied to a similar meta-semantic project. Neither the prototype theory nor the theory-theory of concepts attempts to say anything about how meaning arises from non-meaningful stuff. Neither do most modern versions of the definitional theory – see Laurence and Margolis (1999) for a review. And, after all, why should they? At this early stage, a psychological/semantic theory should be judged on its own merits, not by standards set at some other level of analysis.

2. CA is still a good theory even without the meta-semantic project. First, there is no psychological evidence for definitional structure, and the evidence that drives the prototype and theory theories can be accounted for within CA – the former by supposing that typicality effects arise from a peripheral categorization mechanism, and the latter by supposing that people do have theories that guide their behavior, but that these theories are *about* the concepts they involve, rather than being *constitutive* of them. And second, CA is still the most natural fit to the computational theories of mind that are so popular these days.

Acknowledgements

Will be added for the final draft.

References

Baker, Lynne Rudder. (1991). Has content been naturalized? In Barry Loewer and Georges Rey (Ed.) *Meaning in Mind: Fodor and his Critics*. Oxford: Blackwell.

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

Dretske, Fred. (1986). Misrepresentation. In R. Bogden (ed.) *Belief, Form, Content and Function*. Oxford: Oxford University Press.

Dretske, Fred. (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.

Fodor, Janet D., Jerry A. Fodor, and Merrill F. Garrett. 1975. The psychological unreality of semantic representations. *Linguistic Inquiry, 4*, 515-531.

Fodor, Jerry A. (1987). *Psychosemantics*. Cambridge, MA.: MIT Press.

Fodor, Jerry A. (1990). A theory of content II. In *A Theory of Content and Other Essays*. Cambridge, MA.: The MIT Press.

Fodor, Jerry A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Fodor, Jerry A., Merrill F. Garrett, E. Walker and C. Parkes. 1980. Against definitions. *Cognition, 8*, 263-367

Fodor, Jerry A. and Ernie Lepore. 1999. Impossible words? *Linguistic Inquiry, 30*, 445-453.

Kintsch, Walter. (1974). Lexical decomposition: Compression and memory. In *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Laurence, Stephen and Eric Margolis. (1999). Concepts and cognitive science. In Eric Margolis and Stephen Laurence (Ed.) *Concepts: Core Readings*. Cambridge, MA.: MIT Press.

Margolis, Eric. (1998). How to acquire a concept. *Mind and Language, 13*, 347-369.

Millikan, Ruth Garrett. (1984). *Language, Thought and other Biological Categories*. Cambridge, MA.: MIT Press.

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

Millikan, Ruth Garrett. (1990). Compare and contrast Dretske, Fodor, and Millikan on teleosemantics. *Philosophical Topics, 18(2)*, 151-161.

Millikan, Ruth Garrett. (1998). A common structure for concepts of individuals, stuffs, and real kinds; more mamma, more milk and more mouse. *Behavioral and Brain Sciences, 9*, 55-100.

Putnam, Hilary. (1975). The meaning of meaning. In *Mind, Language, and Reality*. Cambridge, UK: Cambridge University Press.

Usher, Marius. (2001). A statistical referential theory of content: Using information theory to account for misrepresentation. *Mind and Language, 16(3)*, 311-334.