

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

Karen Neander

The Oxford Handbook of Philosophy of Biology

Edited by Michael Ruse

Print Publication Date: Jul 2008

Subject: Philosophy, Philosophy of Mind, Philosophy of Science

Online Publication Date: Sep 2009 DOI: 10.1093/oxfordhb/9780195182057.003.0017

Abstract and Keywords

This article is about the philosophical problem of intentionality. Neo-Darwinian selection theory can be used to solve the problem. The norms that apply to mental representation can be reduced to functional norms. The theories that are based on this idea are known as teleological theories of mental content. The philosophical problem of intentionality is related to but different from the psychological one. Ordinary functionalism is presented as endorsing two main theses: mental states and they are individuated by their “functional roles.” Teleofunctionalism endorses both analogous and homologous psychological categories. The reason that teleofunctionalism holds as a theory of mental content is that content is normative. A second motivation for teleosemantics is the nature of cognitive neuroscience as it forms a bridge between the natural sciences and the so-called interpretive fields of inquiry. The aim of the teleosemantic theories sketched above is to explain how content is determined in terms of nonmental phenomena.

Keywords: intentionality, functionalism, mental, neuroscience, teleosemantic

1. Introduction

Daniel Dennett, ever ready with a good metaphor, likens Darwin's theory of evolution by means of natural selection to universal acid, “a liquid so corrosive that it will eat through anything” (1995, 63). His point: Darwin did not just revolutionize biology. His ideas have profound implications for the social sciences, religion, cosmology, and metaphysics—our whole world view. This chapter concerns one possible strand of this far-reaching revolution; it concerns the philosophical problem of intentionality and whether we can use neo-Darwinian selection theory to solve that problem. The hope that we can do so is based on the idea that the norms that apply to mental representation (the norms that underlie claims concerning, for example, the correctness and incorrectness of representation) can ultimately be reduced to functional norms (the norms which underlie claims concerning normal and abnormal functioning). The theories that are based on this idea are known as

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

teleological theories of mental content (also, teleosemantic theories). The connection with Darwin is that functional norms are thought to (p. 382) depend on the dispositions for which items have been selected by a natural process of selection—Darwinian selection being the paradigm case.

2. The Problem of Intentionality

Thoughts (beliefs, desires, memories, imaginings, and so on) are *about* things (all sorts of things—tables, chairs, the weather, the election, numbers, unicorns). Intentionality is this *aboutness* of our thoughts. For a thought to be about something, it must represent that thing, and so intentionality is the representational power of our minds and brains. For example, if you have a thought about cats, your thought involves a representation of cats. Philosophers want to know what it is about the thought that makes it a thought about cats. Given that it is our brains that are doing our thinking, we can ask this question about neurological processes too. What is it about a neurological process that makes it the case that it represents what it represents?

The philosophical problem of intentionality is related to but different from the psychological one. As it is generally understood today, the main task of cognitive science is to provide an account of the flow and processing of information in the brain and the wider nervous system, to explain how cognitive systems function, to explain how we have the cognitive capacities we do. In doing this, mainstream cognitive science uses a notion of a neural representation. For example, ensembles of neurons are said to have representational functions, and patterns of neural firings are said to represent certain features of the world. There is no point in drawing sharp disciplinary boundaries, but this task is not the traditional philosophical one. Simplifying for the sake of a stark comparison, the scientific task involves explaining cognitive capacities, which involves using the notion of a representation, whereas the philosophical task is to give an analysis of this notion.

Franz Brentano's ([1874]1911) discussion of the problem is famous, so much so that the problem is sometimes known as “Brentano's problem.” He points out that we cannot desire something, such as a sunny day or an ice cream, unless it is “present to our mind,” and he puzzles over the nature of this “presentation.” As he emphasizes, a mental presentation of an item does not require that the item be present in the usual sense. We can yearn for a sunny day even if—in fact especially if—it is raining, and we can even wish for things that do not exist. A child might want to meet a unicorn or a wizard, and a cure for cancer and a better robotic vacuum cleaner are high on my list. Similar things can be said about beliefs, imaginings, emotions, and so on: I can only have a belief about something, or imagine something, or love or fear something, if it is “present” to my mind, but it need not be present in the ordinary sense, and it need not even exist.

(p. 383)

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

Following Brentano, Roderick Chisholm (1957) argues that the relationship between a mental presentation and the thing presented is not a physical relation between two things. In the case of an ordinary physical relation, he says, both *relata* must exist. If it rains on the parade, there must be rain and a parade. If the spider catches the fly, there must be a spider and a fly too. In contrast, Betty can love Santa and Bob can fear God's wrath even if there is no God, no Santa. Chisholm concludes that intentionality is not a physical phenomenon. Few philosophers of mind today would follow Chisholm to this conclusion, and the proponents of teleological theories of mental content do not do so. However, the problem remains: It is hard to say what kind of relation is involved in a mental representation representing its represented.

If we use “thought” in a generic sense, to cover all intentional or representational mental states (e.g., beliefs, desires, memories, perceptions, imaginings, loves, hates, and unconscious representations of features of the world), the key question is this: What obtains between a thought about something and that which it is about *in virtue of which* the former is a thought about the latter? What *constitutes* a mental representation representing what it represents?

This question parallels one in the philosophy of language. With respect to natural languages, such as English or Chinese, we can ask: In virtue of what does a given word have the meaning that it has? The word “meaning” covers a number of related phenomena, but we can ask, for example, what it is about the word “cat” that determines that it *refers* to cats. Similarly, in the case of mental representations, we can ask what it is about a given mental representation that determines that it has the meaning—or, as we say in this case, the *content*—that it has. What determines that my cat concept applies to cats and only cats?

While the two questions—the one relating to words and the other relating to mental representations—are very similar, it is important to appreciate that some of the answers that are plausible in the first case are deeply problematic in the second. Most notably, a Gricean or neo-Gricean analysis has no hope of working for mental representations. A neo-Gricean analysis of linguistic meaning is one that says that our words have meanings because we mean something by them. This seems plausible. However, the parallel response with respect to intentional mental states is not plausible. In order to mean something by a word, we must be able to represent the word to ourselves, and have certain intentions with respect to its use. We cannot give, in their turn, a Gricean analysis of these intentional mental states, for if *these* have the contents that they do because we thinkers have intentions with respect to *them*, an infinite regress looms. If meaning is to get started, some mental representations must get their meaning in a different way. To restate the problem yet again, the problem is to say what that way is.

One possibility is that intentionality is fundamental and inexplicable. However, this proposal does not *explain* intentionality—it abandons the possibility of explaining intentionality. Those who seek what is often called a naturalistic theory of mental content look for a way to explain how meaning could arise from nonmeaning in the natural world. Teleologi-

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

cal theories of mental content are intended (p. 384) to be naturalistic theories. They take seriously the idea that intentionality is a biological phenomenon. Further, their proponents think that, not only have cognitive systems evolved, but—and here is the crucial, controversial claim—in the fact of this evolution lies the solution to Brentano's problem.

3. Functionalism and Functional Roles

Ordinary functionalism (i.e., functionalism of the nonteleological variety) is standardly presented as endorsing two main theses: (1) Mental states (events, processes, properties, and so on) are “multiply realizable”; and (2) they are individuated by their “functional roles.”

The first thesis is negative. It tells us that a creature's being in a mental state, *M*, does not depend on the material (or, for that matter, immaterial) composition of the creature, as such. According to functionalism, it doesn't matter if a system is composed of silicon chips or laser pathways or ion channels through carbon-based molecules, as long as certain other conditions are satisfied. This is where the second thesis of functionalism comes in. According to the second thesis, mental states are individuated (typed or classified) by their “functional roles.” As elaborated in standard versions of functionalism, the functional role of an item depends on its actual causal dispositions. Functionalism says that, in order for a creature to be in a mental state *M*, it must possess an inner state that has the functional role—that is, the causal dispositions—*characteristic* of the *M* state. To be in pain is (crudely) to be in a state that is caused by bodily damage, that tends to cause one to favor the damaged bodily part, that tends to cause one to avoid further damage, that tends to cause the belief that one is in pain, and so on. An attractive implication of functionalism is that, in principle, it allows creatures with radically different biological systems to partake in some of the same kinds of mental states as us. A being from Alpha Centauri could wonder if there is life in neighboring planetary systems and hope for intragalactic peace, even if it lacked a brain like ours considered in chemical and structural, as opposed to functional, terms.

All versions of functionalism agree that *M*'s “characteristic” functional role is a *subset* of the causal dispositions of *Ms*. For example, we might feel burning pain at a lower temperature than Martians do, but that is not supposed to be a difference that could make a difference to whether we experience the same burning sensation. As for how we should try to discover the characteristic functional roles of mental states, opinions differ: Commonsense functionalism, for instance, says that we discover them by consulting the platitudes of folk psychology, and psychofunctionalism, in contrast, says that we discover them by examining the role they play in our best scientific theory. Notice that neither of these proposals, as they stand, gives (p. 385) a sufficient distinction between relevant and irrelevant dispositions. Consider, for example, that it could be both platitudinous and confirmed by scientific discovery that human skin burns in boiling water, whereas Martian skin does not.

4. Teleofunctionalism and Teleofunctions

The biologically literate might have noticed that regular functionalism sees mental kinds as (more like) analogous than homologous kinds. Like analogous kinds, it co-classifies items on the basis of their shared function, rather than, as with homologous kinds, on the basis of shared descent. On the assumption that we evolved separately, we can only share analogous and not homologous kinds of mental states with alien beings.

Teleofunctionalism endorses both analogous and homologous psychological categories. In that sense, it modifies thesis (1): Some but not all psychological categories are analogous. More significant for our purposes is the way it modifies thesis (2): For a start, it employs a different notion of function, one that is said to be “normative” and “teleological.”

Care needs to be taken with respect to our use of both of these words. Let's start with the idea that the relevant notion of function is *normative*. Here, all it means to say that it is a normative notion of function is that it is one that underlies talk of systems functioning normally and abnormally. Most understand this to mean that it allows for the possibility that an item can have a function that it cannot perform. Colloquially speaking, normal functions distinguish between what an item is “supposed to do” and what it is disposed to do. Ordinary functionalism, as ordinarily understood at any rate, uses a nonnormative notion of a function, such that if an item, *x*, has a function to do *Z* in circumstances *C*, then *x* is disposed to *Z* in *C* (Cummins 1975 is a useful elaboration of this notion). In contrast, teleofunctionalism uses a notion of function that allows that an item, *x*, could have a function to do *Z* in circumstances *C* even if *x* is not disposed to do *Z* in *C*. A normative notion of function allows us to coherently say, for example, that all hearts have the function of circulating blood and yet that some hearts malfunction and cannot do so.

Some philosophers dislike this use of the word “normative”, but the word itself is not particularly important and you can put mental scare quotes around it if you want. What is important is that teleofunction ascriptions do not really entail claims about what ought to be done, except in a metaphorical and conditional sense. We may sometimes say that, for example, our pineal glands “ought” to secrete melatonin at nightfall. However, this talk of what “ought” to happen is metaphorical, given that no voluntary action is involved. And all that anyway follows from the (p. 386) function ascription is the conditional: that our pineal glands ought (so to speak) to secrete melatonin at nightfall *if* they are to function normally. If you doubt this, and suspect that moral goodness is required for normal functioning, consider the function of, say, cell binding sites on the HIV virus. We do not want to claim, I assume, that these sites should bind in any moral sense of “should”. We merely mean that they ought to bind *if* they are to function normally, or that they would bind were they to function normally. I return to some of these points about functional norms when I discuss the norms of content a little later.

Different analyses of this normative notion of function are on offer. However, proponents of teleological theories of mental content think that it is *teleological*. It is a teleological notion insofar as it concerns what an item is *for*, or, in other words, what an item of the

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

type was selected for. There is a variant of this idea for the functions of artifacts, but it is natural functions that concern us here. In the case of natural functions, an item's function is said to be what items of the type were selected for by a natural process of selection, neo-Darwinian natural selection being the paradigm. On this understanding, the function of an item, x , is to Z in C , if and only if x s were selected for Z -ing in C by a natural process of selection (Millikan 1984, 1989a and Neander 1991a&b, after Wright 1973). Note that this allows for the possibility that some x s cannot Z in C . Hearts were selected for circulating blood, and hence all hearts have that function, on this analysis, but not all hearts have the wherewithal. Human V4 is adapted for processing information about color, and so all human V4s are for processing information about color, but we don't all have working V4s. This analysis of functions is known as an *etiological* analysis because it refers to the history of items with the function, and it is also known as a *selected-effect* (or SE) analysis because it says (more or less) that functions are the effects for which phenotypes were selected.

This etiological analysis and the notion of a teleofunction that it provides needs more development and defense than I provide here. For example, readers might want to ask: How can we accommodate vestigial loss of function and changes in function over time? How are traits to be typed? (See Neander 2002.) The etiological theory is also controversial, and alternative analyses have been offered. (See Ariew et al. 2002, Allen et al. 1998 and Buller 1999, for discussion). But, while it is controversial, it is the dominant analysis of normal/normative functions in biology. And, although it is an interesting question what kinds of changes would need to be made to teleosemantics, were various different analyses of normal function to be used, there has been little interest so far in using a nonteleological and nonselectional analysis of what have come to be known as *normal* or *proper* functions to ground mental content.

Some proponents of teleosemantics don't like to use the word "teleological" for their theory. In their view, this term should only be used when literal purposes are involved, and natural functions are instead described as "teleonomic", since they are, by hypothesis, not literally purposive. In contrast, in my idiolect and that of many others, natural functions count as teleological, although they are not purposive, because they are what items are *for* (i.e., what they were selected for). (p. 387) Thus, in this idiolect, a *teleonomic* notion is a kind of teleological notion, one that is nonpurposive (and naturalistic). However, this is just a terminological difference again. The important thing is that the teleofunctions (to use a neutral term) that are used in a theory of mental content must not depend on purposes. If they did depend on purposes, the theory of mental content would be circular. Talk of natural teleofunctions does *not* implicate purposes, according to the above etiological or selected-effect analysis.

Recall that thesis (2) of functionalism is the claim that mental phenomena are individuated on the basis of their characteristic functional roles. In some cases, teleofunctionalism involves the claim that mental phenomena are individuated on the basis of their teleofunctions. This would be the simplest way to modify thesis (2). It is a coherent way of speaking, if your current thought or the ache in my right elbow now have teleofunctions

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

courtesy of the teleofunctions of the mechanisms that produce them. However, it is more straightforward to claim that mental phenomena are (in whole or in part) individuated on the basis of the teleofunctions of associated mechanisms. For example, one might claim that the contents of mental representations are determined (in whole or in part) by the teleofunctions of the mechanisms that produce them or use them.

So far, I have described both functionalism and teleofunctionalism as universal theories of mind (as theories of *all* kinds of mental phenomena). Functionalism enjoyed a brief period of popularity as a universal theory, but it is rarely supported as such today. Teleofunctionalism can also be offered as a universal theory for all kinds of mental phenomena, and it has been given for the attitudes, for emotions, and for phenomenal consciousness. However, it has attracted most attention as an account of mental content, and in what follows I focus on this application of it.

5. Motivation 1: The Normativity of Content

One of the main reasons that teleofunctionalism seems to hold some promise as a theory of mental content is that content is normative. Content is normative insofar as it licenses certain semantic evaluations. For example, if a concept is a *cat* concept, then it is correct to apply it to cats and incorrect to apply it to dogs. If a belief has the content *it is raining (here and now)*, then it is true if it is raining (here and now) and false otherwise. If I see something as a bird fluttering its wings, I perceive it correctly if it is a bird fluttering its wings but I misperceive it if it is a sheet of plastic flapping in the breeze. If my brain represents one line as longer than the other, it represents them correctly if they do differ (as represented) in length, but it misrepresents them if they are the same length, as is the case in the Muller-Lyer illusion (figure 16.1).

(p. 388)

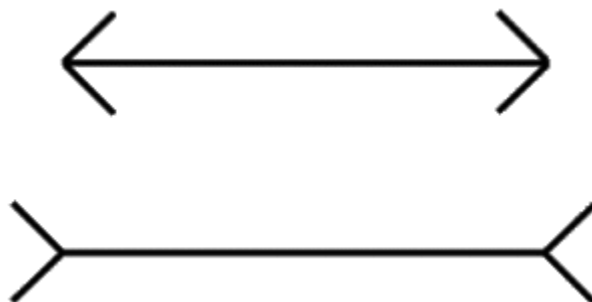


Figure 16.1. The Muller-Lyer illusion.

Earlier, I remarked that function ascriptions are normative, but not in a moral sense. One might be tempted to think that meaning and content ascriptions are normative in a stronger sense. It might seem that, for example, if “horse” means *horse*, then we *should* apply the term to horses, and not to cats, dogs, or grasshoppers. However, all that follows from the *content* of “horse” is the conditional: If “horse” means *horse*, we ought to apply

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

it to horses and not to non-horses *if* we are to use it in accord with its meaning. It is true that agency is involved in the case of word meanings—we use words voluntarily, in the main. Further, it is true that moral issues can arise as a result of misrepresentation. We might, for example, insist that someone should have told the truth: He said that it was a horse but he sold me a donkey; he should not have applied the word “horse” to this donkey and cheated me. In such cases, a genuinely moral “ought” is involved. But these are not the relevant cases to consider when we are considering the general claim that meanings and contents entail truths about how terms and mental representations ought to be applied. The meaning of a term entails truths about how we ought to use that term *if* we are to use it correctly, *if* we are to use it in accord with its meaning. That is what follows from its *meaning*.

This should sound familiar, for this parallels the normativity of teleofunctions. The teleofunction of an item entails truths about how the item (metaphorically speaking) “ought” and “should” operate *if* it is to function properly. The terms ought and should are better suited to linguistic meaning, because in that case there is an agent with voluntary control involved. However, even this difference is lost in many cases of mental content. There is no agent with voluntary control in the case of the Muller-Lyer illusion, for example. It is not appropriate to maintain that we (or our visual systems) *ought* not to produce the illusion, in the literal sense of “ought,” which implies “can.”

Although content ascriptions are, on this view, not prescriptive in a moral sense, they are normative in the way that philosophers have typically had in mind when discussing the normativity of content. That is to say that, when we ascribe content (e.g., when we say things like “this is a *cat* concept” or “this is *adding* and not *quadding*”), we make a claim that has evaluative implications concerning future applications—implications to the effect that this application is correct, this is incorrect, and so on. Similarly, teleofunction ascriptions have evaluative implications concerning (p. 389) future dispositional manifestations—implications to the effect that this is functioning normally or properly, this is malfunctioning, and so on. What matters for our present purposes is that content and teleofunctions are in the same boat, whereas content and the functionalist's functional roles are not. Functional roles do not license evaluations of a parallel sort.¹

Developed as a theory of mental content, ordinary functionalism goes by the name of conceptual role semantics (CRS) or, sometimes, inferential role semantics. According to CRS, the content or meaning of a mental representation (e.g., such as my cat concept, my mental representation of cats) is determined by its conceptual role. Its conceptual role is its functional role insofar as this involves other concepts; it is, in other words, its disposition to interact with other concepts in inferences, and so forth (Block, 1986). We need no more than a rough idea of CRS to see the problem with it that I want to mention here.

Saul Kripke (1982) discusses a version of this problem in his interpretation of Wittgenstein's thoughts on rule following. Kripke is primarily discussing linguistic meaning, but there are parallel issues for mental representations. Kripke speaks, for example, of “plus,” but we can run the same argument for the corresponding mental operator,

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

PLUS. What, Kripke asks, makes mental adding *adding*? When Tommy thinks that six plus thirteen equals eighteen, what makes it the case that he is adding incorrectly, as opposed to performing a different function (following Kripke, call it *quadding*) that happens to deliver eighteen as the value of six plus thirteen? We tend to make mistakes, and so a bare appeal to our actual dispositions—to our representation's functional role—does not provide a basis for assessing some outcomes as correct and others as incorrect.

It is sometimes alleged that CRS is more plausible as an account of (something like) *sense* rather than *reference*. This distinction between sense and reference is motivated by some of Gottlob Frege's ([1892]1960) observations about language. For example, he notes that the ancients who spoke of the Morning Star and the Evening Star did not know that the two names referred to the same heavenly body, Venus. This lack of knowledge on their part is puzzling if all that there is to meaning is reference. For, if all there is to meaning is reference, certain identity statements—for example, *the Morning Star is the Evening Star*—would not only be true, they would be trivially true. Frege answers the puzzle by proposing that words have senses as well as references, and hence that two terms, such as “the Morning Star” and “the Evening Star”, can differ in sense and yet have the same reference.

The idea that meaning involves two aspects—reference and something else that addresses Frege's puzzle—has remained popular to this day. Many philosophers now speak of concepts or mental representations having cognitive contents as well as referential contents. The idea is that two concepts can have the same reference, but different cognitive contents, and it is sometimes proposed that cognitive contents can be types of functional roles. CRS is generally thought to be most plausible with respect to capturing this cognitive content aspect of meaning. (p. 390) However, as Kripke argues, since we are disposed to make mistakes sometimes, we will need to distinguish between right and wrong roles.

The nearest thing to CRS though adapted for reference is what Jerry Fodor (1987) has called the crude causal theory (CCT). According to CCT, a mental representation of red (here denoted as RED) represents something as red if and only if all and only instances of the color red can cause REDs. Similarly, according to CCT, a mental representation of a horse (here denoted as HORSE) represents something as a horse because horses and only horses can cause HORSEs. This is most nearly plausible with respect to perceptual representation. Instances of red do cause REDs and instances of horses do cause HORSEs. However, Kripke's problem also applies to the CCT, considered as a theory of reference.

How can we characterize perceptual contents in terms of their actual causes, or even their actual dispositions insofar as these concern causes, given that we are sometimes disposed to misperceive? Suppose we mistake a donkey in the far distance for a small horse. If HORSEs correctly represent *whatever* causes them to be tokened, as is claimed by the CCT, it would follow from the fact that both donkeys and horses have caused HORSEs, that HORSEs *correctly* represent both donkeys and horses. This is part of what Fodor calls “the disjunction problem.” CCT licenses all sorts of inappropriate disjunctive contents: It licenses the claim that, for example, HORSE refers to horses *or* donkeys in

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

the far distance *or* whatever else is mistaken for a horse. (Fodor 1990 has his own attempt to solve this problem, but this is not the place to discuss it.)

Proponents of teleological theories think we can derive normative content from normative functions. Our dispositions include dispositions to make errors. So it seems that what is needed is a distinction between what our minds are disposed to do and what they are, colloquially speaking, “supposed to do.” The idea is that they are “supposed to do” what they were selected for doing by a natural process of selection. Perhaps mental operators like AND and ADD are used correctly if they play the inferential role for which they were selected. And perhaps perceptual representations like RED and HORSE refer correctly if they are produced in response to the causes they were selected for responding to. This is not quite what any extant teleosemantic theory says, but it is a place to start. I describe a few teleosemantic theories more carefully later.

6. Motivation 2: Scientific Integration

A second motivation for teleosemantics arises from the nature of cognitive neuroscience. Cognitive neuroscience forms a bridge between the natural sciences and the so-called interpretive fields of inquiry. The natural sciences are those (p. 391) ordinarily described as such; they include, for example, physics, chemistry, geology, and biology. The interpretive disciplines include psychology and the social sciences, and they count as “interpretive,” as the word is used here, because they study intentional systems *as such*. Organic chemistry, for instance, does not count as interpretive because, even though it studies intentional systems such as brains, among other things, it does not study them *as* intentional systems, not *as such*. Chemical processes in the brain are studied, but any intentional properties that these processes have do not feature in the descriptions and explanations developed.

Traditionally, philosophers have thought of neuroscience as a nonintentional science. However, this is mistaken. Neuroscience (as well as cognitive neuroscience and cognitive science more generally) is imbued with talk of the representational functions of neurons, ensembles of neurons, and patterns of neural firings, once we rise above the cellular level of functional analysis. There has been much talk of the autonomy of cognitive science vis-à-vis neuroscience, and to some extent it is true that the two disciplines can make progress independently. Nonetheless, the two fields are two sides of the same coin. In general, cognitive scientists study the processing of information in the brain and in the wider nervous system, and they attempt to relate this to its neurological substrate; and neuroscientists study the neurological substrate and attempt to relate this to its information-processing functions. As time goes by, the two have become more integrated, as we would expect.

How does cognitive neuroscience bridge the natural and the interpretive sciences? Put simply, depending on the level of analysis involved, the functional analysis of the operation of the brain is both interpreted and uninterpreted. It bridges what was once perceived by philosophers to be the gulf between the interpretive and noninterpretive sci-

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

ences, because physiology is the study of the functional organization of organic systems and, since the brain uses representations to perform its cognitive functions, physiology becomes an interpretive science when the subject under study is the brain. One cannot give a complete functional analysis of the operation of the brain without eventually discussing its representational functions.

On its own, this might not seem to amount to a motivation for a teleological theory of mental content. However, we have such a motivation if we grant two further assumptions. One is that the key notion of function used in physiology is normative, and the other is that it is teleological, in the senses of these terms outlined earlier. Both assumptions are controversial, but there is really no good basis for doubt about the first. Talk of systems functioning properly or normally, of their malfunctioning, being dysfunctional, being functionally impaired, and so on, is pervasive in neuroscience and cognitive neuroscience, just as it is in other areas of physiology. If you doubt it, just take a look at university texts on the subjects. It is true, as is sometimes alleged in support of skepticism on this point, that research reports do not seem to advert to the notion to the same extent, but a closer observation usually reveals that there is a background presumption that it is a normal (p. 392) or an abnormal system that is under discussion. As for the second assumption, it cannot be read off the page. As I said earlier, the etiological analysis of normal functions is controversial. However, *if* it is granted, we have a reason for taking teleosemantics seriously. Even if it is not granted, we have a reason for taking something like teleosemantics seriously, for we still have a reason to take seriously the idea that the normative notion of function that is employed in physiology—however it is best interpreted—could play an important role in explaining mental content.

The reason is that neurophysiological talk of information-processing functions is already normative in a way that parallels the way that mental content is normative. The functions supply the normativity and the fact that they are *informational* functions makes them representational, as opposed to, for example, circulatory or digestive.

The idea shared by all teleological theories of mental content is that the norms of mental content derive from the functional norms that apply to the mechanisms that produce or use representations. However, different people have developed the idea in different ways. In the next three sections, I look at three ways of doing so. I start with Fred Dretske's indicator semantics and his functional indeterminacy problem, to give that problem an immediate airing, and then move to Millikan's and Papineau's proposals. Following these outlines of particular teleological theories, I develop the indeterminacy problem further, and then close with a look at an entirely different type of problem—the Swampman problem.

7. Indicator Teleosemantics

Dretske (1981) began by proposing a nonteleological version of informational semantics, which he then modified into a teleological theory. Informational semantics begins with the idea that mental representations carry information. On this view, neural signals are inter-

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

nal natural signs of environmental features, just as smoke is a sign of fire and dark clouds are signs of rain.

However, as Dretske (1986), following Paul Grice (1957), points out, there is an important difference between what we might call *mere* natural signs and those with genuine meaning. Mere natural signs cannot misrepresent, whereas genuine representations can misrepresent (in some contexts). To use one of Grice's examples, Johnny's spots are a natural sign that he has the measles only if Johnny has the measles. If Johnny does not have the measles, his spots cannot mean that he does. Mere natural signs can be misunderstood, but they cannot misrepresent. Our words, or the doctor's words, in contrast, can misrepresent Johnny's condition. The doctor's words "Johnny has the measles" represent Johnny as having the measles (p. 393) even if he (Johnny) does not have them. Similarly, the doctor's *thought* that Johnny has the measles can misrepresent his condition. The doctor's thought that Johnny has the measles has the content *Johnny has the measles* even if he does not have the measles. Thus, the fact that neural signals are natural signs cannot alone account for their meaning, since it does not account for their misrepresenting.

A theory of content needs to allow for the possibility of misrepresenting, not just because we are, as a matter of fact, fallible, but because the logical possibility of error is entailed by (most all) contents. Even if we never made mistakes, it would still be the case that the following conditional would be true: *Were* we to use CAT to label a non-cat, we would be in error. It is this logical possibility of error that a theory of content must accommodate. For, if it were not the case that, were a CAT used to label a non-cat it would be used in error, CAT could not in fact refer to cats and only cats.

How does Dretske's theory accommodate this possibility of error? On his view, neural signals are natural signs, but they are not *mere* natural signs. One difference between such things as smoke and fire, on the one hand, and neural events and what they represent, on the other, is that the latter kinds of pairings are the product of evolution and/or learning. Dretske appeals to this fact to explain psychosemantic norms (the norms pertaining to mental content). His idea is that neural signals do not merely carry information, but are recruited or selected for doing so. In other words, they have information-carrying *functions*. This allows for error, or so goes the initial thought, since things do not always perform their functions. Neural events can fail to carry the information that it is their function to carry.

To be more precise, Dretske defines a notion of *indication*. As he defines this term, instances of one type of event *indicate* (or carry information about) instances of another type of event if and only if, within the relevant period and place, whenever there is an instance of the first, there is also an instance of the second. There can be no *misindication*. One event cannot misindicate another, because if *Rs* indicate *P*, then there must be an instance of *P* whenever there is an instance of *R*. Indication is therefore not representation. However, Dretske suggests, a neurological event, *R*, *represents* a situation as *P*, just in case *Rs* were recruited (i) for indicating *P* and (ii) for producing some bodily movement,

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

M. Error is possible because, having been recruited for indicating *P* (and producing *M*) *R*s can subsequently fail to do so.

This is Dretske's initial teleosemantic proposal, but he no sooner makes it than he notices a problem with it. Here enters his (now) famous example of the magnetosomes—small magnetic sensors in anaerobic bacteria. These sensors detect local magnetic north, and hence geomagnetic north, and point the bacteria in the direction of deoxygenated water. Dretske says that, although we might want to say that the function of the magnetosomes is to indicate the direction of deoxygenated water, we can with equal plausibility say that their function is to indicate the direction of geomagnetic or, for that matter, local magnetic north. Dretske's interim conclusion is that he has not yet found “nature's way of making a mistake.” (p. 394) This has (and variants on this have) become known as the functional indeterminacy problem.

Dretske (1986) initially concludes that functions based on natural selection cannot provide determinate content. However, he argues that determinate content is nonetheless possible for cognitive systems with sufficiently sophisticated learning capacities. Anaerobic bacteria cannot have representations with determinate contents, but Pavlov's dogs can. What is required, on this view, is an ability to recruit indefinitely many new episodic pathways for detecting the same phenomenon. Consider Pavlov's dog. Prior to conditioning, only food (let us suppose) causes the dog to produce an inner state, R^f , which causes it to salivate. But, by associating food with other stimuli, a researcher can train the dog to salivate to the sound of a bell, to her wearing a red sweater, to someone's stroking its paw, and so on. According to Dretske, R^f refers to food, and not to the other learned associations, because R_f has the *time-invariant* indicator function of indicating only food.

I shall mention a couple of problems with this proposal before moving on to Millikan's theory. One problem relates to the strictness of Dretske's definition of indication: *R* indicates *P* only if, given an instance of *R* then there is an instance of *P* (i.e., with a probability of 1). As Joelle Proust (1994) points out, conditioning does not require a reward on every occasion that the reinforced response is elicited, and conditioning might be even more effective when a reward is not provided every time the reinforced response is made. This is a problem for Dretske, given that he wants to say that R^f was recruited for indicating food, for it might not have *indicated* food during the recruitment (conditioning) process. A second problem is raised by Barry Loewer (1987), who points out that, once the dog is dead, we can retrospectively identify a time-invariant closed disjunction of other stimuli (the smell of food or the sound of a bell or the appearance of a trainer wearing a red sweater or someone's stroking its paw, etc.) that R^f was recruited to indicate.

Further, as Dretske himself makes clear, this proposal does not permit representation in the absence of learning. This might also be a serious problem because, arguably, learning requires representation, and “innate” concepts and knowledge are now considered to be serious possibilities by developmental psychologists. Later, Dretske (1988, 1991) repudiated this part of his earlier theory, and allowed that natural selection can underwrite con-

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

tents. Unfortunately, he does not return to the indeterminacy problem, except that in places he seems to suggest that we disambiguate contents by linking them to what was “maximally indicated” during recruitment. This clearly modifies the original definition of indication, which did not allow of degrees. But *that* it does so is clearer than *how* it does so or, in other words, how the new notion of indication is to be read. Note also that it will disambiguate contents in the case of the magnetosomes, if it does so at all, in favor of the magnetosome having the function to indicate magnetic north, rather than anaerobic conditions. I do not regard this as a problem for Dretske (Neander 1996), but some do not like this result. (It entails that in this case virtually no error can be made in the absence of malfunctioning—a reasonable result for magnetosomes, I contend.)

(p. 395) 8. Consumer-Based Teleosemantics

Millikan's theory of mental content (Millikan 1984, 1989b, 1990, 1991, 1993) is perhaps the best-known version of teleosemantics. Her writing on the topic has been prolific, and her view is rich and complex, although sometimes obscure. As with the other two authors, I must ride roughshod over complexities and matters of interpretation to give a thumbnail sketch of it. I hope that a thumbnail sketch will nonetheless be useful.

Millikan maintains that, if we want to discover the content of a mental representation, we should ignore the information processing that produces it (see especially Millikan 1989b, 500–502). **It is the use of representations that determines their status as representations and that determines their contents, she says.** To discover the content of a representation, on her theory, we look at the functions of its “consumers,” the devices that have historically used the representation to perform their proper functions (to perform, that is to say, that which becomes their function once they have been selected for it). A consumer system has a proper function if past tokens of the type did something that contributed to the preservation and/or proliferation of such a system in the relevant population.

This appeal to consumers is intended to provide a solution to the problem of error and the problem of functional indeterminacy. Consider the much-discussed example of the frog, which snaps at anything suitably small, dark, and moving, regardless of whether it is frog food. A frog cannot discriminate between moving flies and small plastic pellets tossed in front of it no matter how many pass its way. All such stimuli are apt to cause a certain response in the frog's optic tectum, which in turn causes a motivated frog to engage in catching behavior (e.g., snapping its tongue). What is the content of the relevant neural signal? We can describe the function of the frog's detection device as helping the frog to catch frog food, but we can also describe its function as detecting small, dark, moving things. Both are legitimate descriptions of its function.

According to Millikan, we can discover the content of the neural signal by considering its consumers, such as, in the case of the frog, its tongue-snapping system or even its digestive system. Ancestral frog digestive systems, for example, contributed to the preservation and/or proliferation of such systems in frogs by using the representation. That is, **they made this contribution in collaboration with the representation. The explanation of**

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

this preservation and/or proliferation—i.e., the explanation of how the frog digestive system contributed to its own preservation and/or proliferation—determines content in this case, says Millikan. To determine the content of a type of representation, we look at the past occasions when the consumer contributed to its preservation and/or proliferation by using the representation, and we ask what mapping between the representation and the world was required for such a contribution to take place. According to Millikan, (p. 396) the content of the relevant perceptual representation is frog food, because it was only when tokens of this type of representation were correlated with frog food that the frog got fed.

Millikan claims that her appeal to the consumers rather than the producers of representations solves the problem of error and the problem of functional indeterminacy. However, this is surely wrong. It is hard to see how it can be right if the functions of consuming systems are as “indeterminate” as the functions of producing systems are—and, of course, they are. The function of the frog's tongue-snapping mechanism is to catch food, but it is also its function to snap where and when it is instructed to snap. The function of the frog's digestive system is to feed the frog, but it is also its function to perform certain processes on whatever is ingested. What does more work for Millikan is her emphasis on what she calls the “Normal” condition—the environmental feature onto which the representation mapped that was *required* for past contributions to fitness. It is only when *R* mapped onto frog food that the consuming system's collaboration with *R* contributed to the replication of the frog's genes (Millikan 1991). Consider that, if the frog had caught something that was small, dark, and moving but that was not food, there would have been no contribution to its gene replication, and hence no positive selection pressure for the (production or) consumption of the representation. Whereas, if the frog had caught something that was food, but that was not small, dark, and moving, there would still have been a contribution to its gene replication, and hence the same positive selection pressure for the (production or) consumption of the representation.

Another simple system case that Millikan discusses is a male hoverfly's mate-detection system. Millikan tells us that the male hoverfly cannot distinguish between a female hoverfly and some distant birds and jet planes, and yet, she says, the hoverfly follows the rule *if you see a female catch it* because only female hoverflies, and not birds or jet planes, “actually causally contributed to the proliferation of his ancestors” (Millikan 1990, 330–35). Note that this is true whether we consider contributions by his mate-detection system (the producer of the representation) or whether we consider contributions by his reproductive system (a consumer of the representation).

Some argue that Millikan's theory makes for implausible content ascriptions. Paul Pietroski (1992) asks us to imagine some creatures that, we would intuitively say, can see red, but whose only benefit from doing so is that it attracts them to the top of a hill (to see the sunrise or sunset) and thus keeps them out of harm's way while some predators (the snorf) are prowling the valley below. The content of their perceptual representation is something like *snorf-free zone*, according to Millikan's theory. Pietroski argues that Millikan has confused the evolutionary explanation of their behavior with the intentional ex-

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

planation of their behavior. Millikan rejects Pietroski's appeal to intuitions here. However, even if one thinks, with Pietroski, that something has gone wrong, it would be too hasty to conclude that the problem lies with teleosemantics, as such. Arguably, Millikan's strategy for solving the functional indeterminacy problem is wrong and yet some version of (p. 397) teleosemantics might still be right. Dretske's revised theory, for example, goes Pietroski's way.

Others (e.g., Hall 1990) have pointed out that Millikan's strategy for determining content lacks brakes. Back to the male hoverfly: He will only reproduce if it is a female hoverfly that he has in his sights, and not a bird or a jet plane. But, by the same token, he will also only reproduce if she is a fertile female hoverfly, and only if she is catchable, and not about to be eaten by a bird, or die in a rainstorm, and only if her eggs are not eaten or washed away. To put the point another way, Millikan's theory of mental content seems to entail that every mental representation has the content that an opportunity for replicating the relevant replicators (whether they are, in a given case, genes, memes, etc.) has presented itself. Depending on which way you look at it, contents have become far too specific, or far too nonspecific. How are we to find the middle ground?

Discussions of mental content often focus on (relatively) simple cognitive systems like the frog and the hoverfly. It is a good strategy to start with simple systems, and work toward understanding more-complex ones, but in the end we want a theory that works for humans too. Dretske's theory has not been developed for anything other than simple perceptual representations. Millikan, however, maintains that her theory also works for more-sophisticated representations. This is a large topic, but it is worth noting that Millikan has several interesting strategies for going beyond simple systems. For example, as early as in her first (1984) book, she appeals to meme selection to determine the meanings of words, and in that book and elsewhere she speaks of general rules for mapping representations onto their representeds, which can have novel applications and confer novel contents under novel circumstances.

9. Desire-First Teleosemantics

Like Millikan, Papineau (1984, 1987, 1990, 1993) thinks we need to focus on the effects of representations—what they are used for—rather than on the processes that produce them. Perhaps the most significant difference between his view and Millikan's is the importance he places on a belief/desire psychological structure and on the role of intentional attitudes in determining mental content. He claims that creatures that lack such a psychological structure do not have determinate content. For instance, he questions whether frogs (let alone bacteria) have a belief/desire psychological structure, and he maintains that, if they do not, there is no problem of error in that case, since there is no fact of the matter about whether a mistake has been made (Papineau 1997).

In Papineau's theory, the contents of desires are primary and those of beliefs are secondary. He begins by characterizing a desire's satisfaction condition as “the (p. 398) effect which it is the desire's biological purpose to produce” (1993, 58–59), by which he means

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

that “[s]ome past selection mechanism has favored that desire—or, more precisely, the ability to form that type of desire—in virtue of that desire producing that effect” (1993, 59). Desires are in this case typed by their content. Thus, the desire to eat pizza is one type of desire, and the desire to eat pasta is another. Our desires, he says, have the function of causing us, in collaboration with our beliefs, to bring about certain conditions that enhanced fitness in the past. A general capacity for desire was selected for causing us to bring about conditions that contributed to our fitness, and more particular desires were selected for causing us to bring about more particular conditions that contributed to our fitness. These conditions are their satisfaction conditions, and they are the contents of the desires.

The truth condition of a belief, Papineau then tells us, is whichever condition must obtain if the desire with which it collaborates in producing an action is to be satisfied by the condition brought about by the action. A desire that has the function of bringing about the quenching of my thirst has the content that I will quench my thirst, since it was selected for bringing about thirst quenching. And if this desire collaborates with a belief to cause me to go to the fridge, the belief has the content that there is something thirst quenching in the fridge if it is the case that my desire for my thirst to be quenched would be satisfied by my going to the fridge only if it were true that there were something thirst quenching there.

Papineau's theory is not, in the main, a combinatorial theory. According to a combinatorial semantics, the meanings of complex representations are a function of the meanings of their parts and the syntactic (certain nonsemantic) relations between them. English and other natural languages are said to be combinatorial because we can create novel expressions (phrases, sentences, paragraphs, and so on) by combining smaller expressions (words, phrases, and so on) and because the meanings of the larger expressions are a function of the meanings of the smaller expressions plus their syntactic relations. Thus, “John kissed Mary” is a function of the meanings of “John,” “kissed,” and “Mary” plus their syntactic relations (the latter explaining why “John kissed Mary” differs in meaning from “Mary kissed John”). As we just saw, Papineau takes a top-down approach. The meanings of whole propositional attitudes are derived first. The meanings of concepts (the mental equivalents of words) are then derived from their roles in these attitudinal states.

This might seem to suggest that all of our beliefs and desires must have distinct evolutionary histories. That is to say, it seems to suggest that the desire for a pizza and the belief that there is one in the freezer must have a content-specific evolutionary history, distinct from that of the desire to eat pasta and the belief that the restaurant is open. However, Papineau does not hold fast to a top-down or noncombinatorial approach. He allows that there is some need for a combinatorial semantics, for example to accommodate novel beliefs and desires. He also wants to allow the relevant selection history to be either evolutionary or ontogenetic (1993, 59). Thus, individual learning can underwrite content ascriptions in Papineau's view. Psychosemantic norms are derived in part by bootstrapping

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

up from innate (p. 399) desires and beliefs, and in part by looking to individual learning histories, and in part by looking to some limited form of combinatorial semantics.

Millikan sometimes seems to hold a noncombinatorial view, and some of the early objections to her theory are based on this interpretation of her view (see, e.g., Fodor 1990, 64–69). As well as drawing our attention to the problem of novel belief, these objections draw our attention to the fact that some desires do not contribute to their own satisfaction (e.g., a desire to be immortal, or a desire that a hurricane change course) and to the fact that some desires do not tend to contribute to our fitness, even when they contribute to their own satisfaction (e.g., the desire to smoke cigarettes, or the desire to kill our children).

We have looked briefly at three different teleosemantic theories. They by no means exhaust the possibilities (see, for further examples, Dennett, 1995; Jacob 1997; McGinn 1989; Price 2001). Still, they give some idea of the potential variety. Each is highly controversial, and I have mentioned a few objections in passing. There are also objections to the general enterprise of trying to base psychosemantic norms on functional norms. Such objections are much more ambitious in their scope as they try to show that *no* teleosemantic theory can succeed.

10. Functional Indeterminacy

The aim of the three teleosemantic theories sketched above is to explain how content is determined, ultimately in terms of nonmental phenomena. As we have seen, simple systems are often used as illustrative cases. But one problem that we face when we try to assess how well these theories handle these simple system cases, is a lack of pretheoretic agreement as to the content of the representation under discussion. In relation to the case of the frog, for example, Dretske (in his later writings) seems happy with the idea that it is representing its stimulus as something small, dark, and moving (see also Neander 2006, who argues that this is the appropriate content ascription), whereas Millikan thinks that it is representing it as frog food and Papineau thinks that there might be no determinate content in this case. I'll put this problem—this lack of independent criteria—aside until the end of this section. It is distinct from, although related to, the functional indeterminacy problem, as I explain.

If the content of a mental representation is vague, then a theory should deliver content for that representation that is, to that degree, vague. To the extent that the content of a mental representation is determinate, it should deliver content that is as determinate. “The” problem of content indeterminacy is really many different problems, but in this section I look at the claim that teleosemantic theories have a *special* problem in this respect. The problem, introduced by Dretske and also addressed by Millikan and Papineau, is thought by some other philosophers to (p. 400) doom the general teleological enterprise. The problem is thought to be that function ascriptions are too indeterminate to make content ascriptions suitably determinate, and so teleosemantics must fail.

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

At this point, it helps to say a little more about a key notion in teleosemantic theories. Recall that they use a notion of a natural teleofunction—what an item was *selected for* by a natural process of selection. This is often elucidated by means of a contrast between selection *of* and selection *for*, the rough idea being that there is selection of something if it is preserved and proliferated in a population by means of selection (as opposed to, e.g., drift), and there is selection for something if it causally contributes to such selection (Sober 1984). It is notoriously difficult to precisely characterize selection, let alone selection of and selection for. Subtle issues are involved in distinguishing selection from drift, for example, which I do not try to explore here. Further, as Peter McLaughlin (2007) argues, a number of different notions (both ordinary and technical) can go by the name of “selection for,” and discussion can easily conflate them. However, somewhat more precisely (and borrowing to some extent from McLaughlin, although what follows is a bit different from what he says), the relevant distinction seems to me to be this.

- There was selection *of* a trait's tendency to Z in circumstance C iff the trait that did Z in C was fitter than available alternatives, and as a result it was preserved and/or proliferated in the population more than it would otherwise have been.
- There was selection *for* a trait's Z-ing in C iff the trait's Z-ing in C caused positive selection pressure in favor of the trait and its Z-ing in C.

This makes room for two possibilities. A trait's Z-ing in C can have been selected without selection *for* it or *for* its Z-ing in C (there can be free riders, which are fit but not adaptive). And there can have been selection *for* a trait's Z-ing in C without selection *of* it or *of* its Z-ing in C (there can, as McLaughlin emphasizes, also be paying customers who don't get a ride).

Fodor (1990) has led the attack on teleosemantics, and although he has revised his argument over time, his original claim is that teleosemantics cannot work because selection is “extensional” whereas content is “intensional.” (The appropriateness of the terms I've placed in scare quotes is questionable, and I drop them from the following discussion.) The claim, more precisely, is that, if doing Z in C^1 was adaptive, then doing Z in C^2 was equally adaptive if C^1 and C^2 co-occurred in the place and period in which the selection took place. This latter claim is true, as far as it goes, but what follows for teleosemantics? Fodor claims that it follows that teleosemantics cannot disambiguate contents. For example, in the case of the frog, he claims that it cannot disambiguate between the frog's representing the stimulus as something small, dark, and moving and its representing the stimulus as a fly, on the simplifying assumption that flies and small, dark, moving things co-occurred in the frog's natural habitat. If snapping was adaptive in the presence of a fly, then (p. 401) snapping in the presence of something small, dark, and moving was equally adaptive, on the simplifying assumption that flies and small, dark, moving things co-occurred.

In reply to Fodor, those who support teleosemantics point out that teleofunctions involve selection *for* rather than selection *of*, and that selection *for*, concerning as it does the causes of selection, can distinguish between C^1 and C^2 . On reflection, we can see that this

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

is a little misleading, because Fodor speaks of a trait's *adaptiveness*, whereas (on the above way of drawing the distinction) selection *of* just concerns fitness and not adaptiveness. Still, selection *for* concerns not just adaptiveness but also what *makes* something adaptive. This is crucial.

Suppose, for the moment, that we go along with Millikan and consider, in the case of the frog, one of the consumers of the frog's perceptual representation: its tongue-snapping mechanism. Let *Z* be tongue snapping. In addition, let C^1 circumstances be those where frog food is present and C^2 circumstances be those where something small, dark, and moving is present. Suppose also, for the sake of staying with Fodor's formulation, that C^1 and C^2 reliably co-occurred. Even so, it is correct to say that there was selection *for* *Z*-ing (tongue snapping) in C^1 rather than in C^2 , since it was *because* the circumstances were C^1 and not *because* they were C^2 that snapping in those circumstances was adaptive. Note that the preceding sentence is also true if C^1 and C^2 did not reliably co-occur. It is because enough of the C^2 circumstances were C^1 circumstances that snapping in those circumstances was adaptive. If you doubt this, imagine the counterfactual situation where frogs had always snapped in circumstances that were C^2 but not C^1 . That is, suppose that they always snapped in the presence of small, dark, moving things that were not frog food. The frog's tongue snapping in such circumstances would not—other things being equal—have been adaptive. At best, it would have been a waste of time and energy. So, at least in its original form, Fodor's objection fails. (I leave readers to explore his attempts to revamp the argument in Fodor 1996 and 2007.)

However, while selection *for* is more fine-grained than Fodor's argument requires, an appeal to it *on its own* definitely does not make content determinate. It may be a fine-enough tool, but we can use it in different ways. This is because traits are selected for complex causal roles that can be described in different ways, and there are, besides, many different mechanisms with somewhat different functions involved in each case (Neander 1995).

Fodor's original objection fails because it assumes that the appeal to selection is blunter than it is. Selection *for* is not so blunt, but it leaves a great deal of leeway. Merely linking content to teleofunctions is but a first step toward developing a full theory. That is just to say that the notion of selection *for* does not decide between different teleosemantic theories: We have to do that. But to say that we have to choose *between* teleosemantic theories is not to say that *within* each teleosemantic theory an implicit appeal to our choice must be made. Unfortunately for those who want to ring the death knell for teleosemantics, this means that there is no successful general argument against teleosemantics here. Each theory has to be considered individually.

(p. 402)

I shall shy away from discussing whether teleosemantics can in the end answer Kripkean arguments (see Millikan 1990; Boghossian 1989). The point here is that we have no reason to think that teleosemantics cannot provide content that is at least as fine-grained as causation is. However, Kripkean skepticism pushes beyond this, to argue that this will not

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

be fine-grained enough (it is therefore aimed at, for example, Fodor's own theory, as well as at teleosemantics). In the end, as I see it, a response to Kripke involves a solution to Nelson Goodman's grue problem (1955), but this is not something that I want to tackle here.

To return to the problem mentioned at the start of this section: It would help if we knew which determinate contents to aim for. With different people aiming for different contents, it is hard to make progress, and it is hard to assess what progress has been made. Some readers might think that we should restrict our attention to cases where the contents are not in dispute, but these belong to concepts that humans consciously possess, and they are controversial for other reasons. We understand much less well their neurological substrates, and much less well their structures, their cognitive roles, and their acquisition.

11. Swampman

We end with a look at a very different kind of objection to teleosemantics. There is much difference of opinion as to its significance. This is the so-called Swampman objection. Some are persuaded by it; others will object to my even deigning to discuss it. Those in the latter camp find it not just unpersuasive but, worse, irrelevant. My own view lately lies midway. I don't believe Swampman should persuade us to abandon teleosemantics if we are convinced of its worth on other grounds, but I do think he is a troublesome creature when he is presented in a certain way. In this section, I will set out in which ways he is and isn't troubling.

First, what is a Swampcreature? Swampcreatures pop into existence accidentally, through random collisions of particles. Christopher Boorse (1976) raises the possibility of such creatures against Wright's etiological theory of functions. Swampman, in particular, is Donald Davidson's (1987) creation, introduced as a possible objection to his own theory of mental content, which involves an appeal to a history of causal interaction with the environment. I'll return to Boorse's Swampcreature in a moment. First, let me introduce Swampman.

As the story goes, Swampman pops into existence as a result of lightning hitting a swamp. At his inception he is, in all intrinsic (nonrelational) and synchronic (at a time, as opposed to over time) respects, physically identical to Davidson. For a moment in time, they are molecule-for-molecule—in fact, subatomic particle/property for subatomic particle/property—identical. Assuming physicalism, there is no intrinsic and synchronic qualitative difference between them. (p. 403) Crucially, Swampman is not *copied from* Davidson, either intentionally or accidentally. Their being doppelgangers is just a stupendous coincidence. Swampman has *no* history of selection, either ontogenetic or phylogenetic, nor is he copied from anything that has such a history. Nor is God nor a mad scientist involved. Contra both Paley and Darwin, this is a case of apparent design without a designer *and* without natural selection. The case is imaginary but, according to those who

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

raise the Swampman objection, it is consistent with the laws of nature, and hence it is a nomological possibility, though the probability is, to say the least, miniscule.

Is Swampman possible? I leave this to the scientific experts. In my view, the truth of teleosemantics does not stand or fall with that question: Even were Swampman to exist, teleosemantics could still be true. So let's assume that Swampman is possible, for the sake of the argument. In fact, let's assume he has arrived, has washed off the mud, and is knocking at the door. Does he have intentionality? Of course, there is no doubt that he has something that is superficially and in some respects quite deeply *like* intentionality. But does what he has count as *intentionality* properly so called? His having intentionality—his having normal normative mental content—is inconsistent with teleosemantics. Is this a problem for teleosemantics?

Let's sit on that question for a short while and in the meantime return to Boorse's creatures. These are less troubling than Swampman in my view, and it helps to consider them first. Imagine that a doppelganger of a rabbit pops into existence by the same process—an instantaneous saltation, a coincidence of cosmic proportions. Do its parts have normal normative functions? Their having such functions is inconsistent with an etiological theory of them, but is this a problem for the theory? Boorse's claim is that we can ascribe normative functions to the parts of Swampbunny even though Swampbunny has no selection history. Now, of course, the question is not whether we could or would ascribe such functions to Swampbunny, but whether we should do so, whether we would be correct to do so. And correctness, as we know, is not merely a matter of how we are disposed to apply terms. Intuitions about what we could or would say do not really prove anything either way. Some readers might want to respond that Swampbunny intuitions at least demand an explanation, but that explanation is not hard to find for a supporter of an etiological theory. On the theory, functions belong to systems that are the products of selection, and Swampbunny, by hypothesis, has all of the hallmarks of something that can only—for all practical purposes, at least—have resulted from selection.

Nonetheless, this is not the end of the argument. Hypothetical examples can play a useful role, if they can help to elucidate the theoretical roles of the theoretical notions that are under analysis. (David Lewis's 1980 mad pains and Martian pains are cases in point.) Those who press Swampcreature objections are, I think, mostly trying to use them to do something along these lines, and Swampcreature objections are most charitably assessed in these terms. Consider Swampbunny again. It is a doppelganger, let's suppose, of a real bunny called Snow. Each part of Snow corresponds to a part of Swampbunny, and each part of Swampbunny has the same (p. 404) causal dispositions as the corresponding part of Snow. That's all true, by hypothesis. Their corresponding parts therefore have the same functional roles, although they do not have the same teleofunctions, since a history of selection is required for teleofunctions, and Swampbunny does not have this. This much is definitional. Now, one line of thought goes like this: Swampbunny shows that we do not *need* teleofunctions. It provides a vehicle with which we can probe what is needed in the way of functions for explanation and prediction in physiology. If Snow's heart can pump blood, so can Swampbunny's "heart," regardless of the fact that Snow's heart has the

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

teleofunction of pumping blood and Swampbunny's "heart" doesn't. Ditto for every part of Snow and its corresponding part in Swampbunny. So, functional roles should suffice for physiology, or so the argument goes, because whatever operational explanations apply to Snow must equally well apply to Swampbunny. What we need to know, in order to explain or predict their survival, for instance, is whether their heart and "heart" will pump blood, and not whether they are "supposed" to do so. Think about what would happen were we to take Swampbunny to the vet: The vet would not know the difference and could treat Swampbunny just as effectively as she could treat Snow. On the plausible assumption that biology is in the business of explanation (and prediction), the argument concludes, physiology only needs functional roles, not teleofunctions.

In my view, the problem with this argument is not that it appeals to an, at best, wildly improbable hypothetical example. As I see it, the problem is instead that it relies on an understanding of physiology that is superficial. It fails to understand the need for normative functions in physiology. In physiology, the normative notion of function is used to provide an *idealized* functional analysis of a *normal* system, which permits a useful kind of general description and explanation that would otherwise be unavailable. It is true that a vet could treat Swampbunny just as effectively, but that is just an artifact of the example. All rabbits at the vet are treated as better or worse approximations of a normal rabbit system, and there is nothing that would stop Swampbunny being so treated. Still, that doesn't speak to the question of how we *get* the idealization.

Boorse understands the need for an idealized functional analysis. However, he claims, in effect, that the parts of Swampbunny have the same *normative* functions as Snow's parts do, on his ahistorical and statistical theory of function.³ According to Boorse's theory, the normal function of a biological component (such as a heart) is the typical contribution of components of that type to survival and reproduction within a species or an age group or a sex within a species. The Swampbunny kind of case is intended to bolster his claim that his theory is preferable to Wright's etiological theory, though it is not intended to be decisive on its own.

Let's consider the question of whether Boorse's theory bestows the same normative functions on Swampbunny as it bestows on Snow. To apply his theory to Swampbunny, we must first decide if Swampbunny is a member of an actual rabbit species or if it is *sui generis*. Let's consider both options to see where they lead. First, suppose that Swampbunny counts a member of an extant species. Two (p. 405) problems with that come to mind. For one, we are being asked to abandon all historical interpretations of the species concept, and there are well-worn problems with that. For another, this concession won't help all Swampcreatures. It is only an artifact of the specific example that Swampbunny is so similar to some real creatures. Imagine a thlunkiquot popping into existence. The thlunkiquot, not being similar to any real creature, cannot take advantage of this first tack. Now, let's consider the second option. Suppose that we are meant to count Swampbunny as *sui generis*, as an individual who constitutes its own new species. Then, in that case, the function ascriptions that we make with respect to Swampbunny's parts are no longer normative, because there is no distinction to be drawn between what its parts do

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

and what they are “supposed” to do. Putting this aside for the moment, Swampbunny's parts would anyway not have all of the same normal functions as Snow's, if some of Snow's are malfunctioning. Suppose some part of Snow would do *Z* in *C* if it were functioning normally, but it is malfunctioning and cannot do *Z* in *C*. Then, in that case, the corresponding part in Swampbunny cannot do *Z* in *C*. In its case, though, there is no basis for saying that it has the function to *Z* in *C* or, hence, that it is malfunctioning. For that type of part does not typically do *Z* in *C* in Swampbunny's species: 100% of them do not do *Z* in *C* in Swampbunny's species. Thus, it is not so easy to bestow normal functions on Swampbunny, even if one wants to. Superficial appearances can be deceptive. Perhaps only the products of selection possess normal functions because functional norms depend on what there was selection for. Swampbunny leaves that open.

What of Swampman? Some of the same considerations apply. The question is not whether we could or would ascribe contents to Swampman; the question is whether we would be right to do so. The most that an appeal to Swampman can do is to cast light on the theoretical role that intentionality plays. However, we ought not dismiss too quickly the possibility that Swampman can do this in interesting and even potentially damaging ways.

Consider, for example, whether it would be okay to roast Swampman on a spit.⁴ Put aside any doubt as to Swampman's origins. Suppose we somehow *know* that the being knocking at the door lacks the kind of pedigree that teleosemantics requires for intentionality (just suppose). Would it be wrong to barbecue him? Some think that we should be unconcerned about Swampman's roasting if he lacks normative mental content, and so they think that our repugnance at the idea of the roasting reveals something about Swampman's psychology—i.e., that he must have intentionality. The example at least suggests that adjustments might be necessary somewhere. But to my way of thinking, the wrongness of the barbecue reveals that normative mental content is not crucial for moral concern. It is true and appropriate that ethicists give an important place to preferences, but arguably Swampman shows that, if there were an individual who merely possessed the narrow content equivalents of preferences, this individual would be of moral concern as well. In this discussion, it should be kept in mind that Swampman, by hypothesis, has narrow content. This is trivial given that narrow content is defined as whatever most closely approximates normal mental content that supervenes on the intrinsic (p. 406) and synchronic properties of individuals. What is not trivial is what this amounts to. If teleosemantics is true, it does not amount to normal normative content.

Another argument that is troubling, although also arguably inconclusive, points to the relation between intentionality and phenomenal consciousness. The argument is best put like this:⁵

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

P1	Davidson (D) and Swampman (S) are intrinsic, synchronic, duplicates at time <i>t</i> . (by hypothesis)
P2	The phenomenal consciousness of an individual supervenes on his or her intrinsic, synchronic, properties at the time in question. (assumption)
C1	The phenomenal consciousness of D and S are identical at <i>t</i> . (from P1 and P2)
P3	If an individual is in pain, then it will seem to him that he is in pain. (assumption)
P4	D is in pain at <i>t</i> . (by hypothesis)
C2	S is in pain at <i>t</i> , and it will seem to S that he is in pain. (from C1, P3, and P4)
P5	If it seems to someone that something is the case, then that individual has an intentional state. (assumption)
C3	Therefore, S has an intentional state. (from C2 and P5)

Some proponents of teleosemantics reject P2. For example, Fred Dretske (1995, 1996) has a representational theory of phenomenal consciousness that entails that Swampman lacks phenomenal consciousness. For my part, I accept P2, but suspect a problem with P3 (or possibly P5). I question whether, if Swampman experiences pain, it must seem to him that he does (or, if it must in *some sense* seem to him that he does, whether this seeming so is a full-blooded intentional state—one to which the usual psychosemantic norms apply). Either approach to defending teleosemantics against this objection will take a considerable amount of work. (For further discussion of the Swampman objection, see issue 11.1 of *Mind and Language*, Braddon Mitchell and Jackson 1997, and Papineau, 2001)

The motivations for teleological theories of mental content are powerful, and in my view some sort of teleological theory is needed. However, there are devils in the details. None of the theories described here seems entirely successful, even if we restrict our attention to relatively simple cases. In addition, there are objections raised against the general enterprise. While they are inconclusive, there are challenges that remain.

References

Allen, C., Bekoff, M., & Lauder, G. (eds.) 1998. *Nature's Purposes: Analyses of Function and Design in Biology*. Cambridge, Mass: MIT Press.

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

Ariew, A., Cummins, R., & Perlman, M. (eds.) 2002. *Functions: New Readings in the Philosophy of Biology and Psychology*. Oxford: Oxford University Press.

Block, N. 1986. "Advertisement for a Semantics for Psychology." In P. French, T. Uehling, & H. Wettstein (eds.), *Studies in the Philosophy of Mind: Midwest Studies in Philosophy*, 10. Minneapolis: University of Minnesota Press.

Boghossian, P. 1989. "The Rule-Following Considerations," *Mind* 98:507–49.

Boorse, C. 1976. "Wright on Functions." *Philosophical Review* 85:70–86.

Boorse, C. 2002. "A Rebuttal on Functions." In A. Ariew et al. (eds.), *Functions: New Readings in the Philosophy of Biology and Psychology*, 63–112. Oxford: Oxford University Press.

Braddon-Mitchell, D., & Jackson, F. 1997. "The Teleological Theory of Content." *Australasian Journal of Philosophy* 75(4):474–89.

Brentano, F. [1874]1911. *Psychologie vom empirischen Standpunkt*. Leipzig: Duncker & Humblot.

Buller, D. (ed.) 1999. *Function, Selection and Design*. Albany: State University of New York Press.

Chisholm, R. 1957. *Perceiving: A Philosophical Study*. Ithaca, N.Y.: Cornell University Press.

Cummins, R. 1975. "Functional Analysis." *Journal of Philosophy* 72:741–65.

Davidson, D. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60:441–58.

Dennett, D. 1995. *Darwin's Dangerous Idea*. New York: Simon & Schuster.

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

——. 1986. "Misrepresentation." In R. Bogdan (ed.), *Belief: Form, Content and Function*, 17–36. New York: Oxford University Press.

——. 1988. *Explaining Behavior*. Cambridge, Mass.: MIT Press.

(p. 408) ———. 1991. "Dretske's Replies." In B. McLaughlin (ed.), *Dretske and His Critics*, 180–221. Cambridge, Mass.: Blackwell.

——. 1995. *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.

——. 1996. "Absent Qualia." *Mind and Language* 11(1):70–130.

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

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

——. 1990. *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press.

——. 1996. "Deconstructing Dennett's Darwin." *Mind and Language* 11:246–62.

——. 2007. "Why Pigs Don't Have Wings." *London Review of Books* 29(20):19–22.

Frege, G. [1892]1960. "Über Sinn und Bedeutung." Translated as "On Sense and Reference," in P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell.

Goodman, N. 1955. *Fact, Fiction, and Forecast*. Cambridge, Mass.: Harvard University Press.

Grice, H. P. 1957. "Meaning." *Philosophical Review* 66:377–88.

Hall, R. 1990. "Does Representational Content Arise from Biological Function?" *Philosophy of Science Association* 1:193–99.

Jacob, P. 1997. *What Minds Can Do: Intentionality in a Non-Intentional World*. Cambridge: Cambridge University Press.

Kripke, S. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, Mass.: Harvard University Press.

Lewis, D. 1980. "Mad Pain and Martian Pain." In N. Block (ed.), *Readings in the Philosophy of Psychology*, 1:216–22. Cambridge, Mass.: Harvard University Press.

Loewer, B. 1987. "From Information to Intentionality." *Synthese* 70:287–317.

McGinn, C. 1989. *Mental Content*. Oxford: Basil Blackwell.

McLaughlin, B. (ed.) 1991. *Dretske and His Critics*. Cambridge, Mass.: Blackwell.

McLaughlin, P. 2007. "On Selection Of, For, With, and Against." In P.

Machamer and G. Wolters (eds.), *Causation: Historical and Contemporary Perspectives*. Pittsburgh, Pa.: University of Pittsburgh Press.

Millikan, R. 1984. *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.

——. 1989a. "In Defense of Proper Functions." *Philosophy of Science* 56(2):288–302.

——. 1989b. "Biosemantics." *Journal of Philosophy* 86:281–97.

——. 1990. "Truth, Rules, Hoverflies and the Kripke Wittgenstein Paradox." *Philosophical Review* 99:232–53.

——. 1991. "Speaking Up for Darwin." In B. Loewer & G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*, 151–65. Cambridge, Mass.: Blackwell.

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

——. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, Mass.: MIT Press.

——. 2000. *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.

Neander, K. 1991a. "Functions as Selected Effects." *Philosophy of Science*, 58(2):169–84.

——. 1991b. "The Teleological Notion of Function." *The Australasian Journal of Philosophy*, 69(4):454–68.

——. 1995. "Malfunctioning and Misrepresenting." *Philosophical Studies* 79:109–41.

——. 1996. "Dretske's Innate Modesty." *Australasian Journal of Philosophy* 74(2):258–74.

——. 2002. "Types of Traits: The Importance of Functional Homologues." In A. Ariew et al. (eds.), *Functions: New Readings in the Philosophy of Biology and Psychology*, 390–415. Oxford: Oxford University Press.

(p. 409) ——. 2006. "Content for Cognitive Science." In G. McDonald and D. Papineau (eds.), *Teleosemantics*. Oxford: Oxford University Press.

Papineau, David. 1984. "Representation and Explanation." *Philosophy of Science* 51:550–72.

——. 1987. *Reality and Representation*. Oxford: Basil Blackwell.

——. 1990. "Truth and Teleology." In D. Knowles (ed.), *Explanation and Its Limits*, 21–44. Cambridge: Cambridge University Press.

——. 1993. *Philosophical Naturalism*. Oxford: Blackwell.

——. 1997. "Teleosemantics and Indeterminacy." *Australasian Journal of Philosophy* 76:1–14.

——. 2001. "The Status of Teleosemantics; or, How to Stop Worrying about Swampman." *Australasian Journal of Philosophy* 79(2):279–89.

Pietroski, P. 1992. "Intentional and Teleological Error." *Pacific Philosophical Quarterly* 73:267–81.

Price, C. 2001. *Functions in Mind: A Theory of Intentional Content*. Oxford University Press.

Proust, J. 1994. "Naturalizing Intentionality through Learning Theory." In R. Casati, B. Smith, & G. White (eds.), *Philosophy and the Cognitive Sciences: Proceedings of the 16th International Wittgenstein Symposium*, 233–45. Vienna: Verlag Hölder, Pichler-Tempsky.

Sober, E. 1984. *The Nature of Selection*. Chicago, Ill.: University of Chicago Press.

Teleological Theories of Mental Content: Can Darwin Solve the Problem of Intentionality?

Wright, L. 1973. "Functions." *Philosophical Review* 82:139–68.

Notes:

(1.) Some like to claim, in this context, that the notion of a functional role can be modified and can be made normative without collapsing it into the teleological notion. This is true, although then it would be a different notion. There are alternative accounts of normative function to be found in the philosophy of biology literature: For example, as I mention in the last section, some think functional norms are statistical. However, the people who push the line that functional roles can be modified usually dislike this suggestion, without offering any developed alternative. The work must be done before such a vague claim can be evaluated. In any case, I here draw a stark contrast between, on the one hand, theories that employ a nonnormative notion of a functional role and, on the other hand, theories that employ a normative notion that is construed as teleological or, in other words, in terms of what things were selected for.

(3.) Boorse (2002) does not like this way of speaking. In Boorse-speak, traits that are malfunctioning do not *have* the functions they cannot perform. However, in Boorse-speak, we can still speak about the normal functions of traits of the type, which malfunctioning traits lack and which count as malfunctioning due to this lack.

(4.) I have heard Frank Jackson pose this question, though he has not pressed the argument in print.

(5.) Something similar was once suggested to me by David Chalmers.

Karen Neander

Karen Neander, a professor of philosophy at Duke University, has published numerous papers in philosophy of mind and philosophy of biology. These have been on functions, selection, typology, teleosemantics, and other naturalistic theories of mental representation, pictorial representation, and consciousness. She has a forthcoming book titled *Mental Representation: The Natural and the Normative in a Darwinian World* in which she argues that a Darwinian perspective can solve the problem of intentionality. Her dissertation was completed at La Trobe University, Melbourne, Australia.