

Misrepresenting & Malfunctioning

Author(s): Karen Neander

Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Aug., 1995, Vol. 79, No. 2 (Aug., 1995), pp. 109-141

Published by: Springer

Stable URL: <https://www.jstor.org/stable/4320604>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*

MISREPRESENTING & MALFUNCTIONING*

(Received in revised form 5 December 1993)

I. INTRODUCTION

Consider the following explanation of how you determine the direction of sound. Your brain exploits the fact that a sound coming from your left hand side enters your left ear fractionally before it enters your right ear, and it does so by each of your ears recording the pattern of sound entering it, passing this information on to the auditory centre, which then matches the two patterns to determine which ear received which pattern first.

It might be thought that this process, described as information processing, involves intentionality, but I don't want to assume this at the outset. Rather, I want to consider a question that bears on its intentionality. Many people have thought that a minimal condition for a process being intentional is that it involve states – which we can call representations – that are capable of *misrepresenting*, and my question is whether a process like the above could satisfy this minimal condition. To simplify matters, I am going to assume that any information bearing states involved in this processing are discrete information bearing states. While this is controversial (radical Connectionists and Gibsonian psychologists would protest that it does not) its denial only complicates my account in inessential ways (except at one point which I will mention later). So, assuming that there are discrete information bearing states involved, the open question is whether these states are capable of misrepresenting, or whether they are more aptly thought of as natural signs with only 'natural meaning.'

I chose the opening example as an illustration of information processing that is presumably pre-linguistic and innate, yet also human,

Philosophical Studies 79: 109–141, 1995.

© 1995 Kluwer Academic Publishers. Printed in the Netherlands.

although probably not specifically human; this being where biologically inspired teleological theories of content most plausibly enter the semantic story. One strategy for explaining intentionality is to try to ground basic intentional properties in biological facts and to build up somehow from there.¹ It is a plausible overall strategy in my opinion, but I am not offering a general defence of it in this paper. This paper is confined to the groundwork, in particular it is about whether the capacity to misrepresent can be grounded in the biological facts.

Perhaps surprisingly, my main aim is not to provide a solution to what has become known as the ‘disjunction problem.’ Although this problem needs solving, I argue that we don’t need to solve it in order to solve the misrepresentation problem, more broadly construed. However, I hope this paper does make a useful contribution to the debate on the disjunction problem, for it teases that problem apart from another problem, also deriving from ‘functional indeterminacy,’ with which it has become muddled. The bad news is that there is more than one source of functional indeterminacy, and the different sources give rise to different problems for teleological accounts. This is the burden of the first half of this paper. The good news is that once we have a better understanding of these problems, finding solutions looks easier, and in the second half of this paper I suggest a novel approach to teleosemantics, one that, I think, emerges quite naturally from an improved understanding of the misrepresentation problem.

II. BIOLOGICAL NORMS

First, a few introductory words about functions. Elsewhere I have argued that the notion of a ‘function’ that plays a special and central role in biology is a teleological notion, best described by an etiological theory.² It helps to have this theory explicitly before us, so I will begin by briefly outlining it – details and defences are to be found in other places.

The basic idea of the theory is this: a function (a proper function, otherwise sometimes called a teleo-function) is an effect for which a trait or item was selected. In straightforward biological cases, the selection process involved is natural selection, but other selection processes

generate functions as well. Intentional selection is involved in the case of artefacts, for instance. However, given that our interest in this paper is with innate information processing in natural cognitive systems, we can ignore other selection processes and focus on functions which derive from natural selection alone. For these functions the following definition seems to be most conservative and explanatory:

Def: Some effect (Z) is the proper function of some trait (X) in organism (O) iff the genotype responsible for X was selected for doing Z because doing Z was adaptive for O's ancestors.

So, on this definition, the proper function of a kangaroo's pouch is to protect her joeys because this is the effect of ancestral pouches that was adaptive and that caused the underlying genotype to be selected.

As many people have remarked, this notion of a 'function' is normative, a fact that is reflected in the questions we often like to ask about biological traits, such as: Is it functioning *properly*? What is it *supposed* to do? Is it *malfunctioning*? The main normative distinction is between proper functioning, on the one hand, and malfunctioning (or dysfunctioning), on the other. As you will see, this distinction is significant in this paper. Roughly speaking, a biological part functions properly when it can do what it was selected for and malfunctions when it cannot. (The matter is a little more complex, as I will explain in a while.)

It might help to note that the normativity of biological functions is neither simply evaluative or statistical. To see this consider a few examples. Teenage fertility is biologically normal, but it does not follow that teenage fertility is a good thing: on the contrary, if we could induce (temporary and reversible) infertility in all girls under the age of twenty, that would probably be better [Boorse, 1975]. Judging that something is functioning properly is not the same as judging that its functioning is good.³ Nor is the judgement that something is functioning properly just a statistical abstraction, as epidemic and pandemic diseases testify. If we were all struck blind it would still be the function of our eyes to see. Sight, not blindness, would remain biologically normal proper functioning, and blindness, not sight, would remain dysfunctional. Not surprisingly, we can't cure diseases just by spreading them around.⁴

This normative notion of a ‘function’ has an important theoretical role in biology (which I briefly sketch later in section 3A). But my main aim in this paper is not to examine its present role in biology, it is rather to examine its potential role in semantics and psychology. The basic idea behind teleological theories of content is that this normative notion – and its distinction between proper functioning and malfunctioning – might somehow underwrite the normative notion of content – and its distinction between representation and misrepresentation.

Since Darwin, biological norms can be analysed without appealing to any unanalysed intentional or theological notions (like Divine Design), so biological norms are in the relevant sense naturalistic. Moreover, assuming realism about intentionality, the brain is a biological organ selected for information processing, which makes information processing one of its biological functions. So it’s to be expected that biology, with its normative talk of functions, has the wherewithal to describe something so clearly within its domain – and to describe it, moreover, *as* information processing. At this point in intellectual history, it would be foolish not to take a long hard look at whether biological norms can underwrite semantic norms, at least in the case of innate representation.

III. SOURCES OF INDETERMINACY

This is all very well, you might think, except that some philosophers believe they have already looked long and hard enough and concluded that biological norms cannot underwrite semantic norms. They have a number of objections to the general teleosemantic strategy, principally three: (i) the swamp-baby (history-less doppleganger) objection (ii) the general worry about how teleosemantic theories can be extended to more sophisticated cases than those tackled so far, and (iii) a cluster of problems that go under the umbrella of ‘indeterminism problems.’ (i) and (ii) are serious objections, but they are not my concern here: here I am focussing on (iii), and I don’t aim to exhaustively treat all aspects of this. My interest is principally with the indeterminism which is thought to plague teleosemantic theories due to the corresponding indeterminacy of biological functions.

We need a better understanding of why function attributions (that is, statements that attribute functions to things) are indeterminate, so let's start by setting aside two dimensions of 'indeterminacy' that are irrelevant to the present discussion.

One irrelevant feature of function attributions is the vagueness that is due to borderline cases. Take for instance the classic case of the speckled moth. There was an initial population of individuals with light coloring as a result of selection for camouflage. A recurring mutation produced an occasional darker moth, but these were more easily spotted by predating birds and were selected against. When the trees on which the moths rested darkened during the Industrial Revolution the selective situation reversed, until the genotype responsible for the darker version was selected instead. Now, there is no precise point in time at which the light coloration loses its function of camouflage and the dark coloration acquires it, but this vagueness is not the kind of indeterminacy that troubles people in the teleosemantic debate.

Nor is it to the point that a particular trait can have a plurality of functions. Traits can be selected for more than one effect and have more than one function without any conflict in biological norms arising. Feathers can have the function of aiding flight, regulating temperature and looking sexy to conspecifics, for example. Deformed feathers might function properly with respect to temperature regulation while malfunctioning with respect to flight, but we can simply relativise the normative judgement to the specific function.

The genuinely problematic cases are not so easily dealt with. The problem is that for any given trait and any given function there seems to be more than one way to describe that function, and conflicting judgments concerning biological norms can apparently be derived from the different descriptions. It helps to have an example before us, so let's consider the anaerobic bacteria described by Fred Dretske. Certain marine bacteria in the northern hemisphere have internal magnets that align themselves to the Earth's magnetic field. As long as the bacteria remain in the northern hemisphere, this directs the bacteria downward, away from the oxygenated waters at the surface where the bacteria cannot survive, since to be aligned to the magnetic field is to be headed toward deep water. Dretske asks whether these magnetosomes have

the function of indicating anaerobic conditions or whether they have the function of indicating the prevailing magnetic field and adds that until we have an answer to this question, we have no solution to the indeterminacy problem in teleosemantics [Dretske 1988, p. 63].

An important but overlooked fact is that there are several sources for the indeterminacy encountered in this example. I am going to carve them into two, and dub them the ‘biological problem’ and the ‘philosophical problem’: – both are problems for philosophers but you will see why the names are descriptive. The biological problem is that *there is always more than one description of a function which even a biologist would find plausible*. The philosophical problem is that *beyond the biologically plausible descriptions there are a whole host of other descriptions which no biologist (qua biologist) would take seriously for a moment, but which, perhaps, can only be excluded by appealing to intentional notions (and so the attempt at naturalising intentionality fails)*.

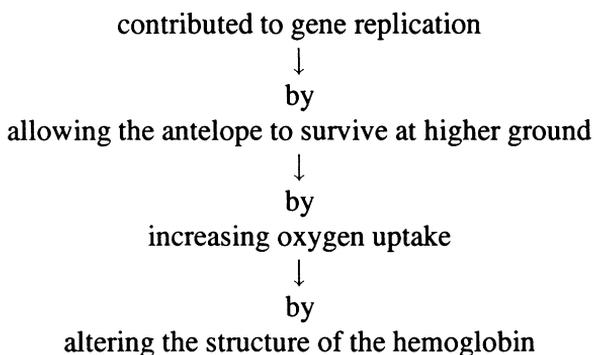
These two problems are orthogonal I believe, and I will discuss them in turn, the biological problem first.

A. *The Biological Problem*

According to the definition favoured in this paper, the function of a trait is to do whatever it did in ancestral creatures that was adaptive and caused the underlying genotype to be selected. Now consider Dretske’s question in the light of this. What we want to know is what the magnetosomes did in ancestral bacteria that was adaptive and caused the underlying genotype to be selected: align themselves with the Earth’s magnetic field or direct the bacteria to anaerobic conditions? The answer seems to be both, for both were done, both were adaptive, and both caused the underlying genotype to be selected. Of course, they were not done independently, nor did they independently cause the underlying genotype to be selected. Rather, it was *by* doing the one that the other was done.

A second example will prove useful. Suppose a trait in an antelope population altered the structure of the hemoglobin, which caused higher oxygen uptake, which allowed the antelope to survive at higher ground to which they were forced to retreat. Suppose also that, as a result,

this trait was selected. What did it do that was adaptive and caused the underlying genotype to be selected? Answer: all of these things. One and the same trait can have the function of contributing to survival and reproduction, allowing the antelope to move to higher ground, increasing oxygen uptake and altering the structure of the hemoglobin, because all of these are correct descriptions of what the trait did which caused it to be selected.⁵ We can present the case diagrammatically as follows (the trait):



As I hope the diagram makes clear, these different descriptions are linked by a 'by' relation of some kind – the trait does one thing *by* doing another. What is this *by*-relation?

Co-extensional descriptions can generate something that might go by the name of a 'by' relation. For example, as I write, Paul Keating is the Prime Minister of Australia, so I could wave at the Prime Minister of Australia *by* waving at Paul Keating. Perhaps this is why people sometimes think that functional indeterminacy derives from co-extensionality. But this is *not* the source of the different descriptions in the diagram above. Notice that a *by*-relation that depends on co-extensionality is a symmetrical relation (I can wave at the Prime Minister *by* waving at Paul Keating and *vice versa*) whereas the *by*-relations in the diagram above are not symmetrical (the trait causes a greater uptake of oxygen *by* altering the structure of the hemoglobin, but it does *not* alter the structure of the hemoglobin *by* causing a greater uptake of oxygen). Nor are the descriptions in the diagram co-extensional. If this isn't immediately obvious, just reflect on the fact that a second trait

(e.g., bigger lungs) might also increase oxygen-uptake, and a particular antelope might have this trait without having the other, thus having a trait that increases oxygen uptake without having a trait that alters the structure of the hemoglobin.

So co-extensionality cannot be the source of the plausible variations in descriptions of functions. The by-relation illustrated in the diagram is rather a relation between mechanism and 'purpose.' The lower descriptions describe the underlying mechanism (how something is done) and the higher levels explain why doing that was adaptive and selected for.

I also want to suggest that plausible function attributions generally have a place in a functional analysis of the organism, but this suggestion needs some preliminary discussion, because familiarity with the idea of functional analysis (at least amongst philosophers of mind) is due mainly to Robert Cummins [1975] who argued that the notion of function involved in functional analysis was *not* the teleological/etiological notion.⁶ I disagree and I will explain why.

Up to a point, Cummins is right about the nature of functional analysis. As he explains, when biologists provide a functional analysis of an organism they conceptually decompose it into simpler and simpler parts, describing what each part contributes to the overall activities or capacities of the organism. So in physiology, the human organism is decomposed into the major 'systems' (digestive, circulatory, respiratory, reproductive, immune, nervous, and so on) and the contribution made by each of these systems to the functioning of the whole is described. These systems are then decomposed into their parts (e.g., the digestive system into the mouth, oesophagus, stomach, liver, pancreas, and so on) and the contributions that each of these parts to the functioning of that system are described. These parts are then further decomposed into their parts (e.g., the mouth into the saliva glands, teeth, and tongue, and so on) and these into theirs (e.g., the tongue into muscles, sensory receptors, and so on) down to the level of individual cells and their sub-cellular components, the causal contribution of each part being described along the way.

But Cummins also argued that a non-teleological notion sufficed for this kind of analysis (see n. 3. this paper). And while it would suffice

for the recursive decompositions found in, say, chemistry or astronomy (although these are not standardly called ‘functional analyses’), biology is different. As well as being massively more complex, different kinds of biological systems (e.g., humans, dogs, frogs) are repeated millions or billions of times, with countless idiosyncratic variations both between organisms and also within organisms across time. In practice, selection over randomly generated mutations is what makes this complexity possible, and this selection also makes a particular kind of idealized description possible, one that abstracts away from idiosyncratic differences. This level of description involves the teleological notion of function, the notion of what something was selected to do. Its involvement in functional analysis is twofold.

Firstly, the parts into which the physiological system is conceptually decomposed are categorised in terms of their proper functions. (Just which description of their function is definitive I will come to in a moment). Most (if not all) physiological categories are functional categories: for example, a heart is a heart because it has circulating blood as its proper function [Beckner, 1959]. No purely morphological or dispositional characterisation of hearts will do the job because the category must embrace pathological diversity and differences across species as well. A badly diseased heart need not look like a normal heart, nor be capable of doing what a normal heart does. But it is still a heart in virtue of the fact that it has the proper function of circulating blood (This should seem a familiar idea, because categories of artefacts are similar: a brake is a brake in virtue of what it is supposed to do – was intended or designed to do – not in virtue of having some specific structure or disposition. Brakes come in different designs, and a brake that fails is still a brake, albeit a malfunctioning one.)

Secondly, within the primary functional analysis of physiology, the aim is not to describe the contributions that these various components presently and actually make to the overall functioning of the system, except in so far as this overlaps with something else that we are describing. The aim is to describe the contributions they were *selected* for making; that is, the aim is to describe their proper function. (Again, *which* description of their function is employed *where* is something I will come to in a moment). The primary physiological analysis neces-

sarily abstracts away from maladaptive environments (like dried ponds that render webbed feet useless) and the infinite possible pathways of pathology (such as what John's liver did when it had cirrhosis) by describing what the components do when they are functioning properly. Descriptions of what happens when things are maladaptive or malfunctioning are given against this background. (I.e., medical descriptions of disease processes tend to assume normal functioning of the body, except where abnormalities are explicitly described or are inferable from those explicitly described.)

If you fail to see the intimate connection between functional analysis and the teleological notion of a function, you might well be bemused by so much fuss about teleology in biology. Having seen the connection, however, it should be evident that it would be hard to exaggerate its importance, for *only with its aid can we conceptualise the organisation of extremely complex and idiosyncratically diverse organisms with sufficient generality*. The generalization requires the idealization. If cognitive analysis requires a similar kind of functional analysis, then it too might require a teleological notion of a function. But this is getting too far ahead of the present argument.

To return to the antelope. The hierarchy of functions for the antelope's trait can be seen as one strand of a fuller functional analysis, except that the components are not identified, only their contributions are, which obscures an interesting fact. That is, as we move up the diagram, we are describing the functions of larger and larger systems. The trait alters the structure of the haemoglobin within individual cells, but the lungs and circulatory system need to become involved before there is an increased uptake of oxygen, and the whole of the female antelope (more or less) is involved in bringing a calf to term and replicating genes. In a similar fashion, aligning themselves with the magnetic field is something the magnetosomes do more or less on their own, moving the bacteria to anaerobic conditions is something they do only in collaboration with other components.⁷

It might be suggested that only the lowest level describes the function of the trait in question, the others describing the functions of the super-systems of which that trait is a part. But while the lowest level describes the function that is most specific to the trait, it is also consistent with

the definition of 'function' being employed here that it is the function of the trait to *help* achieve the higher functions (e.g., to help increase oxygen uptake and to help the antelope survive at higher ground). In any case, what turns out to be critical to the present debate is that there is a determinate order to the different descriptions.

Notice that the different function attributions for a given trait are *complementary* when it is a matter of saying what a trait is supposed to do, because *the trait is supposed to do all of these things*. Everything is simple as long as a trait either does all of what it is supposed to do, or for that matter, none of it. But suppose it performs its function as described at one level but not as described at another? This is the worry that bothered Dretske. As he tells us, if we shifted the bacteria to the southern hemisphere, the magnetosomes would still align themselves with the prevailing magnetic field but they wouldn't help direct the bacteria to anaerobic conditions. Instead, the bacteria would be directed to the oxygenated water where they cannot survive. This is where the plurality of function attributions threatens to cause problems, for it looks as though we might be obliged to say that a given trait at a given time is functioning properly and is not functioning properly.

How to settle this potential conflict in biological norms? Let's continue to consider the problem as it arises for biology by considering the following case. Suppose a woman's Fallopian tubes are blocked, and as a result she is unable to conceive: sperm cannot reach the ova to fertilise them. Now here are two ways of describing her condition: (1) we could say that her reproductive system is malfunctioning, and further that all of its parts are malfunctioning, because none of them can achieve their higher 'purpose' of bringing a child to term and replicating her genes, or (2), we could say that her reproductive system is malfunctioning, but not all of its parts are malfunctioning, only her Fallopian tubes. In other words, on this second option, her ovaries don't malfunction (just) because her reproductive system malfunctions because her Fallopian tubes are blocked.

The first option gives preference to the highest level of the functional analysis. Nothing is functioning properly unless it actually contributes to the proliferation of genes. The second option gives preference, in contrast, to that description of a trait's function that is most specific

to the trait in question. On the second option, if a trait can perform the function that is most specific to it, it functions properly. All the functions above that on the hierarchy are things that its proper functioning may or may not achieve.

We should choose the second option because it makes talk of malfunctioning more informative than the first option does. Think of what we learn if we are told that her ovaries are malfunctioning. On the first option, when we are told this we still don't know whether her ovaries are the cause of her infertility. All we are told is that some larger system of which her ovaries are a part (the reproductive system, or even the organism as a whole) is not working properly. Under the second option, in contrast, when we are told that her ovaries are malfunctioning, we learn something much more specific about the source of the problem. It is the ovaries themselves that are the problem. Not surprisingly, biologists do not maintain that the ovaries are malfunctioning because conception is impossible because the Fallopian tubes are blocked.⁸ Not surprisingly, they choose the second option.

In sum: there is more than one plausible way to describe a trait's function, but the different descriptions have a determinate order. They are connected by an asymmetrical by-relation, and generally have a place in a functional analysis. Moreover, biologists have a principled way of arbitrating potential conflicts in normative judgements about proper functioning: a part malfunctions when it cannot perform its most specific function. This suggestion will find its parallel when I turn more directly to the problem of determining content, but first, a few words on the philosophical problem.

B. *The Philosophical Problem*

Consider again the definition of functions favoured in this paper. It tells us that the function of a trait is what it did in ancestral organisms that caused it to be selected. The philosophical problem is that as long as we capture *de re* whatever the trait did that caused it to be selected, we seem to have a description of its function that is strictly warranted by the definition, and there are too many such descriptions. Most notably, Jerry Fodor has argued along these lines, claiming that selection processes

are sensitive to what is done, not to how that which is done is described: natural selection, he says, is extensional [1984].

With this problem in mind, look again at Dretske's example. Since (in the environment in which the bacteria evolved) the direction of the Earth's magnetic polarity is the same as the direction of de-oxygenated water, directing the bacteria to the one amounted to directing the bacteria to the other (in the environment in which the bacteria evolved). This is why, on Fodor's analysis, both descriptions of the trait's function are warranted by the definition.

Now I have argued that there is an alternative diagnosis of Dretske's example, and one which has nothing to do with the co-extensionality of the descriptions, which is incidental on that diagnosis. I have argued that both descriptions have a place in a functional analysis, and given this, they can be handled as I have recommended all such plausible cases be handled. We have a principled reason for giving priority to one level of description over another, so we have a way of avoiding potential conflicts in functional norms. This suffices for the purposes of biology, at least, and the hope is that something similar can be done for semantics. But, while this reply might seem a promising start for the *plausible* function attributions, it doesn't bear on the host of *implausible* cases which Fodor might equally have used to illustrate his problem.

People rarely notice that Fodor's principle – that natural selection is extensional – generates a host of function attributions that no biologist (qua biologist) would take seriously for a moment. To illustrate the disjunction problem we might as well have said that since there were no mermaids in the environment in which the bacteria evolved, the kind *de-oxygenated water* was (and is) co-extensive with the kind *de-oxygenated water or mermaids*, so that directing the bacteria to de-oxygenated water or mermaids was equivalent to directing the bacteria to de-oxygenated water; doing the one amounted to doing the other. This case, unlike Dretske's original example, has only one source, and it is the one which Fodor identifies as *the* source of functional indeterminacy – the description insensitivity of natural selection.

Picturesquely, Fodor has said, 'Darwin cares how many flies you eat, but not what description you eat them under,' [1990, p. 73]. In doing so he is referring to the infamous case of the frog and the fleabee, which

I discuss in a moment. But let's stick with the bacteria for a while. Presumably, Fodor is also prepared to say that, for all Darwin would or should care, the magnetosomes can be described as directing the bacteria to de-oxygenated water or mermaids. But frankly, *this* doesn't sound plausible, and Darwin would have thought it absurd.

Please notice that I am not denying that natural selection is extensional, in one specific sense. That is, if doing something in the presence of *As* enhances fitness, and *As* and *Bs* are co-extensive (at least locally), then doing it in the presence of *Bs* must equally enhance fitness (at least locally). This is perfectly true, but having acknowledged this, we can still deny that it's all Darwin would or should care about.

He would rightly be fascinated by all sorts of other things. For instance, he might wonder which feature of the *A/B* situation was beneficial to the organism, and why it was beneficial, and he might wonder how the organism detects that it is an *A/B* situation, and which feature of such situations the organism is responding to. And the answers to these questions are not description insensitive. Perhaps *A* was the benefit and *B* was the stimulus, or perhaps *B* was the benefit and *A* was the stimulus, or perhaps *A* was both benefit and stimulus, or perhaps *B* was. The point is that these are different answers and one can be true while the other is false, despite the fact that, by hypothesis, *A* and *B* are co-extensive. Determining the answer involves considering which properties are causally efficacious at different places in the functional analysis.

The simple claim that we need to isolate *the* property which is causally efficacious in the selection process ignores the fact that there is usually more than one such property (as above, where one is the benefit and the other is the stimulus). However, it seems sensible to require that any of the properties that are mentioned in descriptions of functions be amongst those that were causally efficacious. This eliminates descriptions of functions whose only claim to legitimacy is that they are co-extensional with some other description which is independently legitimate.

Can I help myself to this notion of 'causal efficacy?' Ruth Millikan also appeals to this notion via her use of Elliott Sober's distinction between selection for and selection of [Millikan, 1991]. (Traits that are

selected for are causally efficacious in the selection process.) And Fodor argues that Millikan cannot help herself to this notion. But his reason for denying it to her is peculiar to her, or rather, to someone who insists on maintaining what Fodor says Millikan insists on maintaining. (She doesn't, she says (in correspondence).) He attributes to her the claim that counterfactuals are irrelevant ("... as Millikan is always rightly reminding us, on her theory the determinants of content are historical and not counterfactual" [Fodor, 1991, p. 295]). And he runs a strong line against this position, insisting that the inverse is true. "So what makes the difference between being selected and being selected for is not *history* but *counterfactuals*," he says [p. 294]. But neither Fodor nor Fodor's Millikan is correct in this case. If history includes more than a sequence of events and a series of statistical correlations, if, that is, it includes causal relations between events, and if causal relations between events depend on counterfactuals, then it follows that history depends on counterfactuals. Millikan and I can allow counterfactuals to be relevant because (not despite the fact that) history is relevant.

Synchronic counterfactuals can determine what something *is* selected for, but not what something *was* selected for. Consider a variation on Sober's toy [Sober, 1984, Ch. 3]. Sober's toy is a clear cylinder, in which there are balls of different sizes and colors. The toy is divided horizontally by several sieves that become more fine-grained toward the bottom, and it is getting to the bottom that counts as getting selected in this case. Since all and only the smallest balls are green, it is only the small green balls that get to the bottom. Sober's point is that the small green balls are selected *for* being small, not for being green, because being small was the property that was causally efficacious in the selection process (there was merely selection *of* being green). But let's modify the toy. On one side, it is the same as before, but on the other side it has a color sensitive device that allows only green balls to pass through to the bottom. You place the balls in the top of the toy and shake it, as before, but now some balls go one way, and some the other. And when they are at the bottom, there is no way of knowing which way they went. Maybe you could watch very carefully and follow an individual ball's trajectory, but if you lose sight of it, you won't be able to figure it out afterwards. So some of the balls were selected for being green and

some were selected for being small, and there is nothing whatsoever in the synchronic counterfactuals that will tell us which is which.

So Fodor hasn't shown, by this argument at least, that "... counterfactuals do all the work and Darwin goes out the window" [Fodor, *ibid.*, p. 295]. If we want to know whether teleology is doing any useful work, we must ask ourselves, not whether causal and counterfactual claims are doing all the work, but rather whether the causal and counterfactual claims involved are those that underwrite teleology.

That's all I am going to say about the disjunction problem in this paper. Back now to the biological problem, for in what follows I restrict my concern to the plausible descriptions of a trait's function. I have argued that we have a principled reason for giving preference to one plausible description of a trait's function over another, a reason which serves us well in biology. The next question is whether there is a principle which will do the same job in semantics.

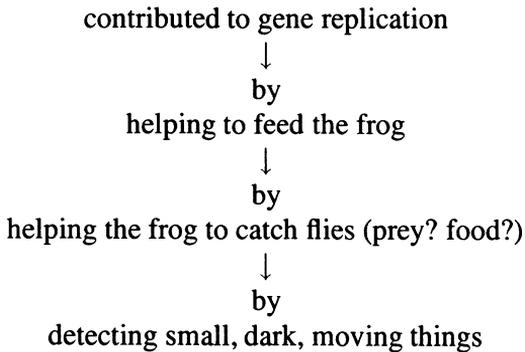
IV. TWO APPROACHES TO TELEOSEMANTICS

Remember that we are considering the claim that teleological theories of content cannot provide for the possibility of misrepresentation, the problem being that function attributions are apparently indeterminate, there being more than one acceptable description of a trait's function which leads to conflicts in associated normative judgements.

To see the problem as it relates to specifically semantic concerns, let's start with the frog (*Rana Pipiens*) which responds to anything appropriately small, dark and moving past its retina by snapping and gulping down the catch [Lettvin et al., 1959]. The proximal stimulus is a particular pattern on the retina which falls within specifiable parameters corresponding to such things as size, contrast with background and motion. The distal stimulus, in the environment in which the frog evolved, was usually (or often enough) a fly, but these frogs will respond indiscriminately to other small, dark, moving things which are not flies, such as bee bees (small pellets). The example is overly familiar, but as my treatment of it departs radically from the standard teleological treatment, it helps to have this common ground as the point of departure.

Let's call the frog's representation of its target 'R'. Here then is the kind of question I want to focus on. Can we use teleology to determine the extension of *R* in a way which allows for misrepresentation? If so, what is the best general strategy for doing so? And how do we choose between the various permissible descriptions of the function of the frog's prey-detector? More concretely, suppose that the frog sees a bee-bee whizzing by. Is the bee bee in the content/extension of *R*, or would *R* be a misrepresentation of the bee bee? And why?

In the remainder of this paper I am going to contrast two broadly teleological but different strategies to these questions. It helps to see these strategies in relation to a by-relation diagram which can be drawn up for the frog as follows. The frog's optic fibres:



The two possibilities I am going to consider are these: we could give priority to the top level of description, which is the standard teleosemantic strategy, or we could give priority to the lowest level of description, which is the strategy I am suggesting we should adopt. This choice doesn't cover all the options. There are some teleosemantic proposals which effectively employ several levels together [e.g., Agar, 1993] and proposals that favour a middling, more ecologically perspicuous, description [e.g., Sterelny, 1990]. While there are some interesting ideas here, in the end I think they don't work, and for reasons of space, I will save my critical comments for the main opposition.

A. *The High Church*

Ruth Millikan [1984, 1986, 1990(a&b)] says, in effect, that the relevant level of the functional analysis is the highest level of description. She identifies the content of a representation with the environmental feature which was historically responsible for the consumer of the representation performing its proper function at the highest level. For example, as Millikan tells the tale [1990(b), pp. 330–335], female hoverflies cruise at a fairly standard speed while male hoverflies hover until they detect something suitably sexy, whereupon they zoom off in hot pursuit and attempt to mate. Like the frog, the male hoverfly is indiscriminate in its choice of targets: large distant objects – birds and even jet-planes – can trigger his amorous instincts. In discussing this example, Millikan claims that the male hoverfly abides by 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.’⁹

Even when other teleosemantic theorists offer arguments ostensibly defending other levels of the analysis, it is often an argument that would better be given for the top level of description. The top level of description is, in that sense, the most robust of the alternative proposals. For instance, Kim Sterelny has urged that the content of *R* is *fly* rather than *small, dark, moving thing* because, he says, the frog is only interested in small, dark, moving things because they were often enough flies and flies were what fed it. However, as Paul Griffiths and Richard Goode have pointed out [1995], the same line of argument can be given against *flies* in favour of *frog food*, for the frog is only interested in flies because they were nutritious, and this is what contributed to the proliferation of the frog’s genes. The argument doesn’t run out of steam until it reaches the top.

High Church Teleology is often thought to provide intuitively appealing attributions of content, but it doesn’t do so if rigorously applied. This is a problem others have noticed before [Hall, 1990], and it is more obvious in the case of the hoverfly. Ask yourself, which environmental feature did ancestral hoverflies actually track on each and only those occasions when the hoverfly made an actual deposit to the gene pool of future generations? Clearly, it had to be a female hoverfly, certainly not a

bird or a jet-plane. But why stop there? It was only the female hoverflies which were caught and fertile and survived long enough to lay eggs, which in turn hatched into viable offspring, which were also fertile, and so on, that were actually causally efficacious. In other words, taken as is, the standard teleological approach implies that our lusty hoverfly misrepresents if he chases an infertile female or one who is soon to be the dinner of some passing bat.

It is no use suggesting that the represented kind should be a natural kind. Even if we are liberal about what counts as a natural kind, it remains unclear what account of natural kinds could possibly shave the unwanted detail without shaving too much else as well. Birds have an innate preference for ripe fruit, and sometimes for mates with vivid red tails, so presumably they have representations of colors, and so color kinds had better count as natural kinds, or we will have shaved too much. But in virtue of *what* would they count as natural kinds? If featuring in an evolutionary explanation is sufficient qualification for being a natural kind, then they would count, but then we would have returned to where we had started and would have shaved nothing.

Another thought is that the content should not include details to which the creature is perceptually insensitive. The male hoverfly cannot possibly be sensitive to whether his prospective mate is about to be gobbled up, for instance. And nor is he sensitive to whether she carries nasty diseases or is infertile. But this reply does too much, for the hoverfly isn't sensitive to whether she is a conspecific either. That's why he stupidly chases jet-planes.

Paul Pietroski [1992] has argued that teleosemantic theories – but in fact it is just the High Church variety – give unintuitive content in a further way. As he indicates, a creature might fall asleep when it sees night-fall coming on and thus be protected from predators. On Millikan's theory, the creature *sees night-fall as a predator!* For the perceptual representation involved means predator, not night-fall, because it was the avoidance of predators that was responsible for the consumer of the representation performing its proper function, and ultimately contributing to the replication of the creature's genes.

Why then should we go for High Church Teleosemantics? The arguments are scarce at this point. Sometimes Sober's distinction between

selection *for* and selection *of* is appealed to again [e.g., Sterelny, 1990]. The claim is that there is selection for the detection of frog food (or some such thing) whereas there is only selection *of* the detection of small, dark moving things. This would be relevant, given that we are attempting to identify content by appeal to what a detection device was selected for, were it not that it misapplies Sober's distinction, as Goode and Griffiths again have argued. The distinction is supposed to be between properties that were causally efficacious in a selection process (those that were selected for) and those that were not (the free riders) [Sober 1984, Ch. 3] but several properties can be causally efficacious in the same selection process: in the case of the frog, one as the benefit (being frog-food) and another as the stimulus (being small dark and moving). If so, Sober's distinction doesn't cut between them. (Those who recall Sober's toy, will notice the difference: in that case the balls were selected for being small, and the greenness of the balls is entirely inert.)¹⁰

It is a mistake to suppose that the higher levels of a functional analysis are 'more teleological' than the lower, more 'mechanical' levels. All levels of description are descriptions of what a given trait did that caused it to be selected, and all levels of description are, in that sense, normative. That's why their plurality presents us with the *prima-facie* problem of conflicting normative judgements. It is also a mistake, if my earlier arguments are on the right track, to suppose that giving priority to the highest level of description is more in keeping with the biological paradigm. Recall that I argued that, with respect to a given part (P) of an organism, biologists should (and do) give priority to that level of description which is most specific to P – the lowest level of description in the functional analysis before we go sub-P.

Recently, Sterelny has argued that the main point of content attributions is to allow us to make generalisations about representations that abstract away from differences of mechanism. We need this in biology, he says, to make cross-species generalisations and classifications – of and about prey and predator detectors, for example. This, he believes, shows that the content of the frog's *R* should be an ecologically perspicuous kind, like *prey* or *fly*. But this doesn't follow. Compare the needs of the biologist with respect to digestive systems. The biologist also needs to make cross species generalisations and classifications here. But they

can do this without supposing that the level of description employed in doing so also determines our fine-grained judgements about biological norms. Both sheep and humans have mammalian digestive systems, about which certain generalisations can be made, but what is normal functioning for the one need not be normal functioning for the other. We don't malfunction just because we can't digest grass. Similarly, the frog's detection device can be categorised as a prey-detector without committing us to any fine-grained judgements about biological norms, or semantic norms either. We can still ask what it represents its prey *as*.

What the High Church needs is an argument to the effect that psychology needs it. But I don't think psychology does. I'd like to make that claim on behalf of my alternative suggestion, but I don't pretend to establish it in the few pages that follow. Instead, large and complex issues are raised that cannot be properly treated here. However, I do want to sketch some reasons for thinking that, contrary to the common presumption, the Low Church offers a better way.

B. *The Low Church*

I suggest that the frog doesn't misrepresent as long as its representation *R* is caused by something appropriately small, dark and moving^{11,12}. This preference for the lower level of description parallels biological practice in the non-semantic case, but that's not my only motivation for making the suggestion, as I will explain.

But first, a few words of clarification. By 'the lower level of description,' I do *not* mean to refer to the *lowest* level of description in the functional analysis (in this case that's somewhere down amongst the sub-cellular neuronal components). What I mean to refer to is *the lowest level at which the trait in question is an unanalysed component of the functional analysis*. So what counts as 'lowest' is relative to the trait in question.

Recall that in functional analysis we recursively decompose a system (*S*) into its parts. It is decomposed in stages: in stage one, *S* is decomposed into the parts *S*_a, *S*_b ... *S*_n; in stage two *S*_a is decomposed into *S*_a¹, *S*_a² ... *S*_aⁿ (and *S*_b into *S*_b¹, *S*_b² ... *S*_bⁿ, etc.); in stage three *S*_a¹ is decomposed into *S*_a^{1a}, *S*_a^{1b} ... *S*_a¹ⁿ, and so on. *S*_a¹ is an unanalysed com-

ponent at stage two, for at that stage it is used to analyse the system of which it is a part, but it has not been analysed into the parts that compose it. (Just as the ovaries don't malfunction just because the reproductive system malfunctions, nor do they necessarily malfunction just because one of their sub-cellular mitochondrial organelle malfunctions.)

We have been vague about just where *R* occurs in the frog's perceptual processing. Sometimes the representation is referred to as 'the triggering of the frog's detection device': meaning, I suppose, whatever state the frog's optic fibers and associated neural processors are in when they cause tongue snapping.¹³ Given that this is the representation we are interested in, my proposal is that the level of functional analysis that should be given priority is the lowest level at which the detection device as a whole is an unanalysed component. Now, the whole detection device can be described as detecting frog-food, detecting flies, or detecting small, dark moving things, but the last of these descriptions is 'lower' than the others, for it is *by* detecting small dark moving things that the frog detects frog-food and flies.

There are various questions that arise at this point. You might wonder: (i) whether this suggestion allows for the possibility of misrepresentation; (ii) whether it allows for it adequately; (iii) whether there is any motivation for thinking it is the correct level for semantics and psychology; and (iv) whether it can resist sliding toward proximal content. The first worry can be answered decisively, but for the rest I can do little more here than trace some connections to wider issues. I hope this will be enough, however, to keep the suggestion I am making on the table. I'll discuss each of them in the order mentioned.

(i) *Any Misrepresentation?* I claim that the frog doesn't misrepresent if it *R*-tokens at something small dark and moving, and this implies that it doesn't misrepresent when it *R*-tokens at bee bees. But, you might ask, wasn't the aim of the exercise to justify the claim that the frog misrepresents when it *R*-tokens at bee bees? The worry is that in failing to allow for this, I may have failed to allow for misrepresentation altogether. If so, I have merely captured something like natural meaning (or indication) instead. (As Dretske, following Grice, defines the notion of natural meaning, '... nothing can mean that *P* in the *natural* sense of meaning if *P* is not the case ...' [1988, p. 155].)

Curiously, it has been an assumption of this debate that the frog is misrepresenting the bee bee, and given this assumption, the goal of the game has been finding a principled justification for *this* answer: thus there are various attempts to show that *R* means *frog food* or *fly* or some such thing. The assumption is based on the perceived need for a causal theory of reference to provide a distinction between the causes of *R* that are properly in the content of *R* and the causes of *R* that are not. The thought is that, in the absence of such a distinction, there can be no misrepresentation, for, in the absence of such a distinction, anything that causes *R* must be in the content of *R*. This is a real enough need, but it doesn't follow that unless the frog misrepresents when it *R*-tokens at bee bees it cannot misrepresent.

Let me explain. On the account I favour, if the frog *R*-tokens at anything which reflects onto its retina a pattern that falls outside of the specified parameters, then it misrepresents. The images cast by snails, for example, will fall outside these parameters, even when the snails are at their smallest and most sprightly, so *R*-tokening in response to a snail is a misrepresentation. So the frog *can* still misrepresent.

But, you might say, the frog can't actually *make* these mistakes. Well, that raises an interesting question. In what sense must it be possible that these mistakes can be made? For a representation (such as *R*) to have content, it is enough that there be something which, were it the case that *R* was tokened in response to it, *R* would be misrepresenting. Since there are plenty of things that qualify (snails, galoshes, lily pads) this condition is satisfied on my story.

Anyway, although there is a kind of mistake that *we* can make that *frogs* can't (more on this in a moment), frogs can make the mistakes mentioned above: some of them, anyway. It's true that they can't/don't when they are functioning properly, but they don't always function properly. A sick frog might *R*-token at a snail if it was dysfunctional in the right way. Damaging the frog's neurology, interfering in its embryological development, tinkering with its genes, giving it a virus, all of these could introduce malfunction and error. Therefore, the theory I am defending does not reduce content to the non-normative notion of indication or natural meaning. The frog can token *R*, and *R* can mean *Y*, even though *Y* is not the case.¹⁴

(ii) *Enough Misrepresentation?* So far, so good, but now a second wave of worries along similar lines needs to be addressed. Having been persuaded that misrepresentation (i.e., something, in some sense, meaning P when P is not the case) is possible, you might still be bothered by the fact that the account I'm defending doesn't allow for *adequate* misrepresentation. And you might think it is inapplicable to human cognition for two reasons: (a) we are capable of misrepresenting without malfunctioning, and (b) the contents of our representations go beyond the physical parameters of the environmental features that our perceptual systems measure. You might argue, therefore, that this account, whatever its merits for frogs, must be irrelevant to human cognition, which should, after all, be our paradigm case of intentionality.

Now, naturally, I agree with statements (a) and (b), but (a) and (b) don't show that the account I am defending is inapplicable to humans, since it doesn't follow that all human representation is of this kind. We need to consider whether we have these capacities at the level of pre-conscious and pre-linguistic perceptual processing, or whether they emerge later in the cognitive story. It seems plausible that they emerge later.

Consider the case where we see a skinny cow in the dim distance and mistakenly represent it as a horse (Fodor's example). Here, we may suppose, we misrepresent without malfunctioning, and clearly the content of our perceptual representation goes beyond the physical parameters of the environmental features measured. But this sophisticated representation occurs after much visual processing has already taken place, at least, this is so on computational theories of vision. In such theories, early visual processing does not represent the cow as a horse (or as a cow) but as something which *looks a certain way* – as having a certain outline, texture, color, and so on.¹⁵ That is, according to conventional computational theories of perception, *initially there is a representation of the physical parameters of the environment as measured by the visual system*. It is much more plausible that there is no misrepresentation without malfunction at this level.

For ask yourself this: if there is no malfunctioning, where does the error occur? There could be many different explanations in this particular case. Perhaps the cow really looked more like a horse from that distance and that perspective. Or perhaps it looked more like a cow

than a horse, but we thought Farmer Brown owned only horses, and our expectations overruled the less than compelling perceptual data. In other words, we are capable of misrepresenting without malfunctioning precisely because our perception involves inference, and inference is a fine way to make a mistake in one's conclusion without making a mistake in one's method.

As Fodor says,

... what must go on in perception is that a description of the environment that is *not* couched in a vocabulary whose terms designate values of physical variables is somehow computed on the basis of a description that *is* couched in such a vocabulary. Presumably this is possible because the perceptual analysis of an event is determined not just by sensory information but also by such background knowledge as the organism brings to the task. [1975, pp. 45–49]

If so, it follows that some of our representations must be 'couched in a vocabulary whose terms designate values of physical variables: those involved in early visual processing, for example. And since the frog's perception of its target is notably immune to correction in the light of further experience, it is reasonable to suppose it never gets beyond this.

While it seems fair to say that an inferential (often called a constructivist) theory of vision is most popular amongst psychologists, the claim that perception is inferential is by no means uncontroversial, and someone with Gibsonian sympathies (such as Millikan) would reject this account of the matter. According to Gibsonian psychology, the first representation that makes its appearance in visual processing – the first state that is semantically interpretable – is one with an ecologically significant content, and no inference, because no representation, takes place before we represent what is seen as one of an ecologically significant kind (e.g., food or cow). In some sense the visual data is supposed to be unambiguous, not merely constraining the perceptual interpretation, but fully determining it. If this were so, my account of the difference between us and frogs could not be sustained. However, I line up with Fodor on this issue, it is implausible that perceptual recognition is non-inferential in this way [see, Fodor and Pylyshyn; and Dretske, 1990].

(iii) *The Needs of Psychology?* I am not particularly interested, here at any rate, in deciding where psychology begins, and whether there is anything that deserves to be called 'frog psychology.' Perhaps psy-

chology proper begins with creatures that have propositional attitudes, and perhaps frogs have contentful states but no propositional attitudes. I want to leave these issues open. But if frogs do have contentful states, our account of their content should be one that flows through to creatures that have a psychology.¹⁶ And in attributing contentful states to frogs, we should keep an eye keenly directed ahead, on nothing less than the needs of human psychology. From that point of view, here are a few reasons for preferring Low to High Teleology.

The first reason follows on directly from the foregoing discussion of inferential perceptual processing. If the capacity for misrepresenting without malfunctioning depends on perceptual recognition being an inferential process (specifically, an abductive inference in the light of background knowledge), and if it's implausible that any such inferential process occurs in the frog, as seems to be the case, then High Teleology bestows on the frog semantic capacities it doesn't have. A theory's allowing for a certain capacity in frogs is only a virtue *if* frogs really have that capacity, otherwise it is a vice, for otherwise it obscures the basis of the capacity in us, and obscures fundamental differences in complexity and sophistication between us and frogs.

Here is a second reason that will trouble some. Unlike Millikan and Sterelny, and with Colin McGinn [1982], I believe that a two-factor theory of content is needed: that is, we need something like sense as well as something like reference. Moreover, Frege's claim that sense determines reference, in the revised form given it by Kaplan [1978] – that is, that sense determines reference in a context – also seems correct. There are Quinean difficulties with this view, but all things considered, it looks as though it could and should be salvaged somehow, and there are a number of promising attempts to do so (see, for example [Peacocke, 1992; Rey, 1993]).

Now, if a notion of sense (as well as reference) is needed by psychology, that need seems to favour Low Teleology, and not High Teleology, since Low Teleology has something to say about the criteria by which a cognitive system identifies that which it represents, and how it represents it, whereas High Teleology has nothing to say about this. Following this insight, Nenad Miscevic (in an unpublished manuscript) has suggested that senses follow (what I call) Low Teleology while reference

follows (what I call) High Teleology. But on his account, sense does not determine reference, not even partly: it is context, albeit an historical context, that does *all* of that work. While this will not bother Millikan and Sterelny, it will rightly bother many others.

A further (speculative but plausible) reason, is this. As Cummins [1983] argued, psychology should aim to provide an interpreted cognitive analysis. This analysis decomposes a cognitive system into parts that are identified in terms of the cognitive tasks they perform (so something might be identified as an *adder*, for example). This analysis is carried out recursively, just like the (uninterpreted physiological) functional analysis described earlier in this paper. We begin with the main cognitive task to be analysed and (a la Dennett) decompose it into the simpler tasks, invoking more and more stupid homunculi, that together achieve it. We continue decomposing the task into simpler and simpler tasks until we bottom out at tasks so basic that they can be done by dumb mechanism.

The brain, then, will have an uninterpreted physiological analysis and an interpreted cognitive analysis, if this story is correct. How the two are to be related is of course a difficult and complicated issue. However, it seems desirable that they be isomorphic, where possible, for that will give us easy access to information about which physiological processes are performing which cognitive processes. And given that the principles of physiological analysis are well established and successful, cognitive analysis should follow its principles where possible, unless a good reason to do otherwise becomes evident.

Furthermore, it is unclear how discrete these analyses would ultimately be, for as earlier described, a functional analysis uses *functional categories* to decompose a system into its component parts. If some of the information-processing functions of the brain are specifiable in intentional terms, then we may not have two discrete analyses (an interpreted and an uninterpreted analysis) of the same system, we might have one, or two intimately interlocking ones.

(iv) *Distality?* This paper is already long, but a comment on distality seems to be needed before finally closing. I have up until now been assuming that the content of the frog's representation is *distal* (i.e., that it is something that is small dark and moving, and *out there in the*

frog's environment). But I haven't provided a principled reason for this. Worse, I have provided a principle for determining content that would, as it were, seem to fly in the face of this. For if the principle being applied tells us to give priority to the lowest level of description in the functional analysis, then, arbitrary prohibitions aside, it seems to drive us to proximal content. For any description I give that speaks of distal objects (e.g., small dark moving things) can be trumped by one that speaks only of a proximal object (i.e., a retinal pattern of a particular kind). It is, after all, *by* responding to a retinal pattern of a particular kind that the frog responds to small dark moving things.

Now, it's true that I haven't provided a principled answer to the distality problem, but I haven't precluded one either, and, in fairness, it was not the problem being tackled. It's also true that the logic of my suggestion seems to imply that we should slide all the way down to the proximal, but it doesn't compel us to do so. All we need is a further principle, compatible with the suggestions made in this paper, which is specifically tailored for handling the distality problem.

In other words, determining content might be a two (or more) step process. Added to the position defended in this paper could be something like Dretske's principle of triangulation [Dretske, 1986]. Or perhaps, on teleological grounds, we could determine the distal range of whole classes of representations on a modality wide basis, for they do seem to be thus characterisable: the role of the visual system is to provide information about objects which reflect light onto the retina of the eye, the role of the auditory system is to provide information about sources of sound, the role of the kinaesthetic system is to provide information about bodily movements, and so on. This is not the place to defend any particular solution to the distality problem, and I don't particularly want to defend either of these suggestions. My point here is really a logical or formal one: I haven't precluded such a solution. What will do the trick is, of course, a very hard question to answer (and a matter for another paper).

V. CLOSING REMARKS

With respect, and improbable as it might seem, I like to think of this paper as the philosophical marriage of Fodor and Millikan. This is teleosemantics for those who love their Language Of Thought and Computational Theory of Mind.

To sum up. I have separated the (philosophical) ‘disjunction problem’ from another source of functional indeterminacy, one that is responsible for the (biologically) plausible variations on descriptions of functions. The plausible cases, I have argued, are not connected by a symmetrical ‘by-relation’ that depends on co-extensionality, but by an asymmetrical ‘by-relation’ that depends on functional analysis. These plausible descriptions have a determinate order, and when there is a potential conflict in biological norms, we should give priority to that description of a trait’s function that is the lowest level in the analysis (most mechanistic) at which the trait in question remains an unanalysed component. Standard teleosemantic theories give priority to the highest level of analysis (contribution to gene proliferation) but the content derived from these theories looks psychologically implausible on closer examination. A more mechanistic teleology is more sensitive to the likely demands of a detailed cognitive analysis.

NOTES

* This paper has gone through a number of drafts, and an early version was first presented at the Australasian Philosophy Conference, Melbourne, July 1991. It has benefited from discussions with many people, and I am especially grateful to my colleagues at the ANU, RSSH, and to the staff and students at Rutgers University. I am much indebted to Georges Rey, for his invaluable comments on several drafts, and to Ruth Millikan for some very enjoyable and useful discussions during The Blizzard of '93.

¹ One hope is that we can give a biological account of innate primitives, out of which more sophisticated representations are constructed. There are notorious difficulties with such an approach, but we shouldn’t despair of it, in my view. If only because the alternatives are worse [but see Fodor, 1981]. If the constructivist is to avoid certain problems (like Verificationism) non-perceptual primitives will also be needed, but this is a task for another day.

² See Neander [1991 (a)&(b)]. I first defended this view in my Ph.D. [La Trobe, 1983] having been greatly influenced by Wright [1973, 1976]. My theory is similar to Millikan’s [1989] and (less obviously) her [1984, Chs 1 & 2].

³ To be fair, it takes a lot more argument to show that function attributions are not determined *in part* by our values. Robert Cummins, for instance, could be taken as arguing for a theory that is part descriptive and part evaluative in his 1975. According to this theory, a function is a causal contribution to a complexly achieved overall activity of the system, and the relevant activities are *the ones we are interested in*. Boorse has a similar interest-relative clause in his theory. I explain why neither of these theories is satisfactory in the papers mentioned in n. 2.

⁴ Suggestion: extend the relevant population back or forward in time and then do the statistics. This idea doesn't work for various reasons. Briefly, the problems are (1) that function attributions will depend on irrelevant demographic fluctuations, (2) we can get different attributions for the same trait of the same individual by moving the beginning of the time period around, and (3) there is no non-circular way to specify a suitable length of time.

⁵ This is reminiscent of William Lycan's, "One and the same space-time slice may be occupied by . . . an unlocker of doors, an allower of entry into rooms, a facilitator of adulterous liaisons . . ." [1990, p. 82.].

⁶ I also make this claim in my 1991 (a) & (b).

⁷ Dretske describes a 'by' relation in his 1988. He describes John, who stains the carpet by spilling his wine by toppling his glass by moving his arm. As Dretske explains it, we have here a set of embedded processes, each defined by the event with which it ends. These events are causally related (the movement of John's arm causes his glass to topple which causes his wine to spill which causes the carpet to stain). The processes themselves, however, are not causally related, Dretske argues, since they have a part/whole relation to each other. (E.g., John's moving his arm is part of his spilling his wine.) This 'by' relation, therefore, is not a causal relation but a compositional one.

Dretske is interested in exposing the structure of behavior at this point. Although he says that digestion lends itself to this kind of analysis, he misses its relevance to the problem of indeterminism. This is not so surprising given the original form of his question: does the magnetosomes have the function of indicating the earth's magnetic polarity or of indicating anaerobic conditions? So described, we are not offered a choice between two processes, one embedded in the other, and each characterised by reference to a distinct culminating event. Rather we have just one process, and one culminating event, for the indication of anaerobic conditions and the indication of the earth's magnetic polarity are the same event, differently described. I have altered the question so that the options can be seen as embedded processes.

It is illuminating to read Dretske's analysis of the structure of behavior with functional analysis and functional indeterminism in mind. However, some of Dretske's claims about the structure of behavior clearly do not carry over to functional analysis. For instance, according to Dretske, any effect of a basic bodily movement (no matter how remote) can be used to describe a behavior (*qua* behavior) involving that movement. This is not true of biological functioning (*qua* biological functioning). Bovine digestion

leads to manure which leads to good compost and glorious roses, but producing glorious roses is not a function of a cow's intestines. (Of course Dretske doesn't suggest it is.)

⁸ I use 'information' here in the objective sense in which one statement is more informative than another if it excludes more possibilities.

⁹ If we give priority to the highest level of description, it's unclear what difference it makes if we attend to the proper functions of the producer or of the consumer of the representation. Both consumer and producer have to pull together if either is to succeed in achieving their highest function, and the conditions for success are the same in either case. Both the hoverfly's production of its visual representation (its seeing something) and its consumption of its visual representation (its zooming off and attempting to mate) contributed to gene proliferation when a female hoverfly was sighted, not a bird or a jet plane. Both the frog's production of *R* and its consumption of *R* (its snapping, swallowing and digesting, I suppose) contributed to gene proliferation when frog food was sighted, not a bee bee or a wind swept grain of sand. Perhaps the two can be pulled apart in some cases, depending on how the notion of the 'consumer of the representation' is made precise, but this is not a matter which we need to pursue for the present discussion.

¹⁰ Griffiths and Goode's paper [1995] was partly a response to an earlier version of this paper. They argue that Sober's distinction cannot be used to choose between complimentary descriptions of a selection process (at different levels), but only between competing descriptions (at the same level). They also have an interesting discussion of the possible sources of the different levels which differs substantially from my own.

¹¹ More precisely, according to Lettvin et al. [1959, p. 1948] the frog's optic fibres tell the frog,

... whether or not the object has a curved boundary, if it is darker than the background and moving on it; it remembers the object when it has stopped, providing the boundary lies totally within that area and is sharp; it shows most activity if the enclosed object moves intermittently with respect to a background. The memory of the object is abolished if a shadow obscures the object for a moment. [Lettvin et al., 1959, p. 1948]

¹² As far as I know, no one else has previously defended this view. Dretske is hard to classify, however. There are two strands in his account which pull in different directions [see his 1988, esp. pp. 68–70 and Ch. 4]. For *R* to represent *F*, Dretske requires two things: one is that *R* was recruited as a cause of movement *M* because it (*R*) indicated *F*, and the other is that *R* was recruited as a cause of movement *M* because *M* was generally rewarding only when *P* obtained. The frog's *R*-tokening indicates small, dark moving things far more reliably than it indicates flies or food, but snapping at flies or food is more reliably rewarding than snapping at small, dark moving things. Which way we determine content depends on which requirement we tighten or loosen, and Dretske doesn't say.

¹³ This may be too coarse grained an approach. If the visual processing involved is computational, there are multiple representations involved, and the state which represents the conclusion will not be the first state with content that we encounter on the way in from the frog's perceptual periphery.

¹⁴ A nice question raised by Karen Green (in discussion) is this. If malfunctioning occurs (say, by congenital defect) what justifies saying we have a token of the *same* type of representation? If it isn't the same type, the frog's snapping at a snail isn't *R*-tokening (as opposed to *R**-tokening). But if representations are functional categories, as I am supposing, a new type of representation (with a new function, and a new content) can only occur *after* selection for that type. Just as a congenitally deformed kidney is a kidney and is supposed to do what normal kidneys do, so too a congenitally defective *R*-tokener belongs in the same functional category as normal *R*-tokeners.

¹⁵ I am grateful to Colin McGinn for helping me to bring this point into focus.

¹⁶ Of course, it is possible that there is no unified account of content. But until we have good reason to abandon the project, our working hypothesis should be that a unified account of content can be given.

REFERENCES

- Agar, N. (1993) 'What Do Frogs Really Believe?,' *Australasian Journal of Philosophy* 71: 162–185.
- Beckner, M. (1959) *The Biological Way of Thought* (New York: Columbia University Press).
- Borse, C. (1975) 'On the Distinction between Disease and Illness,' *Philosophy and Public Affairs* 5(1): 49–68.
- Cummins, R. (1975) 'Functional Analysis,' *Journal of Philosophy* 72: 741–765.
- Cummins, R. (1983) *The Nature of Psychological Explanation* (Cambridge, Mass: MIT Press).
- Dretske, F. (1986) 'Misrepresentation,' in R. Bogdan (ed.), *Belief* (Oxford University Press: Oxford, UK). Reprinted in Lycan, *Mind and Cognition* (1990) listed below.
- Dretske, F. (1988) *Explaining Behavior* (Cambridge, Mass: MIT Press).
- Dretske, F. (1990) 'Seeing, Believing, and Knowing,' in Dan Osherman (ed.), *Visual Cognition and Action* (Cambridge, MA: MIT Press).
- Fodor, J. (1975) *The Language of Thought* (New York: Crowell).
- Fodor, J. (1981) 'The Current Status of the Innateness Controversy,' in Fodor (ed.), *Representations* (Cambridge, Mass: MIT Press).
- Fodor, J. (1984) 'Semantics Wisconsin Style,' *Synthese* 59: 231–250. Reprinted in Fodor (ed.), *A Theory of Content and Other Essays* (Cambridge, Mass: MIT Press, 1990).
- Fodor, J. and Pylyshyn, Z. (1981) 'How Direct is Visual Perception? Some Reflections on Gibson's "Ecological Approach",' *Cognition* 9: 139–196.
- Fodor, J. (1991) 'Replies' in Barry Loewer and Georges Rey (eds.), *Meaning in Mind, Fodor and His Critics* (op. cit.) (Oxford, UK: Basil Blackwell).
- Griffiths, P. and Goode, R. (1995), 'The Misuse of Sober's Selection for/Selection of Distinction,' *Biology and Philosophy* 10(1): 99–108.
- Hall, R. (1990) 'Does Representational Content Arise from Biological Function?,' *Philosophy of Science Association* 1: 193–199.

- Kaplan, D. (1978) 'Dthat,' in P. Cole (ed.), *Syntax and Semantics 9: Pragmatics* (New York, Academic Press).
- Lettvin, J., Maturana H., McCulloch W. and Pitts, W. (1959) 'What the Frog's Eye Tells the Frog's Brain,' *Proceedings of the Institute of Radio Engineers: 1940-51*.
- Lycan, W. (1990) 'Levels of Nature,' *Mind and Cognition: A Reader* (Cambridge, MA: Basil Blackwell).
- McGinn, C. (1982) 'The Structure of Content,' in A. Woodfield (ed.), *Thought and Object. Essays on Intentionality* (Oxford: Oxford University Press).
- Millikan, R. (1984) *Language, Thought, and Other Biological Categories* (Cambridge, Mass: MIT Press).
- Millikan, R. (1986) 'Thoughts Without Laws: Cognitive Science with Content,' *The Philosophical Review* XCV: 47-80.
- Millikan, R. (1989) 'In Defence of Proper Functions,' *Philosophy of Science* 56, 288-303.
- Millikan, R. (1990(a)) 'Biosemantics,' *The Journal of Philosophy* LXXXVI: 281-297.
- Millikan, R. (1990(b)) 'Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox,' *The Philosophical Review*, pp. 323-354.
- Millikan, R. (1991) 'Speaking up for Darwin,' in Barry Loewer and Georges Rey (eds.), *Meaning in Mind, Fodor and his Critics* (Oxford, UK: Basil Blackwell).
- Neander, K. (1983) *Abnormal Psychobiology*, Ph.D. thesis, La Trobe University.
- Neander, K. (1991(a)) 'Functions as Selected Effects: The Conceptual Analysts Defence,' *Philosophy of Science* 58: 168-184.
- Neander, K. (1991(b)) 'The Teleological Notion of a Function,' *Australasian Journal of Philosophy* 69: 454-468.
- Papineau, D. (1987) *Reality and Representation* (Oxford: Basil Blackwell).
- Peacocke, C. (1992) *A Study of Concepts* (Cambridge, Mass: Bradford.)
- Pietroski, P. (1992) 'Intentionality and Teleological Error,' *Pacific Philosophical Quarterly* 73(3): 267-282.
- Rey, G. (1993), 'The Unavailability of What We Mean: A Reply to Quine, Fodor and LePore,' *Grazer Philosophische Studien* 46: 61-101.
- Sober, E. (1984) *The Nature of Selection* (Cambridge, MA: Bradford, MIT).
- Sterelny, R. (1990) *The Representational Theory of Mind* (Oxford: Basil Blackwell).
- Wright, L. (1973) 'Functions,' *The Philosophical Review* 82: 139-168.
- Wright, L. (1976) *Teleological Explanations* (Berkeley: University of California Press).

Philosophy Program
Research School of Social Sciences
Australian National University
Canberra 0200
Australia