TELEOLOGICAL FUNCTIONALISM: NORMATIVITY, EXPLANATION, AND THE PHILOSOPHY OF MIND

by

JILLIAN SCOTT McINTOSH

B.A. (First Class Honours), Simon Fraser University, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Philosophy

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1997

AUTHORIZATION

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy
The University of British Columbia
Vancouver, B.C., Canada

ABSTRACT

The purpose of this dissertation is to advance our understanding of the intentionality and causal efficacy of mental states. More specifically, the dissertation is intended to help justify an appeal to teleological functions in the philosophy of mind.

I start by examining the disjunction problem as encountered by causal/ information-theoretic accounts of intentionality. Such accounts individuate the content of mental states on the basis of their cause or the information they carry. As a result, they require a principled method of ruling out those cases in which a state is tokened in the "wrong" circumstances. Without such a method, a state's content could be massively disjunctive and error would be impossible. The dissertation then considers one type of purported solution, *viz.*, teleological functionalism. The basic idea is that an analogy between malfunction and misrepresentation will help solve the disjunction problem by invoking a suitably naturalised notion of normativity. A state's content need not be what caused it but, rather, what *should* have caused it.

I argue that there are two legitimate ways of understanding teleological function in this context. Selectionist theories— the current favourites— attribute functions on the basis of selection history; a thing's function is that effect or behaviour for which it has been selected. In contrast, systems-theoretic accounts attribute function on the basis of an analysis of components with regard to the workings of a whole; a thing's function is that effect or behaviour which contributes to the performance of the whole, of which that thing is a part. Upon examination, it becomes apparent that neither notion of function meets all the desiderata one might reasonably expect need to be met. This is explicable— the different notions are suited to two different, though related, explanatory projects.

I argue that selectionist construals of teleological function are appropriate when, roughly, the project is that of explaining why extant features are present in the distribution and form that they are. In contrast, systems–theoretic construals of teleological function are appropriate when, roughly, the project is that of explaining how these features work. Furthermore, I argue that, from the perspective of a causal/ information-theoretic account, the normativity that is required for the project of individuating the content of mental states cannot derive solely from history. Knowing what served one's ancestors is not sufficient for knowing what one is doing now, let alone what one *should be* doing now.

A systems–theoretic (and more specifically, a structural) teleological functional approach to the problem of intentionality, because it is importantly ahistorical, has the merit of incorporating normative considerations into the philosophy of mind without rendering the causal efficacy of intentional states unnecessarily mysterious. It also has the merit of allowing for those attributions of teleological function in biology that would not be overturned by new evolutionary information regarding selection history. Adherence solely to an etiological construal of teleological function is too restrictive in both domains. The dissertation ends with a defence of the structural approach against the charge that it is too liberal in attributing functions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am very grateful to the members of my supervisory committee—Andrew Irvine, Phil Hanson, Martin Hahn, and Tom Patton—for all their help. Special thanks to my supervisor, Andrew Irvine, for his empathetic encouragement and direction; I doubt I would have finished without him. Special thanks, too, to Phil Hanson, for all his years of inspirational philosophical guidance.

Thanks to my family and friends. Everyone has been very supportive. Thanks to my brothers, Graham, Paul, and Steve McIntosh. Thanks to my mother, Sylvia Verity. Thanks to my dear friend, Rob Fenger. Thanks to Anne Harland, Lisa Smedman, Ruth Michaels, Rena Humphries, my friends at the Post Office, my Aunt Pat, Julie Colterjohn, Brent Smart, Diane Whiteley, my vet, Melissa Ochotsky, my friends at SFU, my friends at UBC, and my friends at Kwantlen. Thanks, too, to the SFU and UBC philosophy departments for providing me with office space and technical support, and to SFU and Kwantlen for providing me with such wonderful teaching opportunities.

Writing acknowledgements is harder than I thought it would be. I fear I have left someone out. Even worse, what I have written seems woefully inadequate to express my gratitude. I can only hope that those to whom thanks is due already know how much I appreciate them.

DEDICATION

In memory of my father, who helped me to see the clouds.

Ian William McIntosh April 1931–February 1994

And in memory of my son, who helped me to see the moons.

Joel Nathaniel Fenger June 1973–May 1996

Chapter 1

Introduction

> Holding as we do that, while knowledge of any kind is a thing to be honoured and prized, one kind of it may, either by reason of its greater exactness or of a higher dignity and greater wonderfulness in its objects, be more honourable and precious than another, on both accounts we should naturally be led to place in the front rank the study of the soul.
>
> Aristotle, *De Anima*. 402^{a} 1-5.

How is it that the intentionality of our thoughts can affect our behaviour? It is commonplace to maintain that what we believe and desire affects what we do. Everyday experience suggests that knowing the contents of someone's beliefs and desires can help us explain and predict that person's behaviour. Elucidating and grounding this pre-theoretic conviction is not easy, especially in the light of modern science. In this dissertation, I investigate issues surrounding one approach to explicating this conviction and situating it within the scientific world view, *viz.*, the family of theories which make appeal to the concept of teleological function.

The unexamined life may not be worth living, but the examined one is oddly puzzling. Each of us grows up bombarded by a multitude of experiences of the changing and complex world, accompanied by a constant sense of ourselves—that we each are somehow the core of our own lives, as experiencers, interpreters, planners, and agents. For instance, we believe that, when things go well, we receive information from the external world, we reason about it, and we act successfully in light of the information we have received. What could be more commonplace?

Commonplace or not, such feats are difficult to explain. One of the aims of the philosophy of mind is to explain how it is that we interact with the world so that we can reason and act as a result. More narrowly,[1] one project has been to try to explain how *what* we believe can affect what we do. Is there a way to reconcile causal explanation, the apparent centrality of the content of mental states to psychological explanation, and a naturalistic view of the world? Perhaps the puzzlement this question engenders reflects the effect of Descartes' sundering of the mental and the physical. Be that as it may, the need for an account, or at least a dissolution, remains.

§ 1.1 *An Overview*

A natural first step in any philosophical enterprise designed to explain intentionality is to provide an explication of the individuation of the content of mental states. Such an explication must accommodate all the empirical data (including the fact that sometimes our thoughts do not correspond to the world), in addition to meeting other demands. For example, it must, in current parlance, meet or adhere to naturalism's allegiance to science, and be compatible with physicalism's ontological picture. Any plausible account must not postulate mysterious forces that violate the laws of physics.

---

[1] Currently, the nature and causal relevance of the subjective aspect of experience are issues of much interest and of lively debate. On consciousness, see, e.g., Dennett (1991), Dretske (1995), and Searle (1992). This topic must, for me, be largely deferred. In this dissertation, I restrict myself to issues about intentionality. I do not mean to imply that consciousness is irrelevant to questions about intentionality. However, I am gambling that solving puzzles about consciousness is not required for solving puzzles about intentionality. If I am wrong, then the approaches I criticise are doubly damned.

One family of contenders for providing such an account can be characterised as the causal/information-theoretic approach. Its proponents include Dretske (1981, 1988), and Fodor (1987, 1994). The idea, very roughly, is that our mental states are to be explained by the informational relationship they bear to the world. For instance (and crudely), our perceptual systems, in the right circumstances, give rise to intentional mental states in us, the contents of which are features of the external world. For example, upon perceptual access to a dog, our perceptual/cognitive systems typically operate in such a way that we token a mental state with the content <dog>. An obvious problem for this sort of approach is that of error. Sometimes, we get things wrong; the content of our thoughts does not match the facts. Sometimes, we token <dog> when we see a raccoon. An adequate account of intentionality cannot type tokens solely in terms of the things that give rise to them, on pain of attributing massively disjunctive contents and precluding error.

It is here[2] that an appeal can be made to "teleological functionalism"; knowing the teleological function of a thing tells us what it is "supposed to" be doing, and this, in turn, enables us to characterise its behaviour, even in instances of failure to perform. A thing can be doing something other than what it is supposed to be doing—hence, the possibility of error. The history of the view goes back at least to Aristotle, but its current popularity is due in part to the resurgence of interest in the philosophy of biology. The leading theories of teleological functionalism draw heavily on evolutionary biology and the

[2] The notion of teleological function also arises in the philosophy of mind by following another route, namely, in Millikan's theory (e.g., 1984, 1993a). I discuss her theory, and the explanatory task she sets for the notion, in § 5.21. There I argue that, despite important differences, her theory encounters the same problems as do the causal/information-theoretic approaches when set the task I have outlined.

theory of natural selection, making them attractive to naturalists, and assuring them a legitimate place in contemporary scientific thought.

In this dissertation, I make explicit a variety of theories of teleological function, and adjudicate between them on a number of different criteria not restricted to issues in the philosophy of mind. I argue that each of this set of theories can be placed in one or the other of two basic categories: selectionist and systems–theoretic. Subtleties in the formulations of each category aside, they can be summarised roughly as follows. Selectionist theories attribute functions on the basis of selection history; a thing's function is that effect or behaviour for which it has been selected. Systems-theoretic accounts attribute functions on the basis of an analysis of components with regard to the workings of a whole; a thing's function is that effect or behaviour which contributes to the performance of the whole of which that thing is a part.

It turns out that no one theory can do it all. The failure of any of the theories to characterise a notion of function that meets all requisite criteria is not inexplicable. The different notions are suited to two different, though related, explanatory projects, depending upon which basic category they fall under.

To see this, consider the following four-fold characterisation of the task of "Darwinian enquiry", as elucidated by Rosenberg: "first, enumerate current and past species that have endured; second, attribute comparative fitness-levels on the basis of species endurance; third, explain the fitness by identifying generic fitness-making characteristics; and fourth, describe in detail the way organisms achieve these

characteristics" (Rosenberg (1996, p. 18, verb tense changed) summarising Kitcher (1993b)).[3]

This analysis guides our understanding of the projects undertaken from within a Darwinian perspective. The first step by itself is purely descriptive. What sorts of things exist and have existed? The second step presupposes that selection will be relevant to the explanation of the continuing existence of a species, since fitness is a measure of the relative aptness for selection.[4] The third step involves an understanding of the sorts of factors that were relevant to selection, without special attention to the mechanics of their operation in the case under investigation. For example, being able to avoid predators may have been a generic fitness-making characteristic. The fourth step involves an understanding of how that fitness-making characteristic is instantiated.[5]

Selectionist construals of teleological function are appropriate when the project is that of explaining why extant features are present in the distribution that they are. That is, they are useful in attempts to perform step three of a Darwinian enquiry. In contrast,

---

[3] Rosenberg is summarising Kitcher's discussion in Kitcher (1993b). There, Kitcher is defending realism from a particular attack by van Fraassen (1980) in which he (van Fraassen) appeals to what "the Darwinist" (p. 39) would say. Van Fraassen argues that the success of current scientific theories can be explained by the fact that unsuccessful theories have been culled—there is no need to postulate the reality of the entities to which the theories (allegedly) refer. Kitcher's response is to explicate the full scope of what Rosenberg calls a "Darwinian Enquiry". As Darwinists, in Kitcher's words, "we can distinguish (i) current species and those that have endured for significant periods of time in the past; (ii) the possession of high Darwinian fitness, borne by those organisms that have the characteristics that dispose them to survive and to reproduce most successfully; (iii) the generic characteristics that endow organisms with high Darwinian fitness (e.g., ability to avoid predators); and, finally, (iv) the particular ways in which organisms belonging to particular species achieve these generic characteristics." (p. 156). I return to the issue of explanation in Chapter 5.

[4] For further clarification of the view that fitness is a propensity, see Mills and Beatty (1979).

[5] This fourth step is, in fact, two-fold. One part involves specifying in what way the characteristic is exhibited. (Is it by camouflage? Is it by fleeing speed?) The other involves specifying how that more narrowly defined trait is exhibited by members of the species in question. (How is that chameleons change their colour? How is it that gazelles run so fast?)

systems–theoretic construals of teleological function are appropriate when the project is that of explaining how these features work. That is, they are useful in attempts to perform step four of a Darwinian enquiry. Furthermore, the systems-theoretic approach, I maintain, may be appropriate in some projects which do not presuppose a Darwinian completion of step three at all. Darwin's insight was that natural selection is a significant factor in accounting for the existence of much of what astonished earlier theorists. For example, the natural theologians of the eighteenth and nineteenth centuries, as exemplified by Paley (1802), argued that the complexity and adaptedness of natural phenomena were evidence of a divine creator. Their confusion was the result of a lack of imagination. However, it is some sort of hyper-Darwinism to believe that natural selection explains *everything* astonishing. Darwin need not be saddled with such a view, and neither need we.

There is currently a near-consensus among those who appeal to teleological functionalism (e.g., Dretske, Neander, Godfrey-Smith) that it is to be analysed solely along selectionist lines. I think that this is due, in part, to the fact that the systems–theoretic approach looks metaphysically suspect to the naturalistic sensibilities of many, seeming to invoke occult forces that are incompatible with a scientific world-view. Darwin is thought to have overthrown Aristotle. I argue that it is Darwinism that offers the grounds to begin to mount a vindication of neo-Aristotelianism, in that natural selection justifies our viewing extant organisms as complex systems, and as components of complex systems, characterisable, in part, in terms of their success. The use of evaluative notions is justified, in part, by reference to the mechanistic story that Darwin

provides at step three of Rosenberg's analysis. That is, the theory of natural selection goes a considerable distance toward underpinning our optimism that the systems that exist now have a propensity to be successful now and should be viewed as such. Performing step three is often what permits us to move on to step four. Selectionist accounts identify fitness-making characteristics which explain the presence of creatures that have them, and systems–theoretic accounts investigate how these characteristics operate.

However, not only does Aristotle need Darwin, Darwin needs Aristotle. Selectionists undermine their own project if they maintain that the features to which they make appeal do not figure in step four of a Darwinian enquiry. The selectionist approach is not to be faulted for failing to go far enough, for failing to perform step four, but it must acknowledge the legitimacy of this further step. If it does not, then the performance of steps one through three must be something of a miracle. Step four is the one that tells us how.

Of course, the systems-theoretic approach is not vindicated solely by noting that Darwinian enquiry requires a step four. A more complete vindication would require demonstrating that the systems-theoretic approach can perform step four. I shall argue that it can. In doing so, I do not deny the importance of selectionist considerations. If natural selection were not the case, then finding ourselves regularly in a position to perform step four would itself be something of a miracle.

I return to my initial puzzlement about our mentality and our behaviour. It is, *prima facie*, a desideratum of any account of intentionality that it explain things in such a

way so as to explicate the view of ourselves and our relationship to the world that had us pondering the nature of intentionality in the first place. Both selectionist and systems-theoretic accounts reach beyond providing an account of the individuation of intentional states, since the former attempts to explain why we have them, and the latter attempts to explain how they work. These two approaches are symbiotic, yet offer different sorts of explanations.

Another move is crucial in the attempt to mount a robust defence of an appeal to systems-theoretic considerations in the philosophy of mind. Recall the question above that got us started: how is it that the intentionality of our thoughts has an effect on our behaviour? Obviously, this question presupposes that the intentionality of our thoughts *does* affect our behaviour. At this point, it is necessary to ask ourselves if a systems-theoretic account is compatible with this assumption. Insofar as a systems-theoretic approach can perform step four, does it do so by appealing to the intentionality of our mental states as causally relevant? If it does not, then either it is an inadequate account, or our presupposition must be abandoned.

Note that, whatever the verdict on the systems-theoretic approach, only accounts that perform step four have any chance of vindicating the intentional in this sense. Accounts that stop at step three are not of the right form. I argue that a systems-theoretic approach can perform step four in the philosophy of mind, and that it does so by appealing to the intentionality of our mental states as causally relevant. What it cannot do *unaided* is solve the problem of providing a naturalistic account of the individuation of all

intentional states in terms of their content.[6] I conclude that, although appeals to teleological function, of both the selectionist and systems-theoretic variety, are not sufficiently helpful with regard to the problem that they have been most recently invoked to solve, they do have roles to play in the philosophy of mind. However, metaphysical and epistemological questions about the unity of science and nature will remain. Philosophy of mind needs both Darwin and Aristotle and, perhaps not surprisingly, a great deal more.

In this introductory chapter, I outline the disjunction problem, motivating my initial foray into the concept of teleological functions. I provide a sketch of how appealing to teleological function is intended to solve this problem. I then make some general remarks about teleology, discussing an assortment of situations in which its invocation has historically seemed appropriate and, in so doing, indicating both its multifarious nature and something of its *prima facie* incompatibility or tension with modern naturalism. Then, I offer some remarks which will seem, depending upon the reader's proclivities, to be either unnecessary or cautionary. Finally, I provide an overview of the dissertation by chapter.

§ 1.2 *The Disjunction Problem*

The problem in contemporary philosophy of mind that most often motivates an appeal to teleological function is the problem of misrepresentation. Given certain

[6] The teleological approach is most often advocated by those seeking a solution to what has come to be know as the "disjunction problem". I elaborate upon this immediately below.

causal/information-theoretic semantic accounts, the fact that the content of a mental state can be a misrepresentation is a mystery.

The problem is roughly this: given some sort of information-theoretic or indicator theory of content, how do we account for error? Since an information-theoretic or indicator semantics assigns content to mental states roughly in virtue of their causes, cases of error seem inexplicable— if the content of a state is determined solely by its cause, how could its content be false? Clearly, we need our theory of mental content to permit misrepresentation but, to phrase the problem in terms of disjunction, how could we ever rule out disjunctive content? We seem to be at a loss to explain why some causes are not relevant to the content of the state in question. As a result, we are in search of a naturalistic way to determine that the content of a mental state can be <A>, and not <A v B>, even if, on occasion, Bs cause tokens of the state in question.

To get a better sense of the disjunction problem, consider a standard example in the literature— that of the content of a frog's perceptual state when presented with a moving fly.[7] Information/indication theories tell us (with various qualifiers) that what is represented in a frog's perceptual state is that which caused the state, or about which the state carries information. In our particular example, the content of the frog's perceptual state could be taken to be <fly>, since we are supposing a fly was what caused the mental state. Suppose, instead, that the frog is presented with a moving beebee. It will snap at it just as it would snap at a fly. Intuitively, we would want to say that the frog has made a mistake; it is misrepresenting the beebee as a fly. Even so, since the beebee *did* (as surely

[7] The paper in which this issue is initially discussed is Lettvin et al (1959). There is a problem with this example, ubiquitous though it (the example) may be. See the next two notes.

as any fly *would*) cause the particular perceptual state in question, there would seem to be no grounds for deciding that the content of that state is mistakenly <fly> as opposed to <fly-or-beebee>. Yet, if it is <fly-or-beebee>, then the frog is not mistaken.

One might be temporarily tempted to say, in response to a case such as this, that the frog's mental state has the content <black dot>. After all, the frog is right in registering the presence of a moving black dot in the environment, but wrong in trying to eat it. I do not mean to introduce the so-called "depth problem", at this point, although it too is alleged to be solved by an appeal to teleological functions (e.g., Neander (1991b), Shapiro (1992), Sterelny (1990)). That problem concerns the question of the appropriate level for the stimulus class. In the frog example, one might wonder whether the content of the frog's mental state is <ambient light reflection>, or <dot-like retinal image>. This, like the disjunction problem, is a problem of the indeterminacy of content. However, it is not a problem to which the concept of error so obviously applies; at the very proximal end of the depth chain, error is rare if not impossible. At that level of closeness, there is no appearance/reality distinction to be made. However, as one moves "out" the chain, the possibility of error increases until, at a certain distance, error is nearly guaranteed. Sufficiently distant features are beyond our discriminatory powers. Teleology may be invoked for this depth problem in the same way it is for the "breadth" problem (which I discuss in the following section). For the purposes of this introduction, I shall assume that the appropriate level at which to characterise the stimulus is external to the perceiver in some intuitive sense, and worry about *how or as what* it is represented.

My point at this juncture is that one might try to dissolve the disjunction problem by maintaining that the mental state is unequivocal and that it is the connection between percept and movement that gets the frog into trouble. If <black dot> were the content of its perceptual state, then the frog would be making an error of content only if it tokened this state when presented with something that did not produce a black dot, e.g., a large and nearby truck.

Unfortunately, the best this move can do is delay the onset of the problem. Let us suppose that the frog's perceptual system is, in this respect, error-free, in that it invokes this mental state only when presented with the appropriate sort of stimulus. This, of course, is a large assumption. It commits us to saying that there is no possibility of error on the part of the frog's perceptual system with regard to this class of stimulus: <black dot> is tokened only when a black dot is perceived (in some factive sense of "perceived"). Yet even if this assumption were justified in this particular case, the problem of error would not be solved. Clearly, a mistake is being made when the frog snaps at a beebee. If the mistake is not in the content of its perceptual state, then perhaps it is in some other state between percept and movement.[8] Besides, surely we do not want a theory of mental content that commits us to perceptual states never being in error.[9] To leave the frog

---

[8] I do not mean to glorify the cognitive capacities of frogs. Millikan (1986), for one, thinks that there simply is no teasing apart a frog's belief from its desires or from its behaviour, and she may well be right. My point is merely that one cannot avoid the problem of error by putting it somewhere else. See the following note.

[9] Fodor (1990c) thinks that the content of a frog's perceptual state when it perceives a fly is indeed <black dot>. The frog has no disjunction problem, but it still has *a* problem of error. The content of its mental state is determinate and, in the scenario envisioned, veridical, ergo the error must be elsewhere.

The disjunction problem is, I take it, peculiar to an information/indicator approach to cognitively sophisticated organisms (such as ourselves). This is most certainly not to deny some continuity and similarity between human perceptual states and those of other animals. The frog

example aside, it would be theory-driven madness to think that we could never token a state with the content <cat> when seeing a raccoon on a dark night. Nobody is so theory-driven as to think this a solution to the problem—hence, the voluminous literature.

§ 1.3 *The Analogy between Dysfunction and Misrepresentation*

The driving force behind an appeal to biological considerations in solving the disjunction problem is the observation that, in order for a token state to *mis*represent, there must be something it is *supposed* to represent. The biological analogue to this situation is the case of dysfunction. In order for a thing (e.g., a token of a trait) to *dys*function, there must be something that it is *supposed* to do. What a thing is supposed to do is its biological or teleological function. Roughly speaking, if a thing does not do what it is supposed to do, it fails to function properly. If a representation has content other than that which it is supposed to have, it is a case of misrepresentation.

The idea is to delineate a sub-class of things that a trait or state does which are the things that the trait or state is *supposed* to do. Crudely put, one guiding intuition is that our biological organs are supposed to enhance our survival and our reproductive success. Particular tokens of biological organs may, in virtue of their performance, thwart or fail to further this end— such instances are cases of dysfunction. (Refinements of this idea appear in Chapter 3.) Mental states are supposed to guide us in successful interaction with the world. Particular tokens of mental states may, in virtue of their performance,

---

example is ubiquitous but somewhat unfortunate in that talk of frogs having false beliefs may turn out to be largely metaphorical.

thwart or fail to further this end—such instances are also cases of dysfunction. Those tokens of intentional mental states which dysfunction in virtue of their contents are misrepresentations.

An appeal to normativity is intended to do more than just say which contents are misrepresentations. It is intended to help us *individuate* contents. One characterisation of a normative analysis of content is this: "if the (would be) representational state R produces some behavior b, and b works to the evolutionary advantage of the animal when and only when external condition C obtains, then R represents (means, has the content) C" (Hall, 1990, p. 194).[10] The reference to the behaviour working to the evolutionary advantage of the animal is not the only route one could take in appealing to biological function (as will become clear below). Nonetheless, this captures one popular avenue. Hall summarises the idea as follows: "R represents F iff the biological function of R is to produce behavior b when Fs are present" (p. 194). Papineau (1987) writes that "the point of emphasizing outputs is that it allows us to fix the truth condition of a representational state as *that circumstance in which the behavioral output of the state produces successful results*" (p. 71). Dretske (1988), though anxious to resist the conflation of output and behaviour, takes essentially the same line. In the face of either natural selection or learning, certain indicator states are favoured, leading to them being "recruited" for their indicational capacities, hence acquiring the function of indicating, and hence acquiring the status of full-fledged intentional states.

[10] It is clear from his article that, despite the differences in typographical case, Hall intends that all the variables range over types, as opposed to tokens. I should also note, lest it appear that he is an advocate, that Hall is critical of the appeal to biological function in individuating content.

Appealing to the normativity of a teleological function is intended to allow us to fix the content of a representation in such a way that misrepresentation is possible. By way of foreshadowing, I shall mention here two *prima facie* difficulties with this approach. First, there seems to be some lurking circularity. We need to know what something is supposed to do, in order to allow for the possibility of failure. At the same time, we need to know what it is supposed to do in order to ascertain what it is in fact doing, if what it is doing is malfunctioning.

Second, this approach could be thought to entail too much "semantic optimism" (cf. Godfrey-Smith (1993)), along with an attendant "touching, but quite unbiological, devotion to truth" (Matthen (1988, p. 13)). There is a tension between, on the one hand, adopting a realist attitude about the external world and the posits of successful sciences and, on the other hand, adopting a naturalistic approach that would seem to warrant only attributions of empirical adequacy, not truth, and to lead to skeptical questions about the representational veridicality of even our perceptions.

I address both these concerns by explaining in more detail how an appeal to ahistorical teleological function should operate.

§ 1.4 *Background Considerations*

In this section, I want to do two things. First, I want to acknowledge the diversity of senses of "teleological" which exist and thereby set aside some concerns that I do not need to address. Second, I want to sketch my views on some relevant background issues,

since I do not want to be coy about the role they play in what follows. It is not my goal to defend these views, only to show that they are well-motivated, thereby setting the stage for what follows.

§ 1.41 *Connotations of "Teleology"*

Appeals to teleology are, and have been, made in a variety of contexts and with a variety of explanatory projects in mind. I shall not be cataloguing these here. Instead, I want merely to set up the conceptual space so as to allow for my investigation. In particular, I want to leave my project open to a number of outcomes. In particular, one of the connotations of "teleological" is both common and unduly restrictive, so I want to acknowledge it and set it aside.

There is a tendency, in keeping with the worries I outline in § 2.1 below, to think that appealing to teleology is legitimate only when the items in question have conscious access to purposes that motivate or rationalise their actions. On this view, teleology is reserved for conscious agents, and talk of teleological functions with regard to anything other than conscious agents is merely metaphorical.

I resist this construal. Although it is clear that conscious agency is importantly unlike other phenomena and needs explaining,[11] restricting the legitimate use of teleology

[11] I do not mean to imply that conscious agency is somehow "other-worldly". I mean only to allow that explaining it *may* require appeal to facets of reality that are required for no other explananda. My point in the text is intended to be neutral with regard to questions about how consciousness is to be explained. However, my construal of physicalism (see § 1.43) imposes some minimal restrictions.

to conscious agency leaves a lacuna. If we restrict appeals to teleology in this way, then some other normative notion must take its place.

I argue for this position indirectly in Chapter 3. I mention it here to forestall some general sceptical worries the reader may have about an appeal to teleology outside the domain of conscious agency. The range of concepts demarcated by the purported desiderata discussed in Chapter 3 extend beyond those applicable to conscious agency, but do so in a principled way.

At any rate, the project of appealing to teleology in helping to naturalise the mental should not be seen as self-refuting. Teleology requires neither consciousness nor representational capacities. I elaborate on this latter consideration in § 2.3 below, when I cull the set of contending theories.

§ 1.42 *The Role of Conceptual Analysis*

This thesis is a work in both conceptual analysis *and* theoretical definition. In some quarters, conceptual analysis has a bad name. Millikan, for example, considers it, "a confused program, a philosophical chimera, a squaring of the circle, a misconceived child of a mistaken view of the nature of language and thought" (1989, p. 290). She contrasts this approach with that of theoretical definition and, needless to say, considers herself to be employed in theoretical definition. On the other hand, Neander, who is largely in agreement with Millikan, labels her own position on the etiological theory as a "conceptual analyst's defense" (eponymously, (1991a)).

There are, I think, no defensible and useful characterisations of conceptual analysis and theoretical definition such that uses of the two must be mutually exclusive. There is no reason to think that necessary and sufficient conditions are any more the domain of conceptual analysis than of theoretical definition;[12] there is no more reason to think that conceptual analyses imply holism about meaning than to think that theoretical definitions do. Both are equally compatible with holism, and issues about the view cannot be resolved by denying the legitimacy of either.

Nonetheless, I do not deny there is a difference. Conceptual analysis is, as Neander writes, "an attempt to describe certain features of the relationship between utterances of the term under analysis, and the beliefs, ideas, and perceptions of those who do the uttering" (1991a, p. 170). It is "the search for the criteria of application that people generally have in mind when they use the term under analysis" (p. 171). Theoretical definition, in contrast, "is an attempt to explain some aspect of the thing referred to, or some aspect of the relationship between utterance of the term and the actual world" (p. 171). Thus, theoretical definition occurs when one invokes a term or concept in an attempt to characterise a feature of the world that is taken to be external to a linguistic community. If this feature fails to exist, then the attendant theoretical definition does not define anything; "phlogiston" and "élan vital" fail to refer because they fail to pick out features of the world.[13] Nonetheless, we can offer conceptual analyses of such terms to explain what people who use them are doing.

---

[12] This is a point made by Neander (1991a), its negation having been implied by Millikan (1989).

[13] Neander claims that, in such cases, there is no theoretical definition (1991a, p. 170). This, I think, is misleading. There still is a theoretical definition, just not one that features in any of "our" theories, where "we" are those who know that the term fails to refer.

This may seem to be an esoteric matter, with little bearing on the project at hand. However, it is relevant to an assessment of what sort of data will be considered to be appropriate. Millikan's criticism of conceptual analysis is aimed, in large part, at those philosophers who consider the ordinary person's linguistic intuitions to be decisive in ascertaining or elucidating the meaning of a term, and hence the viability of a theory that makes use of that term. In particular, she is concerned to defend her theory of proper functions from those who would claim that the ordinary person does not use the term "function" with her technical machinery in mind. If her writing involves work in conceptual analysis, and conceptual analysis is concerned to accommodate the ordinary person's understanding of the terms in question, then clearly her work has failed. Neander deflects this worry by restricting the linguistic community to be consulted to the scientists or specialists for whom it is a technical term.[14] Investigating their criteria of applicability for a term can tell us a great deal about the theory they are employing or developing, and significant failure of our theory to mesh with theirs should be a cause for concern. Allowing this to be a legitimate use of conceptual analysis allows us to incorporate such data in developing and refining our own theories, yet does not restrict us to comportment with the ordinary person's linguistic intuitions.

There is another worry about conceptual analysis that may be implicit in Millikan's attack. There are those who would give it what Frank Jackson calls an "immodest" role; that is, those who would give it a role in arguing for conclusions about the nature of the world. There are others who advocate a modest role for conceptual

[14] Amundson and Lauder (1994) make similar conciliatory remarks about the role of a conceptual analysis in science and the philosophy of science.

analysis such that its use is restricted to understanding the relationships between sets of terms or statements. If conceptual analysis cannot be used immodestly, that is, if it is ruled out of court in considering the nature of the world, then appeals to the ways in which a term is used, by either scientists or ordinary speakers, are ruled out. Any considerations of what people say or would say, when confronted with certain scenarios, is simply irrelevant. Our reactions to thought experiments or descriptions of hypothetical situations would be of no use in getting us to the truth, as they would be evidence only of what people think or say. This is, I think, to ignore that what people say is a function of the theories they hold, and the theories they hold are about the nature of the real world. Conjoining conceptual analysis with empirical evidence *does* provide us a tool for making claims about the nature of the world. What it does not and cannot do is provide us with a means of refuting the skeptic about either knowledge or the natural world.

Let me be even more obvious about the recalcitrance of my optimistic realism—I think it the business of scientists to tell us about the way the world is and that sometimes they succeed. Of course, they use language (and formulae, graphs, and other representational devices) to do so. Theoretical definition is one of their tools. A standard case is one in which scientists posit the existence of some entity or phenomenon that would account for the data they have at hand. As evidence for the existence of the posited entity mounts, the definition becomes less "theoretical". Talk of the entity becomes less theory-bound, or at least more multiply theory-bound.[15] What I am defending is the use of conceptual analysis as a tool for clarifying a general theoretical

[15] This is not to say that terms in our language can ever cease to be theory-infected, let alone that there is a sharp theory/observation distinction. See N.R. Hanson (1958).

framework, which in turn provides us a means for ascertaining truths about the natural world.

So, such is one assumption of the methodological approach of this work, along with a declaration of an optimistic sort of realism. In this context, I simply assume that there is a world independent of our beliefs about it, and about which we can learn.

§ 1.43 *Naturalism and Physicalism*

It is common practice nowadays to declare oneself a naturalist and a physicalist. I participate in this practice, and offer some justification in this section. I start with naturalism.

Almost every philosopher of mind seems to want to be a naturalist. The word "naturalism" is bandied about blithely and tends to elicit a flock of affirmative nods of heads, yet the heterogeneity of the views held by those who take themselves to be naturalists is, *prima facie*, shocking.[16] How could such a motley crew all be naturalists? One explanation (which I favour) is that naturalism is a very weak doctrine. Construing it as such lets those who claim to be naturalists continue to so claim, and turns our attention to whatever substantive differences exist between them.

My somewhat stipulative understanding of naturalism in the philosophy of mind is that it is a methodological doctrine that recognizes a continuity between science and philosophy. Empirical findings can help ensure that our philosophical proclamations

[16] Consider this list of self-proclaimed naturalists: Armstrong (1968), Churchland and Churchland (1983), Chomsky (1995), Dretske (1981), Fodor (1994), Millikan (1989), Papineau (1993), Rorty (1979), and Tye (1992).

apply to the most important of all possible worlds—the actual one. Philosophical analysis, by making explicit our assumptions, helps prevent the over–interpretation of data, and lessens the chance of blindered failure of imagination. Good scientific methodology and good philosophical methodology, while not identical, are importantly complementary. Empirical data meets conceptual analysis, so to speak.

All this probably sounds laudable, but it leaves precious little space for anyone to be other than a naturalist. What is there for anybody to deny? In the context of modern philosophy of mind, it seems nothing. Since, as I said above, so many players on the field purport to be naturalists, it is no surprise that my definition fails to discriminate between them. Being a naturalist is pretty much one of the rules of the game.[17]

What then, of physicalism? I take physicalism to be an ontological doctrine. It's a thesis about the sorts of ontologically respectable things there are in the world, and how, roughly, that world is structured. Here's a first run at how I think a moderate version of the doctrine ought to be characterised.[18]

---

[17] Annoyingly for my stipulative definition, Rorty (1979) considers himself a naturalist, but he most certainly does not fit into the category as I have described it.

Someone else who may well not be a naturalist on my weak construal, is Davidson (e.g., (1970)). I take it that, for him, the necessity for, and nature of, radical interpretation, depending as it does upon the principle of charity, removes psychology from the realm of the rest of science. It does so quite apart from concerns about non-strict laws, which affect the other special sciences as well. Even so, his case is not so troubling from the point of view of my definition, as it is not clear that he considers himself a naturalist.

At any rate, within most of current philosophy of mind, the issue is rarely, if ever, whether or not to be a naturalist, as I've defined the position. Rather, it is *how* properly to be one. My definition is intended to get us to move on to focussing on the relevant differences, without getting too bogged down in terminological disputes. (Fodor considers himself a naturalist (1994), yet attacks "naturalism" (1980), without being inconsistent.)

[18] This characterisation owes something to Van Gulick (1992). I was inspired by his attempts to be clear about distinguishing methodological from metaphysical doctrines, but I think that his own analysis fails to capture the relevant physicalist intuitions. Hence, my characterisation is substantially different.

P1. Necessarily, every object that exists, exists in, and only in, space and time.

P2. Necessarily, every instantiation of a property in the world of space and time is as it is in virtue of the instantiations of physical properties in the world of space and time.

P1 rules out entities that have only non–physical attributes. Thus, for example, the existence of Cartesian minds, construed as essentially and constitutively unextended, is explicitly denied.[19] That is, P1 rules out substance dualism.[20] The addition of P2 ensures that physicalism has slightly more bite. The phrase "in virtue of" is intended to tell us something about the nature of the relationship between the physical and the mental (or, more strictly, between the physical and the non–physical). It stipulates that what is the case at the physical level has some sort of bearing on what is the case at any other "non–physical" level.[21]

The reasoning behind the modal operators in the above definition is this: if physicalism is true, it is a fundamental fact about this world and would be true in all relevantly similiar worlds. The necessity involved is not logical necessity, but rather metaphysical necessity. A characterisation of physicalism that did not involve such a modal dimension would be too weak to handle the intuition that physicalism, if true, captures an important structural feature of our world.

---

[19] Descartes' "locating" the point of interaction of the mind with the body at the pineal gland does not tell against this point. Minds could exist, for Descartes, without such a point of interaction. Unembodied and unlocated minds would not be exemplars of oxymorons. I am less certain that Descartes thought that minds are atemporal.

[20] On my construal, Descartes was a naturalist, but not a physicalist.

[21] The formulation as it stands is mute on the issue of whether or not properties at a non-physical level can have a bearing on properties at the physical level.

A weaker notion of physicalism could be had simply by asserting only P1: in effect, no ghosts, souls, or God.[22] P1 in isolation yields a *very* weak notion of physicalism. It is just a way of saying what there is not. Without the addition of P2, there is no picture of science as consisting of different disciplines that are, in some sense and to some degree, both autonomous and interdependent; there is no indication that certain sorts of properties have a status different from other sorts. The idea that there are levels of phenomena corresponding roughly to the various sciences is a pervasive one[23]—a physicist appeals to properties different from those appealed to by a biologist, but that is no reason (no *prima facie* reason) to think that they are not both doing science; no reason to think that the properties to which they each appeal differ in ontological legitimacy. Nonetheless, "our" physicalist intuitions paint a hierarchical picture of some sort. The addition of P2 nudges our characterisation of physicalism toward one in which this view of the world and of science is implicit.[24]

P2 is explicit in treating physical properties as a subset of all properties,[25] and hence requires that there be a way to delineate types of properties. My mention of the notion of a level is premature, as it alludes to some unspecified structure of inquiry and of

---

[22] Unless such things either are in all of time and all places, or somehow fail to be objects. I have nothing useful to say about theology or spiritualism.

[23] Wimsatt (1976), quoting Plato, claims that "the aim of science is to cut Nature at its joints", then quips that "whatever levels are, they are at least major vertabrae among the joints of Nature" (p.237).

[24] As shall become clear below, I do not consider my "first run" at a definition of physicalism to be particularly satisfying. The hard work remains to be done. What I do hope to show is that appealing to supervenience is not doing said hard work.

[25] P2 leaves open the possibility that physical properties are not a *proper* subset of all properties. That is, P2 is compatible with eliminativism with regard to non-physical properties. In that case, if the properties appealed to in, say, biology are non-physical properties, biology is not a science, at least to the realist. However, eliminativism, whilst compatible with P2, requires independent motivation.

nature. Of course, the question remains: what is it for a property to be a *physical* property? One common, and intuitively plausible if unsatisfying, answer is that a property is physical if it is a property that is mentioned by physics or, better, by the ideal complete physics. Physical properties are simply those to which reference is made in physics. Ergo, given P2, the properties picked out by physics are somehow fundamental, which is another desideratum of a characterisation of physicalism. By itself, this is an unsatisfactory response to the question of which sorts of properties it is appropriate to privilege in the way that P2 does, precisely because it leaves open the question of why physics should have such a privileged role in scientific inquiry.[26] Three answers suggest themselves. One answer is that physics is precise, deals in exceptionless laws, and is a model for us all. A second answer is that the properties to which physics appeals are ubiquitous and, if physicalism is in fact an ontological doctrine, surely it must accommodate such a fact. A third answer is that causality occurs only between instances of the types of properties to which physics appeals. Other properties are subsidiary, in that their causal potency is, at best, entirely derivative.

The first option, that physics deals in exceptionless laws, is, I think, the wrong sort of answer since it draws metaphysical conclusions from the (purported) structure of a theory, rather than explaining the structure of the theory from metaphysical truths. Why is it, for example, that the laws of physics are exceptionless, if in fact they are? Let us ignore the fact that all this presupposes a particularly simplistic view of physics. The

[26] It is an unsatisfactory response for two other reasons, as well. Physicists now do not agree upon ontologies, and the future, ideal, complete physicists may not either. That is, issues concerning, say, realism versus instrumentalism may remain unresolved in physics, even in the long run. Furthermore, "physics" may not pick out that science on which to base fundamental ontological claims. Physics itself may be importantly hierarchical or, at least, multi-dimensional.

laws of physics may rarely, if ever, be exceptionless.[27] A yearning for simplicity may be ill-placed, should the world turn out to be as messy as it seems to be. Yet this yearning, coupled with a belief in the laws of physics being exceptionless, generates a sort of "ideal physics envy", and could account for many of the attempts at the unification of science by reducing the special sciences to physics. Exceptionless laws, were they to apply, would be both aesthetically and explanatorily preferable to untidy ceteris paribus laws, or so this response goes. Such an understanding could underpin a reading of P2.

Nonetheless, I think this answer ill-conceived as well as ill-supported. Why should we, as physicalists, cleave to so pristine a view of physics as a matter of ontological doctrine? Furthermore, even if such a view were vindicated, we would still require explanation and elaboration of the "in virtue of" of P2. Let us grant that the laws of physics are exceptionless. This claim goes nowhere at all toward explaining the fundamentality of properties at the level of physics. I think that advocates of such a claim have in mind something like our third reason above—being an instance of a relatum of a causal relation is being a referent of one of the terms of an exceptionless law, ergo, physics is fundamental. Let us postpone evaluating this line of reasoning.

The second answer mentioned above to the question of why one ought to accord a special status to the properties picked out by physics was that, given some atomistic or constitutive assumptions, physics (or a subdiscipline thereof) deals with those properties common to all objects that exist in space and time. No objects exist outside of space and time, hence physics is always, in some sense, relevant. This sort of answer, though

[27] See, e.g., Cartwright (1983) and Pietroski and Rey (1995).

vague, nets us a somewhat privileged position for physics. No other discipline can claim such generality. Of course, we are still faced with the problem of explaining the "in virtue of" in P2. I will say more on this below.

The third proffered answer is that causality occurs only at the level of physics. This would be a very good reason to privilege the properties postulated by physics. If causality occurs only at the level of physics, then, given the plausible assumption that our interests lie in, and the world operates on the basis of, causal relations, we have every reason to defer to physics. There are at least two problems: i) why think that the only causal connections are at the level of physics, and ii) the nature of the "in virtue of" in P2 remains, once again, unexplained.

My version of naturalism would doubtless annoy Quine, what with its casual references to modality and the causal efficacy of property-tokens. However, I see no reason to think that an allegiance to naturalism requires one to come to love desert landscapes. Dewey characterizes his logical theory of inquiry as "naturalistic"... because in it "there is no break of continuity between operations of inquiry and biological operations and physical operations". The continuity is a matter of the fact that "rational operations [of inquiry]... *grow out of* organic activities, without being identical with that from which they emerge." (Dewey (1938, p. 19)). The lesson I draw from this is that naturalism requires a commitment to pursuing open-minded inquiry wherever it leads.

§ 1.5 *The Structure of this Work*

Here is how I shall proceed. In Chapter 2, I sketch broad concerns about the legitimacy of any appeal to teleological functions, as mentioned above. I then offer a taxonomy of recent analyses of the concept of teleological function, each of which is purported to handle these problems in a naturalistically acceptable way. I divide the analyses into two camps, the selectionist and the systems-theoretic approaches, then further divide each camp into two factions. I outline a total of twelve distinct analyses found in the literature. (The abbreviations I use to refer to the schema for each analysis can be found in Appendices A and B.)

In Chapter 3, I compare the analyses on their capacity to accommodate certain, largely pre-theoretic, distinctions, and on their *prima facie* ability to be naturalistically acceptable. My critical evaluation fails to rule out any one of the types of proffered analyses, but does show that no single kind of approach is entirely satisfactory.

In Chapter 4, I show the reason why: different analyses correspond to different explanatory tasks. Very roughly, I distinguish between explanations of why something is where it is as it is, and explanations of how something does what it does. I argue that certain selectionist analyses of function are appropriate to the former sort of explanation, while some systems-approaches are appropriate to the latter. I also demonstrate that the different explanatory tasks, and hence the concepts to which they appeal, are importantly interconnected; in particular, I argue that no source of normativity can be grounded solely in selection history.

In Chapter 5, I return to considerations in the philosophy of mind and mount a further case for the role of ahistorical teleological function. To do so, I address some concerns that have been raised about teleological accounts by Fodor. I then discuss Millikan's approach to the work that teleological accounts are to achieve by addressing her construal of what and how we are to explain in psychology. I discuss what realism about intentionality requires and I demonstrate the need for a synchronic account even with the framework of her own project.

In Chapter 6, I defend an ahistorical construal (specifically, a structural approach) against some alleged counterexamples. Next, I sketch the limitations of such an approach and suggest further avenues worth taking. Finally, I review and summarise what has been discussed in this dissertation, recapitulating my conclusion that appeals to etiological teleological function are often misplaced and that there is a role for ahistorical teleological function in the philosophy of mind.

Chapter 2

Functions: Systems-theoretic versus Selectionist

> Just as our age has already saved from scorn Democritus' corpuscles, Plato's ideas, and the Stoics' tranquility in light of the most perfect interconnection of things, so now we shall make intelligible the teachings of the Peripatetics concerning forms or entelechies, notions which seemed enigmatic for good reason, and were scarcely perceived by their own authors in the proper way.
>
> Leibniz, *Specimen Dynamicum*, 1695

In this chapter, I articulate the general problems alluded to in § 1.1 that face those who claim to be naturalists but who appeal to teleological functions. These problems, encountered in an attempt to provide an analysis, concern the avoidance of commitment to the existence of any of backwards causation, a creator of the natural world, vital forces, or peculiarly mental forces. I then offer a taxonomy of analyses of the concept of teleological function, each of which purports to handle these problems in a naturalistically acceptable way. This chapter is largely expository—a map of the relevant conceptual space. In the next chapter, I evaluate sites on the map in terms of their ability to meet certain *prima facie* desiderata.

Elucidating and exploring the different conceptions of teleological function before those of explanation may seem an ill-advised method of proceeding, simply because different analyses are appropriate only to different explanatory contexts. How can we begin to evaluate a given conception if we do not know the purpose for which it is to be

used? In a way, we cannot. Deciding, for example, whether or not accommodating a given distinction is a desideratum of a conception cannot be decided independently of fixing an explanatory task. However, I choose this method of presentation because it is a good expository device—the structure of each broad type of theory, and many of the details of their variants, are best understood in the context of disputes over their handling of would-be desiderata. Furthermore, starting with explanation would involve similar difficulties. Deciding just which conceptions of explanation are the ones relevant for consideration in an understanding of teleological functions would require a prior understanding of these functions. One must start somewhere.

§ 2.1 *Problems with Functions*

My interest in functions arises out of my interest in the philosophy of mind. Even so, it is clear that appeals to function are rampant in both everyday discourse and science. The function of the heart is to pump blood, as Harvey realized in the early 1600s. The function of the fourth leg on a table is to provide greater stability (let us suppose). The function of a gas gauge is to indicate approximately how much fuel remains in a tank. The function of the kidneys is to separate waste products from the blood. The function of the extra blanket on my bed in winter is to keep me warm and the function of traffic laws is to facilitate the safe and efficient movement of traffic.[1]

[1] I largely rely upon intuition to support the claim that such functional statements are teleological, but see § 1.4 above. See, also, Wimsatt (1972) for a careful delineation of various sorts of function statements, and a demarcation of those that clearly are not teleological (e.g., mathematical function statements).

Questions then arise: how are such statements to be analysed? Are they all fundamentally of the same type? How do the explanations of which they are a part integrate with the rest of science?

Some branches of science, such as most areas of biology, appear to require function statements. There are those who would respond, given the difficulties to be discussed below, "so much the worse for such areas of science". Their view would be that biology either must somehow be reducible in principle to physics and/or chemistry, or it is simply a misguided or ersatz discipline. For advocates of the latter view, there is no puzzle about function statements that cannot be solved by eliminating them—with the exception of the puzzle of the apparent incorporation of such statements into theories with predictive success and apparent explanatory value. I do not intend to argue directly against such eliminativism here and will say only that such a response seems premature.

Of course, the eliminativist is correct in demanding an analysis of function statements that is in keeping with naturalistic constraints on science.[2] Function statements even in ordinary discourse seem to require an analysis; the need is all the more pressing in science.[3] This is so for two reasons. First, it is the business of science to

---

[2] Given my construal of naturalism in Chapter 1, it should be obvious that I do not think a commitment to it entails a commitment to inter-theoretic reductionism. "Reductionism" is a notoriously vexed term. I take it that inter-theoretic reductionism is the view either that the laws of any legitimate special science must map, by means of some sort of bridge laws, to those of basic science, or that the terms of the special science must be definable in the vocabulary of basic science. Nothing said earlier entails either of these views. A different sense of reductionism, let us call it "ontological reductionism", amounts to the view that, in order for phenomena be ontologically respectable, they must conform to physicalist strictures, such as P1 and P2. If this be reductionism, then I am a reductionist in this sense. I prefer to think of myself as a compatibilist.

[3] The degree and nature of the connection between function statements in discourse about artefacts and function statements in discourse about naturally occurring items is a matter of debate. More on this below, § 3. 4.

analyse the claims it makes or, at the very least, this is the business of the philosophy of science. Second, appeals to function have sometimes been thought to entail a commitment to at least one of the following objectionable theses: i) backward causation, ii) the purposive design of natural features of the world, iii) vitalism, and iv) mentalism.[4]

Consider the following function statement: 'the function of the melanin in my tissue now is (at least in part) to absorb the Sun's harmful radiations at some future time'. This may seem to entail backward causation, since we do not think it the case that a thing has a function only while it performs the activity that constitutes its function, or only after it has done so. The melanin in my tissue has its function even when I am indoors, and it had its function even before my first exposure to the Sun. Nonetheless, it is unclear how we are to understand the claim that the melanin in my tissue now has the function of doing something later, unless the future event somehow retroactively bestows a function. The claim that such metaphysical baroqueness is clearly unpalatable does not rely upon any substantive view on naturalism in science. At any rate, the appeal to backward causation appears doomed when we consider that the melanin in someone's tissue may *never* absorb the Sun's harmful radiations, either because of circumstance or defect, yet will have that function nonetheless.

---

[4] Neander (1991b, pp. 455-59) defends an etiological conception of function against the first two of these objectionable theses. Allen and Bekoff (1995) offer a slightly different taxonomy. They do not address the purposive design thesis *per se*, but instead discuss the claim that functional notions are thought to be incompatible with mechanistic explanation. These two concerns are closely related, and a satisfactory handling of one will likely handle the other. Allen and Bekoff add to the list of concerns the methodological worry that teleological notions are inherently empirically untestable. This concern is addressed in more detail in Chapter 4, in which I discuss explanation.

Three other objectionable theses are related. Suppose that, instead of being instances of backward causation, functions are determined by the purposes that are, or could be, or are intended to be, served by the thing in question. On such a picture, the reason that the melanin in my skin has the function it does could be that God put it there with that in mind. In the case of artefacts, an account of this sort does not seem objectionable (see § 3.4), but it does not translate well into the realm of natural phenomena. It may be legitimate to appeal to the intentions of the designer of an artefact to determine its function (e.g., the switch on the wall has the function of turning the lights on or off, since that is the purpose for which its designer made it), but such a move in science seems to commit us to the existence of a Grand Designer. If the melanin in my tissue derives its function from its purpose in *this* way, then our metaphysical picture is, once again, unfortunately baroque. It (nearly) goes without saying that if function statements are to be respectable in modern science, they require an analysis of their teleological nature that does not rely on the existence of a purposeful creator.

Vitalism is the view that biological phenomena contain a force that non-biological phenomena do not.[5] As a view about what distinguishes living from non-living things, it involves invoking a quasi-Cartesian distinction between mere matter and living matter. Living matter is matter *plus* a vital force, just as, to Descartes, humans are matter plus mind. To the vitalist, it is a vital force that gives functional items their forward-looking nature. The melanin in my skin may or may not have been put there by God, but what gives it its function is, roughly, this force. Nevertheless, as an explanation or analysis of

[5] In modern times, something like this view has been held by Bergson (e.g., 1913). It has also been attributed to the emergentists, but they typically consider their project to be in opposition to vitalism. See, e.g., C.D. Broad (1925).

function, vitalism is flawed. It amounts to positing in the explanans the existence of something that is defined solely in terms of the explanandum.[6] Even on a very weak construal of naturalism, such a view seems question-begging and metaphysically suspect, not to mention ontologically extravagant.

Finally, there is mentalism—the view that the vital forces that distinguish the animate from the inanimate, and that ground function are, in some sense, mental. Attributing a function to a biological feature, on this view, is to attribute purpose to it, where purpose is understood to be a mental phenomenon.[7] This sort of mentalism is an attempt to specify the nature of the vital force postulated by vitalism, but it suffers all the defects of vitalism while being even more implausible. It is difficult to see that anything is gained by maintaining that the vital force has mentality.

So, the task of those who would defend the legitimacy of teleological functional talk in the biological sciences is to characterise function and related concepts in such a way as to avoid implying backward causation, the existence of a Grand Designer, or the existence of mysterious animistic forces. This much, at least, is demanded of the naturalist.

---

[6] Historically, vitalism sometimes arose in the context of denying a simple reductionism of the biological to the physico-chemical, on the grounds, roughly, that those two positions exhaust the options and reductionism is clearly false. As such, it is not quite the circular view that I have characterised it to be, but would still be in need of independent support. See Mayr (1976) for discussion.

[7] Someone who may have held this view is Driesch (1908).

§ 2.2 *Some Proposed Solutions*

Despite some disagreement in the literature as to how best to taxonomize the analyses of function, I divide them into two main camps: selectionist theories and systems theories.[8] In what follows, I first provide a brief characterisation of each sort of theory, then offer some details and refinements in the following section by means of a close look at some representative and plausible examples. In the next chapter, I evaluate each theory's success in meeting certain desiderata.

For a quick explication of selectionist theories, it is simplest to focus straightaway on the domain of biology. To quote Neander (1991b), "the central element of the etiological approach [a version of the selectionist view] should be seen as the simple idea that a function of a trait is the effect for which that trait was selected" (p. 459). The function of the trait of having an opposable thumb is to enable its bearers to grasp objects, because it was this effect that was selected for.[9] On this view, it is not the (or even a) function of the trait of having an opposable thumb to enable its bearers to hitchhike, since this effect was not selected for.[10] A trait may have many effects, but only some of them will play a causal role in natural selection. The polar bear's thick coat is selected for its

---

[8] I am not claiming that my taxonomy exhausts all possible positions on function, or even all actual ones. But I hope to have covered most of the important conceptual space.

[9] Despite its grammatical inelegance, the phrase "selection for" and its cognates have become standard in the literature on evolutionary theory. For more detail on the selection for/selection of distinction, see Sober (1984) and § 4.111 below.

[10] Any temptation to claim that this *is* a function of the opposable thumb arises from a tendency to collapse the notion of function into the notion of usefulness, according to this view. The thumb may function *as* a ride-getter, but that is not one of its functions. See § 3.3 for relevant discussion of the function/accident distinction. The central issue of explanation is discussed in Chapter 4.

effectiveness in keeping the animal warm, not its effectiveness in adding to the animal's weight. There is selection *for* warmth, selection *of* heaviness.

The etiological approach can be found in the work of Larry Wright (1973), Ruth Millikan (e.g., 1984, 1989)[11], Fred Dretske (1988), Karen Neander (e.g., 1991a, 1991b), and Peter Godfrey-Smith (e.g., 1993, 1994). More details follow in § 2.211 below.

Another version of the selectionist approach is the propensity theory. This theory can be summarised by the claim that "something has a (biological) function just when it confers a survival-enhancing propensity on a creature that possesses it" (Bigelow and Pargetter, 1987, p. 192). It is not clear whether "survival" in Bigelow and Pargetter's work means mere continued existence, or continued existence and the reproduction of fertile offspring. The theory is more plausible if one assumes the latter in so far as their view is motivated by considerations of the importance of evolution by natural selection to biological items and theories. "Survival-enhancing propensity" may be seen as analogous to contributing to selection for. More on this below, § 2.212.

Initially, I had considered the propensity theory to be in explicit contrast to the etiological approach, which is how Bigelow and Pargetter view their project. Kitcher (1993) convinced me of the merits of thinking otherwise. The two theories do belong in the same camp insofar as they are selectionist. However, I also think there is a not implausible interpretation of their theory that renders it analogous to the systems approach discussed immediately below. As a consequence, the propensity theory appears,

---

[11] Millikan does not think that her view has any similarity to Wright's: "Various published remarks to the contrary, there is no overlap at all between Wright's analysis of functions and mine" (1989, footnote 5). I do not agree, as is obvious from my categorising them together. I shall attempt to rebut her claim indirectly below, by discussing the relevant features of both their views that qualify them as selectionist.

on my analysis, in both the selectionist and systems-theoretic camps. I argue below (§ 4.112) that it best fits into the systems-theoretic family.

The systems approach, which is notably ahistorical, is comprised of two sub-groups. One group, which I shall call "structural theories", is best represented by Cummins (1975). This group also includes Bigelow and Pargetter (1987), Prior (1985), and Wimsatt (1972). To run roughshod over some significant distinctions in approach, the basic idea can be captured this way: feature *A* has function *F*, in system *S*, iff *A*'s performing *F* contributes causally to the capacity of system *S* to do *G*.[12] For example, the heart has the function of pumping blood because this feature contributes causally to the capacity of the system of which it is a part, a human body, to survive. The manifestation of the larger capacity is explained, in part, by the manifestation of the 'component' capacity.

A variant of the systems approach is a goal–directed analysis, and is endorsed by such people as Boorse (1976), Adams (1979), and Schaffner (1993). The idea is similar to that of the structural theories, except that the capacity *G* is considered to be a *goal* of the system of which feature *A* is a part. This restricts the range of the larger capacities, the component capacities of which are to be considered functions. (See § 2.222.) This approach, like the structural, is importantly ahistorical. Even if, as must be the standard case, the goal is a *future* state, it is having the goal *now* that is relevant.

To summarise, it is helpful to distinguish selectionist accounts from systems-theoretic ones. The former sort I have divided into the etiological and the propensity

[12] The mere fact that I have grouped this particular set of philosophers together likely seems cause for suspicion. I motivate my choice by showing, in § 2.221 below, that the views of each can be interpreted as roughly conforming to the schema I offer.

approaches. The latter sort I have divided into the structural and the goal-directed approaches. In the remainder of this chapter, I elaborate upon these general approaches.

§ 2.21 *Selectionist Theories*

Selectionist theories are usefully divided into two families—the etiological and the propensity approaches. I discuss them in turn.

§ 2.211 *Etiological Theories*[13]

The etiological approach has been developed by several authors, five of whom I discuss immediately below. Larry Wright's (1973) and (1976) are considered seminal. He offers his work as an analysis of the general notion of teleological function, and as a correction to analyses that assign functions to anything useful (he has in mind Canfield (1964)) and that fail to accommodate artefacts (he has in mind Beckner (1959)), or that allow for "accidental" functions (both Canfield and Beckner, according to Wright). His view (to which I'll refer sometimes using "LW" below) is this:[14]

> The function of *A* is *F* iff:
>
> (i) *F* is a consequence (result) of *A*'s being there,
> (ii) *A* is there because it does (results in) *F*,
>
> (1976, p. 161, variables changed for consistency)

[13] Godfrey-Smith (1994) refers to this approach as "The 'Wright' Line".
[14] A list of abbreviations and their attendant schemata can be found in Appendix A.

Thus, the function of the heart is to pump blood because pumping blood is what hearts do, and hearts occur in humans because they pump blood. Despite its surface circularity, this account, when properly interpreted, provides a good starting point for the etiological conception.

Wright thinks that his account is univocal with regard to natural and artefactual phenomena (see § 3.4 for discussion), but, for ease of exposition, I shall explicate his account in terms of natural phenomena. If the variable "*A*" is understood to refer to a type (as opposed to token) of a trait, and the "because" clause is intended to refer (indirectly, at least) to natural selection, then we have something very close to the view of Neander (1991b) mentioned above: "a function of a trait is the effect for which that trait was selected" (p. 459). Wright offers an intuitively similar explication: "the function of *A* is that particular consequence of its being where it is which explains why it is there" (1976, p. 78, variable changed for consistency).

Next is Ruth Millikan. Her overall project is to provide a theory of intentionality, and her definition of function is stipulative. She is not concerned to accommodate common usage or other theoretical purposes.[15] Her formal definition of function (to which I may refer as "RM") is this:

> Where *a* is a member of a reproductively established family *T* and *T* has the reproductively established or Normal character *C*, *a* has the function *F* as a direct proper function iff:

[15] As I noted in Chapter 1, Millikan considers conceptual analysis, which she views as the contrast to theoretical definition, as "... a confused program, a philosophical chimera, a squaring of the circle, the misconceived child of a mistaken view of the nature of language and thought" (1989a, p. 290).

(1) Certain ancestors of *a* performed *F*.

(2) In part because there existed a direct causal connection between having the character *C* and performance of the function *F* in the case of these ancestors of *a*, *C* correlated positively with *F* over a certain set of items *I* which included these ancestors and other things not having *C*.

(3) One among the legitimate explanations that can be given of the fact that *a* exists makes reference to the fact that *C* correlated positively with *F* over *S*, either directly causing reproduction of *a* or explaining why *T* was proliferated and hence why *a* exists.

(1984, p. 28, some variables changed for consistency)

In this definition, the term "function" *simpliciter* means effect, the term "proper function" being reserved for what I have been calling "function" or "teleological function". A reproductively established family is, to a first approximation, a family established by some sort of copying. A Normal character is the set of properties common to all members of a reproductively-established family. Clauses 2 and 3 are a spelling out of the afore-mentioned notion of selection for.

These views are not without difficulties, some of which will be discussed below.[16] In light of these difficulties, other etiological conceptions have been proposed. I shall outline three—those of Godfrey-Smith (1994), Griffiths (1993), and Bedau (1991, 1992a, 1993).

Godfrey-Smith offers what he calls a "Modern History Theory of Functions". (Henceforth, it may be called "GS".) It is intended to do the same theoretical work as

[16] The need for modification and refinement may seem, at this point, unsubstantiated. I ask the reader to bear with me.

Millikan's, but also to capture the general biological sense of 'function'. His formal definition is as follows:

> The function of *a* is to *F* iff:
>
> (i) *a* is a member of family *T*,
> (ii) members of family *T* are components of biologically real systems of type *S*,
> (iii) among the properties copied between members of *T* is property or property cluster *C*, which can do *F*.
> (iv) one of the reasons members of *T* such as *a* exist now is the fact that past members of *T* were successful under selection in the recent past, through positively contributing to the fitness of systems of type *S*, and
> (v) members of *T* were selected because they did *F*, through having *C*.
>
> (1994, p. 359, some variables changed for consistency)

On Godfrey-Smith's analysis, a family (*T*) is a set of entities generated by some sort of copying, roughly the same idea as Millikan's reproductively-established family.[17] My heart is a member of the family of hearts insofar as it is one of a long line of hearts, the method of "copying" being genetic. Biologically real systems (*S*) are those systems catalogued by biology.[18] Being selected for is a causal process, the result of both suitedness to an environment, and success relative to one's competitors.[19]

---

[17] This sort of view owes a lot to Millikan (1984), as Godfrey-Smith acknowledges. In Millikan's technical vocabulary, hearts are members of "higher-order reproductively established families" in that they are not reproduced directly one from another, but via the genetic route of the reproduction of the organisms that house them.

[18] According to Godfrey-Smith, creating this catalogue is a matter of ascertaining which sorts of things are "interactors", i.e., subject to selection pressures. I think this stipulation somewhat question-begging, insofar as the issue of the relationship between selection and function needs to be addressed. However, I do not think it an unfair first run at understanding "system" in biology.

[19] Neither condition is separately sufficient for a thing's being selected for. Something could be well-suited to its environment yet be driven from the gene pool, if other things are *better* suited.

There are several other things to note about this analysis. First, the item whose function is in question would seem to be a token, not a trait. It is my (token) heart that is a member of the family (type) of human hearts. Second, the analysis is etiological in roughly the same sense as Wright's. In ascertaining a thing's function, it is necessary to look to the past to see whether or not ancestral instances of it have been selected for. Third, Godfrey-Smith's analysis, like Wright's, requires the possibility of current performance of the effect considered to be its function. What Godfrey-Smith's analysis adds is the notion of reproduction. This addition is helpful in overcoming certain counterexamples, discussed below. Fourth, the definition is explicitly restricted to natural (biological) phenomena.

The fourth variety of etiological approach I want to discuss is that of Paul Griffiths (at times dubbed "PG" below). His view is offered, like Godfrey-Smith's, as a general account of biological function and it, too, is not offered so much as competition to RM, but as a refinement. He writes:

> Where *A* is a trait of systems of type *S*, a proper function of *A* in *S*s is *F* iff a proximal selective explanation of the current non-zero proportion of *S*s with *A* must cite *F* as a component in the fitness conferred by *A*.
>
> (1993, p. 418, requisite variables changed, and syntax amended)

On the other hand, something could be better than its competitors, but still not good enough. Even so, of course, something could be the best available option, meet minimum requirements, but not be selected for because of misfortune, so to speak.

By "proper function", Griffiths means the set of teleological functions (or, perhaps, a proper subset thereof). By "proximal selective explanation", he means an explanation which makes reference to recent episodes of the "selection for" process mentioned above. By "component in the fitness conferred by *A*", he means an action or attribute that is causally relevant to the afore-mentioned selection process—that is, an action or attribute the performance or manifestation of which was causally relevant to creatures with *A* having been selected. The analysis is restricted explicitly to types, not tokens.

The final etiological theory to which I wish to make appeal is that of Bedau (1991, 1992a, 1993). He has a significant point of disagreement with the above authors—he offers an analysis of teleology which requires acknowledging value in the world. He claims that for something to be the appropriate sort of thing to which to ascribe teleological properties, it must have intrinsic interests; it must be the sort of thing about which it is appropriate to say that something can be *good* for it. Bedau makes this move because he thinks that otherwise an etiological analysis will assign functions too liberally. Without an appeal to value, Bedau argues, any thing that has been selected will, counterintuitively, have a function.[20]

Bedau offers the following schema (henceforth, "MB"):

The function of *A*'s *B*ing is *F* iff
*A B*s because [*A*'s *B*ing contributes to *F*ing and *F*ing is good]

[20] A discussion of one of Bedau's examples occurs in Chapter 6.

(based on 1992a, p. 790, the square brackets are intended to show the scope of the "because")

On this schema, the function of the heart's constricting rhythmically is to pump blood if and only if the heart constricts rhythmically because doing so contributes to the pumping of blood, and because the pumping of blood is good. Rather a lot is packed into this latter clause. It is intended to capture the "consequence etiology" (1993, p. 34) of the etiological position, in that the fact that the pumping of blood is good partially explains the rhythmic constricting of the heart. Bedau writes that his theory "can be viewed as a modification of Wright's analysis to which a value condition has been added" (1993, footnote 33). This value condition itself is partially captured by the reference to *F*ing being good. Something (*F*ing) is good for a system or entity (*S*) if and only if (roughly) "(i) *S* is the kind of thing that has its own interests...; (ii) *S*'s good is independent of any value that some third party might place on *S*; and (iii) *F*ing is in the interests of *S*, i.e., *F*ing promotes *S*'s interests or constitutes (at least part of) *S*'s interests" (1992a, p. 45, variables changed for consistency, including a change to progressive from noun). There is no explicit reference in the above schema (MB) to system *S*, but it is clearly a requisite part of the theory. *S* may be some larger system that contains or uses *A*, or may be identical to *A*.

It is not obvious whether the variables are to range over types or over tokens. Perhaps either is appropriate. The schema purports to provide necessary and sufficient conditions for the ascription of function.[21] The function in question need not actually be

[21] Bedau distinguishes between three grades of teleology, and this schema provides an analysis appropriate to his third, the "full-blooded" one.

being performed in order for the ascription of that function to be appropriate, but it is not clear whether or not *A* be able to *B*, hence able to result in *F*.

The most salient difference between MB and the other etiological selectionist theories is the reference to goodness. It is motivated by the fact that MB faces the same difficulties encountered by other etiological theories, insofar as it is etiological.[22] However, attributing unreduced value to features of the world is a troublesome move for a naturalist, and one that should be resisted if possible.

The etiological selectionist approach is currently very popular, which justifies my offering the above five different analyses. I take the first one to be something of a founding paradigm, the next three to be good exemplars of the currently dominant view, and the fifth to be an interesting variant.

§ 2.212 *Selectionist Propensity Theories*

The basic idea of a selectionist propensity view is that a trait has a teleological function when its performing the effect said to be its function is advantageous with regard to current selection pressures for the organism which has the trait. This view has fewer explicit formulations in the literature than does the etiological view. As noted above, Bigelow and Pargetter (1987) write "something has a (biological) function just when it confers a survival-enhancing propensity on a creature that possesses it" (p. 192). Part of Bigelow and Pargetter's motivation for appealing to propensities, as opposed to the actual

[22] See the next chapter and § 6.11 for further discussion.

selection history of a trait, is their desire to explicate the intuition that appeals to function are "forward-looking" in a way in which appeals to causal history are not.[23] As they note, there seems nothing particularly odd about saying that "[t]he function of teeth at time *t* is to pulp food at time *t'*, where *t'* > *t*" (p. 181). On their view, saying what it is that teeth did in the past does not adequately explicate a claim about what it is their function to do in the future. Appealing to propensities is intended to correct this perceived defect.

The interpretation of their view that renders it selectionist is one according to which the relevant propensity is one which furthers the likelihood of, or capacity for, selection. It is, of course, possible to have a propensity view which focusses on something other than selection. One might wish to investigate the propensity of some things to become square given certain conditions, although it is difficult to see what explanatory task one would be attempting to perform. Less fancifully, one might wish to investigate the propensity of some things simply to survive under certain conditions, whether or not they are subsequently selected. With regard to natural phenomena (and perhaps even artefactual phenomena, see § 3.4), the set of things that have a propensity to survive may be co-extensive with the set of things that have a propensity to be selected, if surviving is interpreted broadly enough.

At any rate, here, in keeping with the schema offered above for the etiological selectionist approach, is a schema for the selectionist propensity view (hence, "BP"):

---

[23] This may be because they see the relevant explanatory task to be that of explaining future selection, although it is not clear. More on this in Chapter 4.

> The function of *A* is *F* only if the performance of *F* by *A* confers a survival-enhancing propensity on the system *S* of which *A* is a part.
> (based on Bigelow and Pargetter, 1987)[24]

In Bigelow and Pargetter's work, there is some equivocation as to whether it is traits or tokens that are the topic at hand. I think that their view is most plausible when restricted to types, but shall leave it ambiguous for now. I look at both a type and a token reading below.

§ 2.22 *Systems Theories*

Systems theories are usefully divided into two families—structural and goal-directed approaches. They differ from selectionist accounts both in not appealing to selection and in not being temporally focussed either back in time (as are etiological selectionist accounts) or forward in time (as in the selectionist reading of the propensity approach). For this reason, they can be considered ahistorical accounts and I shall sometimes refer to them as such. However, the term "system" has the merit of focussing us on the schema I am about to present, so I shall use it primarily. In what follows, I discuss the two systems-theoretic families in turn.

[24] This is their explicit formulation, though it is not always clear that they do not intend to provide sufficient, as well as necessary, conditions. If they intend to provide both (as I shall argue in the text that they should), then "only if" should read "if and only if".

§ 2.221 *Structural Theories*

There is a family of theories, exemplified by Cummins (1975), that offers a theory of functions that clearly is not selectionist. Although I am deferring an in-depth discussion of explanation until the next chapter, an explication of Cummins' view requires mention of what he takes to be the explanatory project for which his analysis is designed. Certain types of things exhibit a "dispositional regularity" in certain sorts of circumstances; i.e., they manifest a disposition "given the requisite precipitating conditions" (p. 758). The fact that, say, sugar dissolves in water, requires explaining. Cummins thinks that there are two legitimate explanatory strategies in such cases. One is the "subsumption strategy", in which we explain token manifestations of a disposition *d* by item *a* by ascertaining what sorts of events trigger the manifestations, then subsuming the connection between them and the manifestations under "one or more *general* laws, i.e., laws governing the behavior of things generally, not just things having *d*" (p. 759, emphasis his). The other strategy, the "analytical strategy, proceeds by analyzing a disposition *d* of *a* into a number of other dispositions d_i . . .d_n had by *a* or components of *a* such that programmed manifestation of the d_i results in or amounts to a manifestation of *d*" (p. 759). By "programmed", Cummins means "organized in a way that could be specified in a program or flow chart: each instruction (box) specifies manifestation of one of the d_i such that, if the program is executed (the chart followed), *a* manifests *d*" (footnote 16). To explain the disposition of, say, a type of organism to survive, one would analyze "the organism into a number of 'systems'—the circulatory system, the

digestive system, the nervous system, etc.,—each of which has its characteristic capacities" (p. 761).[25]

The analytical strategy is applicable in the case of functions as shown by the following schema ("RC"):

> the function of *a* in system *s* is *F* relative to an analytical account X of s's capacity to *G* just in case *a* is capable of *F*-ing in *s* and X appropriately and adequately accounts for *s*'s capacity to *G* by, in part, appealing to the capacity of *a* to *F* in *s*.

(taken from p. 762, syntax altered, some variables changed for consistency)[26]

The upshot is, as Cummins states, that "to ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system" (1975, p. 765). Cummins' project is to explain dispositional regularities, which would seem to indicate that the variables are to range either over types of items, or over types of events of a token[27]. In both cases, it is assumed that the item (or items) *has* (or have) the capacity in question. This requirement will be relevant to my discussion in Chapter 3.

---

[25] As Cummins notes, when using the analytical strategy, speaking of capacities as opposed to dispositions seems more natural. He does not elaborate on what may ground this intuition, though intuition suggests it may be something about the "forward-looking" connotations to 'disposition' that do not attend 'capacity'.

[26] As with some other analyses, RC is presented as providing necessary but not sufficient conditions. If a bi-conditional formulation is required, "just in case" would need to be followed by "and when".

[27] Walsh and Ariew (1996) maintain that structural functions apply to tokens only, and that type membership is irrelevant (p. 505). I disagree.

Cummins' approach places few restrictions on our choice of dispositions or capacities on which to use the analytic strategy, as he claims that our explanatory interest is a deciding factor. Any disposition that interests us for whatever reason *could*, in principle, be given a Cummins-style analysis, although the appropriateness of doing so is, he writes, a matter of "(i) the extent to which the analyzing capacities are less sophisticated than the analyzed capacities, (ii) the extent to which the analyzing capacities are different in type from the analyzed capacities, and (iii) the relative ... complexity of the organization of the component parts/processes that is attributed to the system" (p. 764). With very simple systems, the analytical strategy is of no more explanatory interest than is the subsumption strategy, so an appeal to functions is rendered otiose. Whether or not something should be considered a function is a matter of whether or not so considering it has explanatory value. The issue of the range of capacities to which to apply an analytical strategy, and hence to which to assign functions, arises again below (in, e.g., § 4.12).

The above discussion of Cummins was an exposition of a paradigm of the structuralist family of the systems-theoretic approach. I now consider two other views that fall into this same family—a version of the propensity approach discussed above, and the view of Wimsatt (1972).

The propensity view (BP) was categorised above as selectionist because of its focus on survival, and the assumption that survival and selection are closely linked. However, let us look at the schema again, with the structural approach in mind.

> The function of *A* is *F* only if the performance of *F* by *A* confers a survival-enhancing propensity on the system *S* of which *A* is a part.

This view is, I contend, a restricted variant of RC. The variables "*A*", "*F*", and "*S*" play the same roles that they do in RC (though RC permits more sorts of things to fill them, in part because that analysis is not restricted to biological systems). Survival itself is analogous to the capacity *G*, while the "survival-enhancing propensity" may be the same thing as *F* (in the case of simple systems), or another, larger capacity (in the case of larger, more complex systems). For example, the function of the heart is pumping blood, which confers the survival-enhancing propensity to oxygenate organs important to the survival of the organism of which the heart is a part. The "performance of *F*" in BP is the manifestation of *F* implicit in RC in the stipulation that the capacity to *F* be relevant, via an application of the analytical strategy, to explaining the capacity to *G*. The implicit relativization of BP to a normal environment (Bigelow and Pargetter, 1987, p. 192, see § 3.11 below for discussion) is analogous to RC's "requisite precipitating conditions" (p. 758). The BP project could be seen, again analogously, as that of explaining what it is about a complex biological organism that permits it to survive in its normal environment.

Thus, I think BP may be read as advocating a restricted type of Cummins' analysis. Doubtless, there are differences between the two views, and I do not want to press the comparison too hard. However, more support will be offered below, when it is noted that both theories have difficulties handling the same sorts of would-be desiderata and that explanations they offer are fundamentally ahistorical.

Finally, one more systems approach that fits into the structuralist family is that of Wimsatt. Wimsatt (1972) distinguishes many senses of "function", most of which are clearly not at issue in a discussion of teleological function, as he realizes. The three he considers relevant to the present discussion are the "perspectival", "evaluative", and "teleological" senses. The first involves assigning a function to a component in terms of its contribution to what a system is doing, the second in terms of what a part does that contributes to the "good" of its containing system, and the third in terms of what a part does that contributes to "the attainment of some end or purpose of some user or system" (p. 5). Only the last notion is genuinely teleological, according to Wimsatt.

The perspectival sense corresponds to Cummins' analysis, and assigns functions equally liberally. The evaluative sense demarcates a subset of structural functions, *viz.*, those that contribute to the "good" of the organism that has them.[28] The teleological sense is more restrictive still,[29] in that purpose is required. (I will say more on purpose below.)

Wimsatt characterizes the schema of (teleological) function statements as this (hence "WW"):

> According to theory *Y*, a function of item *i*, in producing behaviour *B*, in system *S* in environment *E* relative to purpose *P* is to *F*.
>
> (1972, p. 32, some variables changed for consistency)

---

[28] The evaluative sense bears considerable resemblance to Bedau's approach, though Wimsatt's theory is not, I maintain, a selectionist one.

[29] It is not clear to me whether or not teleological functions are a subset of evaluative ones. Insofar as the analysis is intended to capture artefacts or human actions, it would seem not. Humans are all too capable of acting with self-destructive purposes.

Superficially, the similarity to Cummins' approach is apparent. The function of an item (say, the heart), in producing behaviour (rhythmic constricting), in a system (the organism which houses it), in an environment,[30] relative to the "purpose" of the organism (surviving), is to pump blood.

Wimsatt's notion of purpose *P* is more restricted than Cummins' of capacity *G*; without it, his teleological sense of function is equivalent to his perspectival one, in which any consequence of the operation of an item can be analysed functionally, given a view of what the system is doing. The choice of purpose is not simply a matter of interest, but also, as is the case with Cummins, of complexity. He thinks that there is a principle of theoretical relevance; "*...citing the purpose or end of a functional entity plays a role in explanations concerning the existence and form of that entity*" (p. 5). Despite the fact that this construal of the appropriate explananda would, superficially at least, seem to put Wimsatt in the etiological camp (see Chapter 4), I think his theory is better understood as being largely neutral with regard to which explanatory projects are suitable. This is because he maintains that complexity is a necessary feature of a thing to which functional analysis is appropriate. However, he acknowledges that selection is not "logically necessary" in ascriptions of purpose (1972, p. 15), and may well not be sufficient (1972, p. 16). To remedy the latter problem, he suggests that complexity may be required as well.

---

[30] Wimsatt does not have a detailed analysis of how to understand "environment". Although he recognises that function can vary depending on environment, he offers little clue as to how to delimit relevant options. He mentions "normal" environments (p. 20), but does not give this matter the attention that later writers have.

It is the reference to purpose that allows an interpretation of his analysis to fit into the goal-directed approach outlined below. However, his own construal of his appeal to purpose is deflationary; in the case of biological functions, that something has a purpose can be interpreted, "*within the framework of evolutionary theory*", as meaning that something makes a "*contribution to the fitness of an evolutionary unit*" (1972, p. 8). Thus, many structural functions of biological entities will fail to be teleological functions in Wimsatt's sense.

An attribution of purpose is also understood, in part, to set the limits of functional inquiry. There is no need to situate the purpose in the context of yet another purpose. A purpose is that to which functions are relative; a purpose itself need not be functional relative to anything else.

Since current fitness and future selection are related, there is a sense in which Wimsatt's theory is selectionist. He maintains that the only systems to which attributions of teleological functions are *in fact* appropriate are those that are the result of a process of selection. This is because of his view of the sort of explanation in which appeals to teleological function may legitimately occur. He maintains that teleological explanations concern "the existence and form of [a functional] entity" (1972, p. 5). I shall return to this point in Chapter 4.

However, my claim that his view is not, strictly speaking, selectionist is supported both by the similarity of WW and RC, and by the fact that he reluctantly acknowledges (1972, p. 15) the logical possibility of the applicability of teleological explanations to unselected systems. He claims (footnote 33) that complexity may be the requisite feature.

Thus, even if some complex systems are not the result of selection, their parts may still be teleologically functional.

Finally, it is worth making a few more clarificatory remarks about Wimsatt's schema. He introduces the behaviour variable, *B*, so as to permit one item to have more than one function via its exhibiting two different sorts of behaviour. The theory variable, *T*, is intended to capture the requirement that a functional analysis be grounded in a theory which can provide a causal story of the interaction between the components of the system. Without such a causal underpinning (or the reasonable expectation of being able to discover one), a functional ascription is fanciful. This move is intended to rule out many Aristotelian examples of would-be functional relations. The variables *I*, *B*, and *S* should be understood to range over types, as opposed to tokens, given Wimsatt's explicit relativization of the schema to a causal theory.

At any rate, Wimsatt's approach is systems-theoretic, although his understanding of purpose in biology as contribution to fitness adds an element of selectionist theory.

§ 2.222 *Goal-directed Theories*

A sub-family of the systems approach is that of goal-directed theories. There are many variants within this family, but the central idea can be captured with the RC schema, given the proviso that the capacity *G* is to be understood as a *goal* of the system, not simply a larger capacity. There are a number of ways to cash out this talk of goals. Here, I discuss three theories, those of Boorse (1976), Adams (1979), and Schaffner (1993),

then briefly explain why Wimsatt (1972) does *not* properly fall into this category, despite the superficial appearance of doing so.

Boorse (1976) offers his account of functions in the context of a critique of Wright's (1973) account. He writes that "[f]unctions are, purely and simply, contributions to goals" (p. 77). His schema (hence "CB") is as follows:

> *A* is performing the function *F* in the *G*-ing of S at [time] t, *means*
>
> At [time] t, *A* is *F*-ing and the *F*-ing of *A* is making a causal contribution to the goal *G* of the goal-directed system *S*.
>
> (p. 80, some variables changed for consistency)

Boorse permits both type and token readings of this schema. He uses the progressive tense instead of using "performs", as he thinks that the latter implies that the function in question is performed many times, or over a significant period of time. Both these features of his account (allowing token readings and using the progressive tense) are linked to his view that a token item might perform a function only once and/or "by accident". (This issue is discussed further in § 3.3). A system is goal-directed if a component behaviour is directed to a goal. Following Sommerhoff (1950), Braithwaite (1953) and Nagel (1961), Boorse writes "[t]o say that an action or process *B* is directed to the goal *G* is to say not only that *B* is what is required for *G*, but also that within some range of environmental variation *B would have been modified* in whatever way was required for *G*" (p. 78, some variables changed for consistency, emphasis his). The appeal to goals is intended to delineate the set of items or behaviours with functions from

that of items or behaviours without a function; in other words, it is intended to distinguish functions from mere effects.

Boorse's analysis of what it is for a system to be goal-directed makes it a matter of the organism's "plasticity", to use Braithwaite's terminology. It is a matter of there being a variety of means of attaining the goal state, and the possibility of the persistence of the organism in exhibiting another of those means should the initial one(s) fail. Boorse's example is of a cat catching a bird; "capturing a bird may be the goal of a cat's behavior in so far as this behavior not only is appropriate for capturing a bird but would also have been appropriately modified if the bird had behaved differently" (pp. 78-9). Boorse then discusses two troublesome sorts of cases raised by Scheffler (1959)—i) cases in which the goal object is missing, and ii) cases in which there are many (possible) goals.[31] The first he handles by allowing the system to be goal-directed if its behaviour is appropriate in the relevant ways *had* the goal-object been there. The second, cases in which many end-states could be served by the behaviour in question, he handles in two ways: If there is "an internal mechanism which standardly guides pursuit of [a particular goal] but not the others" (p. 79), then the goal is determinate. If there is no such mechanism, then the attribution of a particular goal at a particular time is determined by contextual features.

A similar approach to an analysis of goals and goal-directed behaviour in terms of the existence of an internal mechanism is given by Adams(1979) and by Falk (1981).[32] Here is the schema of Adams' analysis of function (hence "FA"):

---

[31] These two problems amount to concerns about failure to perform function and indeterminacy of function.

[32] Both acknowledge their debt to a number of people, including Nagel (1961), and Rosenbleuth, Weiner, and Bigelow (1943).

A structure *a* has a function *F just in case*:

(1) *a* does *F* in system *S*;

(2) *F* causally contributes towards *S*'s outputting *G*
(*through the causal feedback mechanism*);

(3) *G* is (or itself contributes toward) a goal-state of *S*.

(p. 508, variables changed for consistency, his emphasis)

Adams offers three necessary conditions on a system's being goal-directed (and hence able to have goal-states) . It must have the following:

(1) an internal representation of the goal-state;

(2) a feedback system by which information about the system's state variables and its output values are fed back into the system as input values;

(3) a causal dependence between the information which is fed back into the system and the system's performance of successive operations which minimize the difference between the present state of the system and its goal-state. (p. 506)

This is a specification of the nature of Boorse's "internal mechanism", discussed above. Its precision, while allowing the analysis to rule out several ersatz functional ascriptions, may prove unduly restrictive.[33]

There are several things to note about this version (FA) of the goal-directed systems theory of functions. First, as with several other of the analyses discussed above,

[33] More seriously, this account, if used in the philosophy of mind, is subject to the criticism of circularity, as is CB. I discuss this issue in § 2.3 immediately below.

it provides only necessary, not sufficient, grounds for ascribing functions. Second, the goal-state must, in fact, be attained, given that the second part of the characterisation is not in the subjunctive. Third, there is no stipulation as to whether to read the variables of the formulations as referring to types or to tokens. Fourth, the problem of multiple end-states mentioned above in connection with Boorse is solved only by appeal to the systems having the appropriate (causal feedback) internal mechanism—without such a mechanism, there are no goals or goal-states, contextually defined or otherwise.

Kenneth Schaffner (1993) also falls under the goal-directed systems rubric. He offers his analysis as one that captures the sense of function relevant to explanation in biology and in medicine. Here is the schema he offers (hence, "KS"):

> To claim that *a* has a function in system *s* is to claim
>
> (1) System *S* has a goal *G* property,
> (2) *a*, via *B(a)*, results in or significantly promotes *G*.
>
> (1993, p. 404, some variables changed for consistency)

He is explicit about the nature of goals—they are properties that *we choose* to consider as goals. He is critical of the liberalism of Cummins' approach (see § 4.12 below), but is seemingly unaware of how closely connected his stipulative solution is to Cummins'. In Cummins' case, the larger capacity to be analyzed is a matter of our choosing; in Schaffner's case, the consequences we are to consider the goal are also of our choosing. Unlike Cummins' account, his analysis provides both necessary and sufficient conditions for functional ascription. The variables may range over tokens or

over types, depending upon an interpretation of "significantly promotes". Schaffner writes that the goal is to be promoted "either universally or in a significant probabilistic way" (1993, p. 404), leaving either interpretation open. However, the type interpretation is more natural, if "universally" is taken to mean across all tokens of a type, or if the "significant probabilistic way" is cashed out in terms of past instances of success. He maintains that "[t]o obtain the *primary biomedical science* sense [of function], we need to restrict the [goal property] to something like species (or deme or "subclan") survival" (ibid), which supports this interpretation. Instances of the goal property may or may not occur, granting, as seems reasonable, that "significantly promotes" does not require that the goal property always occur.

Finally, it is worth mentioning why Wimsatt's theory (WW) should not been seen to fit into the goal-directed category, despite its reference to purpose. As Wimsatt notes, complex systems fall under the rubric of many different sorts of theories. He introduces his purpose variable, in part, to *distinguish* his view from goal-directed theories.[34] What the variable is to range over depends on the theory being used. In the case of evolutionary theory, the purpose variable can be considered to range over a thing's "contribution to the fitness of an evolutionary unit" (1972, p. 8). Here, Wimsatt does *not* have in mind the sort of internal mechanism to which Boorse refers and that Adams tries to characterise, and which grounds their appeal to goals. Rather, in his (1972), he understands the relevant sort of complexity "roughly as the number of levels between the units of variation and heritability (in biology, the individual genes) and the units of selection (in

[34] His 'counter-example' to goal-directed theories involves an artefact. Whether such an appeal is unduly restricted will be discussed in § 3.4.

current evolutionary theory, most commonly the individual organism)" (footnote 33). Thus, in the case of adaptations at least, complexity and selection are conceptually linked.[35] However, the crucial feature for an attribution of complexity is an item's organizational structure. At any rate, these considerations leave purpose an open variable that can be filled in only by stipulating the relevant theory, where such a theory deals with complex systems. Since evolutionary theory is only one of many such theories, purpose remains an open variable. This explains why WW is not an essentially selectionist theory. Furthermore, unless all complex systems are goal-directed, then goals need not play a role in filling in the purpose variable.[36] Hence, WW is not an essentially goal-directed theory.

§ 2.3 *Narrowing the Field*

The previous section completed my initial presentation of a variety of analyses of teleological function. To summarise, I have distinguished two main camps, the selectionist and the systems-theoretic. The former is divided into the etiological and the propensity approaches, the first of which looks back to past selection and the second of which ahead to future selection. The latter, systems-theoretic accounts, can be divided into structural theories and goal-directed theories, both of which are ahistorical in focus. More details will emerge in the discussion that follows. However, we are already in a

[35] This is not, however, the only understanding of complexity, nor is it even Wimsatt's only understanding of it (cf. his 1974).

[36] Note that the goal-directed theorist need not claim that all complex systems are goal-directed—the analysis explicitly restricts attributions of functions to those (complex) systems that *are* goal-directed, whether or not all are.

position to narrow the field of analyses for two reasons. First, there is considerable overlap in the views presented, with some being sophisticated versions of other, earlier ones. Second, some of the views require the prior individuation of a representation, which would render them circular for the purposes of philosophy of mind.

Let me explain the latter consideration first. The goal-directed theories, CB, FA and KS, appeal to goals. For KS, goals are a matter of our own interests. For CB, they are either to be ascribed on the basis of context, or are a matter of the item in question having an "internal mechanism" which tracks the purported goal. For FA, this internal mechanism is further specified—goals require that there be a representation that is used in a causal feedback loop. For the purposes of using an appeal to teleological function to help to individuate representations, this last sort of account clearly will not suffice. It requires the existence of a representation in order for there to be a function. Even though Adams' notion of a representation is quite minimal, the circularity of this account in aiding in individuating representations is glaring. In order for a function to be ascribed, a representation must be ascribed. If this representation is determinate, then the account is circular. If it is not determinate, then something other than its function is what makes it determinate. Therefore, FA and, in its "internal mechanism" version, CB are not in the running, given current purposes.

This narrows the field slightly. It can be narrowed somewhat further by noting the relationships between some of the accounts. In the selectionist camp, I consider GS, RM, and PG to be refinements of the basic schema as sketched in LW. In the case of GS and PG, these refinements are made to allow the etiological selectionist account to avoid

some of the problems to which LW falls victim, as will be discussed in the next chapter. Millikan does not offer her account, RM, as a refinement to LW. Though, if I am right in viewing them both as selectionist etiological ones, there is no harm in considering it as such. However, it is important to keep in mind that she has a specific theoretical or explanatory project in mind. (This will be discussed further in Chapter 5.) The elaborations of GS and PG are specifically designed to meet the difficulties encountered by LW. GS is the more detailed of these two, and warrants (and hence will receive) careful attention. MB is an interesting case, as it mirrors LW but, in its attempt to avoid some of the difficulties encountered by LW by appealing to goodness, it raises troubling questions with regard to the legitimacy of postulating value in the natural world. Such an abandonment of naturalism should be a very last resort. Hence, I will not be treating MB in any detail.[37]

BP, on both a selectionist and a systems-theoretic reading, requires investigation. On a selectionist reading, it is sufficiently different from the other selectionist theories so as to encounter different problems and manifest different virtues. This is also the case, *mutatis mutandis*, on a systems-theoretic reading, with reference to the other systems-theoretic structural accounts. RC, the prototypical structural account, will require attention, but it should be kept in mind that Cummins did not offer his account as one that would capture the normativity inherent in biological discourse. WW, which predates RC, is largely superceded by it. Their *prima facie* difference, the appeal to purpose, is, as I suggested above in my discussion of WW, just that—*prima facie*.

---

[37] In Chapter 6, I do discuss some aspects of Bedau's treatment of views that oppose MB, as they are helpful in illuminating issues there.

As a result, in what is to come, I shall be focussing most closely on the following accounts: GS, RM, BP (in both its guises), RC, and CB. PG will be mentioned when it handles a problem differently from GS or RM. KS is subsumed by CB. WW is subsumed by RC. FA and LW will not be directly addressed, for the reasons just adduced. MB will be mentioned when its handling of a difficulty is different from that of the other selectionist etiological accounts, though it runs into trouble with regard to naturalism, as will be discussed.

Chapter 3

A Preliminary Evaluation of the Proposed Analyses: Handling Purported Desiderata

> But one, who is tormented he knows not why, with the apprehension of spectres in the dark, may, perhaps, be said to reason, and to reason naturally too: But then it must be in the same sense, that a malady is said to be natural; as arising from natural causes, tho' it be contrary to health, the most agreeable and most natural situation of man.
>
> Hume, *A Treatise of Human Nature*, Book I, Section IV.

In this chapter, I compare and contrast the different theories introduced in Chapter 2 to see how well they handle our intuitions with regard to that which appeals to teleological function are *prima facie* supposed to accommodate.[1] Specifically, I discuss i) the function/dysfunction distinction, ii) the function/vestige distinction, iii) the function/accident distinction, iv) univocality with regard to natural and artefactual traits or objects, and v) the constraints of naturalism. The issues of whether or not any acceptable theory of teleological functions *must* make all or any of distinctions i)– iii), or offer a univocal account of natural and artefactual traits, are deferred to Chapter 4. Issues

[1] I acknowledged the difficulties with such an approach at the beginning of Chapter 2. As Adams (1979) notes, this piecemeal approach seems to rely simply upon our intuitions as to when a functional attribution is appropriate and when it is not, and each modification of an analysis seems to leave it open to yet other counterexamples. I agree that intuition-mongering tends to get us nowhere. However, this approach has the virtue of forcing us to articulate said intuitions. The deficiencies of this piecemeal approach I hope to remedy by means of the next chapter on explanation.

about naturalism will resurface there as well, as will particular problem cases for each sort of analysis.

§ 3.1 *Function/Dysfunction*

It is generally considered a desideratum of any theory of functions that it be able to allow for some instances of failure to fulfil function. A view of function that entails that the dysfunctional heart of a victim of a congenital defect does not have the function of pumping blood simply because it cannot, is, *prima facie*, implausible.[2] To claim this would be to claim that either the "defective" heart isn't defective at all, or that the item in question is not a heart. Deciding on the criteria for the membership of a token in a functional type is a difficult issue which will be addressed below, but a requirement that a token be able to perform its function in order to be an instance of a functional type is too demanding, if the possibility of dysfunctional tokens is to be countenanced. In the case of the defective heart (having established that it is, indeed, a heart), we want to say that there is something that the token *should* be doing that it cannot. It is this distinction between function and dysfunction, with its background normativity, that, the reader will recall from Chapter 1, motivates the appeal to teleological functions in the philosophy of mind.

What is needed here is a story about a token's membership in a type that does not require that a token actually be able to perform the function in question. The idea behind

---

[2] Wimsatt writes, as part of an attack on Kauffman (1971) and the inadequacies of the perspectival (Cummins) sense of function, that "... to say that something is functional from a given perspective suggests that it ought to be possible also for something to be non-functional or disfunctional [sic] from that perspective" (1972, p. 37).

this is as follows: if it is possible that there are tokens that are 'aberrant' with regard to their ability to perform a particular function, but are nonetheless tokens of that functional type, then it is possible that there be dysfunctional tokens. The analyses offered in Chapter 2, if taken to range over types, provide necessary (or necessary and sufficient) conditions on a type's being functional. Unaddressed by many of those analyses is precisely the issue relevant here—that of when a token can be said to belong to a (functional) type.[3] If a response to this latter issue entails that a token of a functional type be itself functional, then the function/dysfunction distinction is obliterated.[4]

This distinction is a widely-accepted desideratum, though its range of applicability needs to be spelled out further; principled motivation for considering it a desideratum will be considered in Chapter 4. At this point, it is worth noting that, in order to draw the function/dysfunction distinction, it must be possible that there be tokens of functional types (functional tokens, for short) that do not or cannot perform their function, but it is not necessary that every functional token that does not or cannot perform its function be dysfunctional. If this were a consequence, the distinction would lose some of its intuitive force. We do not think that *every* heart that does not or cannot pump blood is dysfunctional. The heart of a dead person does not and cannot, in the circumstances, pump blood, but this, by itself, does not mean that this heart is dysfunctional.[5] After

---

[3] Faber (1986) offers a goal-directed analysis that ranges over tokens, and only derivatively, by induction, over types. He thinks this a virtue, "for how could one assign functions to a type of thing prior to discovering functions of individual tokens of the type?" (p. 93). This epistemological point does have some force, but his theory does not have the resources to accommodate explicitly a function/ dysfunction distinction.

[4] As I shall argue below, this is why analyses that are taken to range over tokens cannot draw the function/dysfunction distinction.

[5] I do not mean to imply that it is never legitimate to categorise an organ of a dead person as dysfunctional. My point is just that the fact that a person is dead does not make all their organs

some time, there are no circumstances in which the heart in a dead person can pump blood—too many cells will have died. Even in this case, an attribution of dysfunction is inappropriate.

Claiming that a thing is dysfunctional is claiming that it *should* be doing something that it cannot. Of course, this normativity is not restricted to dysfunctional tokens; a heart that can pump and is pumping blood, *should* be doing so. Functional tokens are *supposed to* perform their function; it is instances of dysfunction which bring this fact to our attention. There is, however, a continuum of circumstances along which the normativity shifts from the token item to its environment. At one end of the continuum is the sort of scenario in which a functional token cannot function and it is its *nature*, as opposed to its situation, that accounts for the inability. In such a case, the token is dysfunctional since it is in the sort of situation or environment in which it should function, yet it fails to do so. In these cases of dysfunctionality, there is something about the token itself, intrinsic to it, that results in its failure—it is to blame, it is dysfunctional. At the other end of the continuum is the sort of scenario in which the environment is utterly inimical to the functioning of the token item in question. The functional token cannot function, but this is for reasons extrinsic to it. It is in such cases as these that an ascription of dysfunction to the token is misleading. The blame, and hence (at least some of) the normativity, falls to the token's circumstances—the environment needs to be other than it is in order for the token to function.

---

dysfunctional. We do not transplant dysfunctional organs, but we sometimes transplant organs from a dead body.

To make the ends of the continuum intuitively clear, consider the following examples.[6] In the first case, we have a baby born with a hole between the two ventricular chambers of the heart. The result is, sadly, that this heart is a poor pumper of blood, in that it permits oxygenated and unoxygenated blood to commingle. This heart is defective. In a second case, we have a person whose heart pumps blood well. Pretend this person is me. Then, imagine that I am transported outside the Earth's atmosphere, and summarily removed from the mode of transportation. I would not last long—for one thing, my heart would very soon be unable to pump blood. In this sort of case, it is not the heart that is defective, even though it cannot perform its function. Rather, it is the environment in which the heart is found that bears the brunt of the normativity.

Most cases of a functional token's failure to function will fall on the continuum between these two sorts of extremes, and there is unlikely to be a hard and fast line between the sorts of cases in which an attribution of dysfunctionality would be appropriate and those in which it would not. However, I argue below that the notion of a token's environment is implicit in any attempt to accommodate the function/dysfunction distinction, and is intimately tied to the intrinsic/extrinsic distinction.

§ 3.11 *Selectionist Theories*

Most etiological theories (e.g., PG, RM) try to accommodate the function/ dysfunction distinction by defining functions in terms of types, and understanding type-membership to

[6] These examples are, of necessity at this point, sketches that beg important questions. Nonetheless, they will help focus discussion.

be established by considerations of a token's history, not its functionality. A type of feature *A* has as a function that effect for which instances of it were selected. A token *a* is a member of functional type *A* in virtue of being descended, in the right sort of way, from other tokens of type *A*.

Thus, a particular token organ has the function of, e.g., pumping blood, even when it cannot, because it is a token of a functional type, in this case, the heart. The token heart is a heart because it is a descendent of tokens that were selected because they pumped blood. This particular organ is a defective heart. It has a function, in virtue of being a member of a functional type, but cannot perform it. It is malfunctioning or dysfunctional.[7]

This way of handling the function/malfunction distinction is not immune to difficulties, which shall be discussed below. However, even if it were flawless, it is not available in any straightforward way to any theorist whose analysis is cast in terms of tokens. Consider the etiologist, Godfrey-Smith. Recall that, in order for a token to have the function it is purported to have, it must have a property or property cluster *C* "which *can do* F" (clause iii in GS, emphasis added). Despite this, what makes a token dysfunctional is precisely that it cannot do F.[8] GS does not appear to have the resources to accommodate dysfunctional tokens. This calls for a closer examination, which will involve looking again at the claim sketched at the beginning of this section (*viz.*, not all

---

[7] In this context, "malfunction" should be considered roughly synonymous with "dysfunction". I use "dysfunctional" in addition to "malfunctioning", since the latter seems to imply that the heart is *doing* something. Perhaps it isn't doing anything at all. It may still be dysfunctional.

[8] Actually, the situation is more complicated, in that something may be dysfunctional yet still be able to perform its function in some way and to some extent. Not everyone with a defective, i.e. dysfunctional, heart is dead. Total inability to function is the limiting case, but it must be accommodated if the function/dysfunction distinction is to be accommodated.

instances of failure of a token of a functional type to function are instances of dysfunction), and the attendant notion of a thing's environment.

To recapitulate, a token heart may not be able to pump blood for two broad sorts of reasons—one extrinsic to the heart, the other intrinsic. In an extreme extrinsic case, the heart is in a hostile environment (e.g., in a dead body, not in a body at all). In an extreme intrinsic case, there is something about the heart itself that explains its inability to pump blood—very roughly, that it is defective, damaged, or malformed. Furthermore, we take an ascription of function to be implicitly relativized to a certain sort of environment,[9] which goes some way toward explaining the intuition that a heart in a dead body isn't necessarily dysfunctional. In the extrinsic case, the *environment* should be different. A token heart in a dead body cannot function, even if there is nothing wrong with it, *per se*. My claim here is that perhaps it is not the case that GS need handle those cases in which the causes of failure to be able to perform function are extrinsic, since these are not cases of dysfunction.[10]

Now consider the intrinsic sort of case, in which the environment is somehow appropriate, yet a token heart cannot pump blood. Granted, it would be no mean task to determine what constitutes an environment being the appropriate one, as "appropriate" cannot mean simply that sort of environment in which the item in question *would* function, for then there would be no cases of failure to function that were intrinsic to the

[9] As will become apparent below, relativization to an environment is a feature of *all* the analyses of function introduced. Where they differ is in how they delimit the relevant environment.

[10] Even if one thinks that such extreme extrinsic cases *are* cases of dysfunction, the distinction between intrinsic and extrinsic should still make some sense, and intrinsic cases still count as cases of dysfunction. I am about to argue that GS cannot handle even the intrinsic cases.

item.[11] At any rate, in a case in which the cause of failure to perform function is due to intrinsic features of the item in question, that item is, presumably, malformed.[12] To a first approximation, "malformed" means different to a significant degree from other tokens of its type, whether morphologically, chemically, or in some other regard. More will be said below about the concept of malformity, but it is important to notice that any analysis of this notion should allow for malformed tokens that do not malfunction, either because they are not tokens of a functional type (hence, trivially, cannot be said to malfunction), or because the nature of their "deviance" from the "norm" for tokens of their type does not adversely affect their functioning.[13] Arguably, a malformed appendix does not malfunction (a vexed example, perhaps; see § 3.2), and a malformed tooth may still function just fine. The class of malformed things will be larger than the class of malfunctioning things. Thus "malformed" and "dysfunctional" are not equivalent. The latter is a more strongly evaluative notion. Nonetheless, in the case of dysfunctional tokens, there is something about the item, its malformation, loosely speaking, which underlies the malfunction.[14]

---

[11] I am skeptical that the notion of "appropriate" environment and, hence, the intrinsic/extrinsic distinction can be drawn completely independently of considerations of function. Nonetheless, this lurking question-begging does not worry me, as my goal is merely to highlight the fact that there is often a continuum and that our concerns at one end are different from those at the other. Functional ascriptions are most at home at the intrinsic end of the continuum, as we are focussed on the item more than on the environment.

[12] The category of malformed items is intended to be neutral with regard to the cause of the malformation. Hence, the category includes tokens which should perform their function, but which cannot do so, because of intrinsic properties, whether these properties are the result of 'internal' developmental events, or because of externally inflicted damage.

[13] The terms "deviance" and "norm" are in scare-quotes because I do not want to be committed to a statistical normality approach to this notion.

[14] The notion of malformity might seem as evaluative or normative as the notion of dysfunction, but I want to resist this reading. The idea is that, if one has two tokens of the type heart, and both are in the same sort of environment, yet one can pump blood and the other cannot, there must be

The important thing to note from the point of view of an evaluation of GS is this: if a token heart cannot pump blood due to its being malformed, it simply *does not have* the relevant property or property cluster *C*. If it *did* have *C*, then it would not be dysfunctional, given any plausible reading of the modality of "can do" in clause iii. Pointing out that the "can do" must be implicitly relativized to an environment will not help here. The environment against which to relativize the functional claim cannot be characterised as one in which *no* tokens would be able to perform, since then little sense could be made of the claim that there is a functional type of which they are tokens. Such standards would result in the intuitively inappropriate attribution of a multitude of functions—hearts might be seen to have the function of enabling their bearers to square the circle. Similarly, the environment cannot be characterised straightforwardly in terms of its suitability for the performance of a would-be function—otherwise, hearts might be seen as having the function of allowing their bearers to be concert pianists. Not only would such attributions be intuitively odd, they would, unless restricted further, entail that there would be no, or very few, cases of inability to function. However it is that the environment is characterised, it must be done in such a way that it is *possible* that some tokens function and some do not, if there is to be a function/ dysfunction distinction.

It might seem that a version of GS that is restricted to types would correct the problem.[15] Then it would be the *type* heart that would be said to have property cluster *C*

---

some internal difference between the two. If they are both in the "appropriate" environment, one is dysfunctional.

[15] LW and MB face a similar problem to GS insofar as their formulations are not explicitly restricted to types. Their use of the present tense would seem to commit them to no possibility of making the function/dysfunction distinction (a token that is not performing its purported function would not have that purported function). Restricting the analyses to types as opposed to tokens may circumvent this problem, if a suitable story can be told about type membership. They do not

which can do *F*, not individual tokens. (GS would need to be significantly revised to incorporate this change.) If it is possible that tokens be members of the type heart without having property cluster *C*, then instances of dysfunction are possible. But first—what is it to say that a type "can do" something? How are we to understand the modality of this claim?

This is the sort of question that PG and RM manage to avoid having to answer explicitly since they appeal to actual selection history and make no reference to current capacity. Modal notions appear to play no role. On these views, as mentioned above, the function of a given token is to *F* if and only if ancestral tokens of it were selected for doing *F*. A type of thing is functional if and only if instances of it have been selected for. A token's membership in a type is determined not by its capacities but by its history of descent.

So, a token reading of version GS of the etiological approach is sunk on the function/malfunction distinction. A type reading of GS is snagged (but not yet sunk), awaiting explication of its modal claim, while its less modally committed cousins sail past.[16]

Next in line for scrutiny is the selectionist version of the propensity theory schema (BP). Recall that it was not clear whether BP should be read with the variables ranging over types or over tokens. If it is tokens, then the same problem that befell a token

---

provide one, though perhaps they could tell the one that I about to tell and that is offered by other selectionist theories. The success or failure of a particular selectionist approach lies in the details, which I take theories such as GS, RM, and PG to be attempts to provide. MB encounters specific difficulties with regard to its comporting with naturalism—these will be discussed below.

[16] This modal commitment is not unmotivated in GS, as it goes some way toward helping GS handle the function/vestige distinction, § 3.2 below.

reading of GS strikes BP, at least on a literal reading of "performs". A token heart that does not perform, that does not pump blood, simply does not confer a survival-enhancing propensity on its bearer. According to BP, performing the function in question is a necessary condition for the thing to have a function. Ergo, a non-pumping heart has no function, and the function/dysfunction distinction is not made.

However, there is another way to read BP. Suppose that "perform" is somehow conditionalised and means "*would* perform". Bigelow and Pargetter explicitly state that their theory of functions "must be relativized to an environment" (p. 192), viz, the "usual [or]... natural habitat" of the creature that bears the trait (or of the trait itself). They recognize that there are difficulties in delineating what constitutes a thing's natural habitat, but let us grant that this can be done without generating any problems fatal to the theory.[17] Could this yield a subjunctivized reading of "perform" that would permit a token reading of BP to accommodate the function/dysfunction distinction? The idea is this: a token heart has the function of pumping blood (even when it cannot perform this function) just in case the token *would* perform its function (thereby conferring a survival-enhancing propensity on the organism that houses it) *if* it were in its natural habitat. Unfortunately, this modal move fails in the same way GS does; it would ameliorate the problem *only* in cases in which the failing-to-perform token is not in its natural

[17] I think they underestimate the difficulties. They write "[t]here may be room for disagreement about what counts as a creature's 'natural habitat'; but this sort of variable parameter is a common feature of many useful scientific concepts" (p. 192). This hand-waving at the notion of environment is indicative of a failure to realise the centrality of the notion to an understanding of function.

environment *and* when this is why it fails to function.[18] If it *were* in its natural environment, it *would* perform its function. According to this reading of BP, an instance of intrinsically-caused failure to function would not be a case of dysfunction, since the token would not *have* the function in question and, again, the function/dysfunction distinction is not made.

Thus, a subjunctivized or relativized reading of "perform" will not help on a token reading of BP, as was the case for GS's "can do". How, then, might we read BP, if taken to range over types?

Here is an attempt: the function of the type heart is to pump blood only if the pumping of blood by the heart confers a survival-enhancing propensity on the type of system of which the heart is a part. Exactly why a thing's conferring a survival-enhancing propensity on that of which it is a part is to be a necessary condition for its having a function is not clear. Even so, this requirement is no less plausible than the requirement that the feature have been selected for. Perhaps one (or both) of these ways of understanding a trait to be functional are legitimate, but there must, in addition, be a way to assign tokens to the type. This is the crucial issue that must be addressed if a theory is to be able to draw the function/dysfunction distinction. As noted above, theories such as RM and PG do so in terms of descent; this route may also be open to BP (and a reconstructed-to-range-over-types version of GS).[19] This is an appropriate point at which

---

[18] As noted above, this sort of case is at the end of the extrinsic/intrinsic continuum at which ascriptions of dysfunction seem misleading. It is a case of extrinsically caused failure to function. There is nothing wrong with the item, *per se*— rather, the environment is to blame.

[19] Some etiologists (e.g., Millikan (1987), Neander (1991b)) have claimed that BP, in being a self-proclaimed "forward-looking" theory, cannot legitimately make any appeal to historical facts. I do not see the force of this objection, since it is the function itself that is alleged to be forward-looking, not the individuation of environments.

to have a closer look at this option, which will require an extended bit of Millikan exegesis.

Etiological theories such as RM and PG agree that a token, if it has a particular function, has as that function the effect for which its ancestors were selected. Millikan writes that "[h]aving a proper function depends upon the *history* of the device that has it, not upon its form of dispositions" (1984, p.29). A token's ancestry bequeaths membership in a functional type. Although an intuitively straightforward thesis, rather a lot is built in to the notions of ancestry and descent here. Millikan offers, as part of a sufficient condition for a thing, *x*, to be a member of a higher-order reproductively established family *T*, the following: *x* is "*in some respects like* Normal[20] members of *T* because... it has been produced in accordance with an explanation that approximates in some (undefined) degree to the Normal explanation for production of members of *T*" (p. 25, italics mine, variable changed for consistency here and in subsequent quotations). A first-order reproductively established family is "any set of entities having the same or similar reproductively established characters derived by repetitive reproductions from the same character of the same model or models" (1984, p. 23). Millikan's example is that of the set of tokens of the written word "dog", variously reproduced by printing presses, people's hands, and photocopiers. A higher-order reproductively established family is a set of entities sharing a characteristic because of *indirect* reproduction. Thus, hearts do not form a first-order reproductively established family because no token is copied directly from another, but they do form a higher-order reproductively established family

[20] A capitalised "normal" , i.e., "Normal", indicates that the term is being used in its technical, Millikan, sense; it is not to be used to mean, e.g., statistically usual. I explain this concept in more detail below.

because each heart was "produced under Normal conditions in accordance with the proper functions of certain of [its containing organism's] genes which *were* directly copied from [the containing organism's] parents' genes" (1984, p. 25). Normal conditions are those to which one adverts when giving a Normal explanation. A Normal explanation, in turn, is "an explanation of how a particular reproductively established family has historically performed a particular proper function" (1984, p. 33). Such explanations come in various degrees of "proximity", the most proximate being "the least detailed explanation possible that starts by noting some features of the structure of members of *T*, adds some conditions in which *T* has historically been when it actually performed *F*—these conditions being uniform over as large a number of historical cases as possible—adds natural laws, and ... shows ... how this setup leads to the performance of *F*" (1984, p. 33).

At this point, I have a quibble. Millikan mentions features of the structure of members or tokens, and some conditions that obtained for the family or type. However, it does not make sense to speak of conditions in which the type heart has been historically, though it does make sense to speak of conditions in which tokens of the type heart have been. Taking Millikan literally, in the case of the heart, a Normal explanation would note some structural features of hearts, some of the conditions in which the family of hearts has been when it pumped blood, and some causal information, and hence explain how it is that the family of hearts pumps blood. However, it is not, of course, the family itself that pumped or pumps blood, but members of the family that did or do. Incorporating this insight into Millikan's schema results in an understanding of the most proximate explanation as being "the least detailed explanation possible that starts by noting some

features of the structure of members of *T*, adds some conditions in which [members of] *T* [have] historically been when [they] actually performed *F*—these conditions being uniform over as large a number of historical cases as possible—adds natural laws, and ... shows ... how this setup leads to the performance of *F*". Less proximate explanations provide more specific details about the structure and historical environment of hearts, though such details need only, according to Millikan, to be "usually" (1984, p. 33) true of historical cases. Less proximate explanations apply more restrictedly, i.e., to fewer members of the family of hearts, because they are more detailed.

This careful reading of Millikan makes it clear that there is an appeal to the statistical sense of "normal" built into the explication of what conditions are to constitute Normal conditions. They are those that *usually* obtained when a token performed its function. Millikan calls a Normal explanation "a preponderant explanation" (1984, p. 34), and I take it that this is why. Of course, it is important to distinguish this appeal to statistical normality from one that Millikan clearly does not make. Normal conditions need not be those that are usual *simpliciter*. In the case of sperm, Normal conditions for performing the function of penetrating an ovum are very rare for sperm, so, in that sense, are unusual or not normal. They must, however, have been usual in the cases of successful performance.

With all this technical machinery at hand, let us return to the issue of what makes a token item a member of the type heart, on a selectionist schema such as RM or PG. I said above that this was a matter of being descended in the right sort of way from the right sort of things, but this is not the entire story. For Millikan, being a token of a

functional type is a matter of being a member of a reproductively–established family. Being a member of a reproductively–established family is, in turn, a matter of having been produced by something, the function of which is to produce members of that family that are "*in some respects like* Normal members, [and to be like those members because of having] been produced in accordance with an explanation that approximates in some (undefined) degree to a Normal explanation for production of members" of the family. (1984, p. 25, italics mine). That a token be "in some respects like" Normal members is an ahistorical requirement. The last conjunct ensures that what accounts for this ahistorical similarity is what (to some undefined degree) has accounted for the similarity in the past, and hence that the similarity is not a fluke.[21]

For our purposes here, it is important to note that both causal history *and relevant similarity* are pertinent in establishing that a given token is indeed a member of a reproductively-established family. On Millikan's theory, relevant similarity must come about because of the method of production of the token—there is an explicit link between similarity and causal history. Relevant similarity without appropriate causal history would not be sufficient to render a token a member of a family, but neither would appropriate causal history unless it resulted in relevant similarity. This similarity need not, of course, be identity, since this would preclude the possibility of some tokens being dysfunctional while others are not. Two tokens, if qualitatively identical,[22] would behave identically in the same circumstances; the function/dysfunction distinction requires that it

[21] The motivation for this restriction will become more apparent in Chapter 4.

[22] I do not mean to raise age-old questions about the individuation of identicals. My point is merely that, in order for dysfunction to be possible in a constant environment, it must be possible for tokens of a type to vary to some degree, yet still all be tokens of the same type.

be possible that there be a pair of tokens, only one of which is dysfunctional. That is, the distinction requires the possibility of cases toward the extreme intrinsic end of the extrinsic/intrinsic continuum and, hence, requires a notion of similarity that is not identity.

The causal history component of the claim that there must be "an (undefined) degree" of approximation to a Normal explanation for the production of the token does not present quite the same problem when it comes to identity. It amounts to the requirement that the token have been produced in a similar way and in similar circumstances to those in which its ancestors were produced. Two tokens could be produced in *exactly* the same way that their common ancestors were produced and under *exactly* the same conditions, yet be different from their ancestors and each other (and, hence, possibly be dysfunctional), provided that the method of production was not completely faithful. For instance, a method of reproduction could be 90% accurate.[23]

Whatever the method and conditions of production, relevant similarity of the token to Normal members of the type is required. In the case of an organ such as the heart, this similarity is presumably morphological.[24] A token must currently resemble other members of its (would-be) functional type in certain structural ways in order to be a member of this type.

A sensitivity on the part of etiological theories to morphological considerations in assigning a token to a type should not be surprising, as structural features are one of the components in a Normal explanation. Besides, it is intuitively well-motivated. Suppose

---

[23] The notion of similarity of method of production does encounter problems. See below in this section.

[24] It obviously cannot be *functional* similarity, else there would be no dysfunctional tokens.

the incidence of congenitally defective human kidneys increases sharply. For some reason (not natural selection, presumably), it ends up that many babies born have tokens of this type of impaired "kidney". Let's suppose further that the defect is morphologically dramatic—impaired "kidneys" bear little resemblance to unimpaired ones. It is far from inconceivable that, eventually, we distinguish, and have good grounds for distinguishing, two traits here, instead of one. "I'm sorry, new parents", says the doctor, "but your baby was born with a schmidney instead of a kidney."[25]

Another example that is perhaps more persuasive does not involve distinguishing two different types of traits. A token item may be so ill-formed that, even if situated where a token of a functional type would be, it cannot reasonably be said to be a functional token. That there is vagueness "in some cases, [as to] whether a bit of matter should be called 'a malformed eye' or merely 'a glob of misplaced organic matter on the forehead'" (Millikan, 1984, p. 25) doesn't entail that there are no clearcut cases. Millikan would concur, as evidenced by her requirement for similarity.

One might think that the etiological theory is able to accommodate the intuition that, say, a token item could fail to be an eye even if situated in the usual place in a creature that is of a type that ordinarily has eyes, without appealing to morphological considerations at all. Instead, it would seem to be sufficient to appeal to historical facts; if the item were not a "descendent", in the relevant sense, of instances of a functional type, then it would not be of that type. In this scenario, there may well be, even must be, a "glitch" in reproduction, or conditions that are insufficiently similar to Normal

[25] This, of course, is not to say that there is nothing wrong with infants born with schmidneys. There is—they die. What is wrong is that they fail to have as a component a token of a crucial functional type, not that they have a dysfunctional token of a functional type.

conditions. Here is where we need a story about what it is for a method of production to be similar "to an (undefined) degree" to the (or a) Normal method, and for conditions to be similar "to an (undefined) degree" to Normal conditions.

Normal methods were those that were successful in producing ancestral tokens of the functional type, and Normal conditions were those that obtained when functional tokens performed their function. It is not a requirement that Normal methods be those that have never produced a dysfunctional token, in part because their doing so now would seem inexplicable if the requirement set such high standards. Millikan has stipulated that a *sufficiently* different method of production would produce a token that would not be a member of the family in question, even if it were structurally identical to some token that is a member of the family. Thus, an entity formed "by some cosmic accident [from] a collection of molecules formerly in random motion" (Millikan, 1984, p. 93) and that is physically identical to me, say, would have no functional parts.[26] This method of production is, intuitively enough, sufficiently different from the Normal one, so as to block attributions of function, given an analysis which appeals to history. If it is historical considerations alone, not conjoined with morphological considerations, that are to distinguish between something's being a malformed eye and something's being a glob, then there must be good reason to think that the two things, one of a functional type and one not, could not be produced in *sufficiently similar* ways.

Stating that there must have been a "glitch" will not answer the question, as "glitch" is an evaluative notion. It will not suffice to delineate sufficiently similar

[26] Davidson's (1987) Swampman is, of course, just such an entity. He will return in Chapter 6.

methods of production in terms of that which is produced, if the degree of similarity of method of production were to obviate the need to appeal to morphological considerations. One could appeal to statistically unusual methods of production, but this will not distinguish between malformed eyes and globs, even though they would differ in the methods by which they were produced, if both occur equally often.

This is not a criticism of Millikan's position *per se*. On the contrary, she is explicit in requiring both historical and ahistorical similarity. My point here is that an attempt to circumvent the ahistorical similarity requirement by reducing it to the historical similarity requirement will not work without a way of characterising sufficiently dissimilar conditions and methods of production without individuating them in terms of their output. Appeals to statistical normality are not adequate unless sufficiently different is sufficiently rare, and there is no reason to think that this need be the case.

So, an etiological selectionist approach handles the function/dysfunction distinction by assigning tokens to a functional type in virtue of both morphological similarity and appropriate causal history, where the latter includes similarity to both Normal conditions and Normal method of production. I mentioned above that BP could make use of similar considerations. Before returning to that point, I wish to examine briefly a different way in which BP might be thought to handle the function/dysfunction distinction.

One might think that BP, in either its systems-theoretic or selectionist guise, can handle the problem of dysfunction by appeal to the notion of propensity that is built right

in.[27] Functions are "dispositional in nature" (1987, p. 193), and the failure of a particular occurrence of a trait to enhance survival does not tell against the propensity of such a trait. Actual "frequencies and statistically normal outcomes will be important evidence for the requisite propensities" (1987, footnote 9), but frequencies and propensities do not always coincide.

When applied to a token, this clearly will not do. A token that cannot function does not have a propensity to confer a survival-enhancing propensity on the organism or system of which it is a functional part. These are the grounds on which Neander objects to BP's handling of dysfunction: "The impaired kidney does not have a disposition that is apt for selection, or a disposition to enhance survival systematically, so according to the propensity theory it does not have a proper function" (1991a, p. 183).

Restricting our analysis to types, as opposed to tokens, is, as I have argued, a necessary first step to accommodating the function/dysfunction distinction. BP falls into both the selectionist and systems-theoretic camps, but the story of how to assign tokens to a type need not differ. At one point, Neander considers it a failing of non-selectionist approaches that the type/token distinction is not built right in. She thinks that the appeal to this distinction in the case of biological function is *ad hoc*, on the grounds that it is introduced "to ward off a plethora of counterexamples" (1991a, footnote 11). On my understanding, both etiological and non-selectionist approaches necessarily invoke the type/token distinction and need to do so for the same reason—we want to know about the

[27] Perhaps Bigelow and Pargetter *do* think this, though I doubt it. The discussion cited immediately below occurs in the context of an explication of the subjunctive "forward-looking" nature of their theory and its role in explanation, not in an explication of dysfunctionality or of how tokens are assigned to types.

properties of a type of thing or, if of a token, of a token *qua* member of a type. That appeals to selection also, in and of themselves, invoke the type/token distinction is not evidence of the theoretical superiority of the etiological approach. (Bigelow and Pargetter's failure to be clear about whether their analysis should be read as ranging over types or over token may have contributed to Neander's seeing the distinction invoked in discussing their theory as *ad hoc*.)

Regardless, let us take BP to range over types. Thus, a functional trait is one that confers a survival-enhancing propensity on its bearers.[28] The implicit modality of "propensity" requires an analysis, and snags (but does not sink) BP on the same shoals that have GS snagged.[29] Construing BP as ranging over types tells us nothing, of course, about how to assign tokens to a type. The etiological avenue outlined above (requiring ahistorical and historical similarity) is open to BP, without pressing BP into an etiological mold. I see no reason to demand that an ahistorical theory of function be ahistorical in every respect. Assigning tokens to the type may well require an appeal to historical considerations without thereby turning the theory into an etiological one such as RM or PG.

Godfrey-Smith, though not a proponent of the propensity theory, notes that "it does seem likely that the propensity approach can be developed in a coherent way, at the

---

[28] The notion of "bearers" needs to be sufficiently loose so as to accommodate considerations of inclusive fitness and thereby accommodate "altruistic", or individually destructive but group-preserving, behaviour.

[29] Just as was the case for GS, the modal requirement is not unmotivated. See § 3.2 below.

price of narrowing the gap between it and the historical [etiological] view" (1994, p. 353). This way of developing the theory retains an important feature of its ahistoricity.[30]

At any rate, an appeal to a combination of historical and structural considerations in assigning tokens to types is available both to etiological theories and to BP. Note that these two requirements can be teased apart, though the results may not be satisfying. Using only the historical considerations may deprive us of the tools for distinguishing between Millikan's malformed eye and glob.

§ 3.12 *Systems-Theories*

Let us turn to how a systems-theoretic approach (other than BP) might accommodate the function/dysfunction distinction. Consider first RC. Not surprisingly, a token reading of this schema will not accommodate the function/dysfunction distinction for the same reasons adduced above for selectionist theories. Suppose RC ranges over types (as it was understood to do). Would this permit RC to make the distinction? As above, the question revolves around what it takes for a token to be a member of a functional type. Although Cummins is not explicit in this regard, I suspect that he considers that it is actual capacity to perform the function in question that is required and that, consequently, the function/dysfunction distinction cannot be made.[31]

---

[30] Godfrey-Smith explicitly rejects such a "mixed theory" (1994, p. 353), on the grounds that it fails in its explanatory task. I return to his view in the next sub-section (§ 3.12), and discuss the problem of explanatory task in Chapter 4.

[31] This is not surprising, given Cummins' explanatory task. See Chapter 4.

It might be tempting for a systems–theorist to try to accommodate the distinction by appeal to statistical normality. Wimsatt states that "certainly most and probably all functional relationships involve relatively high frequencies of performances of the functional consequence and of the efficacy of that performance in promoting purpose attainment" (1976, p. 50). If this idea is taken as a means of explicating the notion of dysfunctionality, then, using our example, it is because (in part) most hearts achieve the "functional consequence" of pumping blood that, even if this particular one does not, we can conclude that the function of even this heart is to pump blood.

This is how Neander (1991a) reads Wimsatt's claim; she sees the appeal to statistical normality as an attempt to accommodate the function/dysfunction distinction. She argues against the statistical normality approach by noting that "dysfunction can become widespread within a population, through epidemics or major environmental disasters" (p. 182). Were this to happen, the systems–theorist seems forced to say that a token that was once dysfunctional is no longer so, since it is now statistically normal. Neander states that this means that "if enough of us are stricken with disease (roughly, are dysfunctional) we cease to be diseased, which is nonsense" (1991a, p. 182).

I think it clear that she is right, but equally clear that this cannot be the systems-theoretic answer to the question of how to draw the function/dysfunction distinction. For that task, as I argued above, an explication of how to assign a token to a functional type is required. It may be that most hearts pump blood, and it may even be necessary that at least some do, in order for there to be a functional trait,[32] but clearly this sort of statistical

[32] This will be true if the above kidney/schmidney example captures something about how we individuate types. The point made there would still hold, even if "schmidney" were not a type, since sufficient morphological disparity could bar a token from being a member of a type.

fact is not germane at this juncture. In fact, I doubt it is taken to be germane by the systems-theorist even to the issue of ascertaining what makes a type functional. There is no reason to think that somebody such as Wimsatt is claiming that the function of a type is whatever it is that most of its tokens do. A glance at WW should be sufficient to convince us otherwise.

At any rate, how could statistical facts bear on explaining how a token could have a function that it cannot perform? My liver cannot pump blood, and does not have the function of pumping blood, even if most hearts can pump blood. What most tokens of a functional type do is irrelevant to assigning a token "thing" to a functional type.

Few, if any, of the systems-theorists are explicit about how to assign tokens to types, so my exposition here is brief. However, as in the case of the selectionist theories above, the obvious candidates are morphological and historical similarities. These may not exhaust the options—environmental similarity, non-historically defined, may also be relevant, depending upon the explanatory project.[33] Nonetheless, they seem plausible enough and are available to all the analyses discussed, provided the analyses are taken to range over types, not tokens.

So, by way of brief summary, we have seen that the function/dysfunction distinction requires that there be a way to assign tokens to types other than in virtue of their capacity to function, and that appeals to morphology and etiology may fit the bill and are, *prima facie*, open to all the analyses, though the details remain to be explicated. Also, questions about how it is that functionality is assigned at all remain, but things

[33] See Chapter 4 for further discussion.

should become clearer as we investigate other distinctions considered to be desiderata in this chapter, and turn to issues of explanation in the next.

§ 3. 2 *Function/Vestige*

Many things fail to perform any function, yet this need not mark them as dysfunctional. A case of note is that of vestigial traits; we must account for traits that once had a function but no longer do, without evaluating them as deficient.[34] Perhaps the notion of vestigiality is encountered only with reference to biological items; however, unless we think our analysis of functions applicable only *elsewhere* than in the restricted domain of biology, then it had best accommodate vestigiality.[35]

Consider the human appendix. Our current physiological theory offers the conclusion that it has no function. Our practice lends support—whenever our appendices give us enough trouble, we simply remove them, without worrying about replacing them. Our current evolutionary theory offers the further conclusion that the appendix is vestigial. That is, at one time, it had a function but it no longer does.[36]

Vestigiality is usually attributed to traits, and only derivatively to tokens. This may be because there are few, if any, cases of traits that it is reasonable to think vestigial

---

[34] A trait can be vestigial with regard to one particular function, yet not another. Vestigial traits are vestigial *simpliciter* only if there is no function they now have. The term "vestigial" without modifiers should be taken to mean vestigial *simpliciter*.

[35] I discuss below, § 3.4, the question of the applicability of our analysis to artefacts, and there address the issue of the desirability of univocality.

[36] Apparently, the appendix once had the function of helping its bearers to digest cellulose.

that have some tokens that are "dysvestigial", so to speak.[37] I challenge this custom below, in § 3.22, arguing that such a practice is unduly restrictive. In any case, the task central to accommodating the function/dysfunction distinction, that of assigning tokens to types, does not play as key a role in elucidating vestigiality. Rather, we require a story of the attribution of function such that an item may at one time be a functional item, yet at a later time not. This could range over types or tokens, if the right comparison class can be eludicated.[38] Vestigiality clearly has a diachronic element. Time must have passed, and something must have changed, for an item to be vestigial. Regardless of which theory of function one holds, one must accept that, in order for an item to be vestigial, it must have at one time been functional, but now it is not. With this change in functional status comes an attendant change in normative status. Appendices were once supposed to do something. Now, there is nothing they are supposed to do. There is no task that my (or anyone's) appendix is supposed to be performing.

In order to be a vestige, the item's change in functional status, however mediated, cannot have affected the organism so adversely that the existence of the organism is seriously compromised. If something changes so much that it no longer performs the function it once did, *and* this function is crucial to the continued existence of the organism of which it is a part and there is no other trait that can play the role, then an attribution of vestigiality is inappropriate. A vestigial trait's not performing its erstwhile

---

[37] Another reason may be that tokens tend not to last long enough for the relevant long-term changes that are required for attributions of vestigiality to occur. If a token organism lived for eons, it might be that some of its token parts would become vestigial, as it or its environment changed.

[38] Here, the type/token consideration may again be relevant, assuming that tokens do not last long enough for the changes necessary to occur. (See previous note.) In such a case, it is *past* tokens of the type that had the function, so type membership is important.

function is benign; hence, the loss of normative status. If the change is not benign in this way, and is toward the intrinsic end of the scale, then tokens of it are dysfunctional, not vestigial. If it is toward the extrinsic end of the scale, then tokens of it are neither dysfunctional nor vestigial—they are unlucky.

At any rate, vestigial traits once had a function that they no longer have. An account of functions ought, *prima facie*, to make sense of this notion.

§3.21 *Selectionist Theories*

Selectionist etiological theories attribute functions on the basis of selection history. The basic idea, recall, is that, in order to have a function, a trait must have been selected for. Traits that were not selected for do not have a function. This analysis seems straightforwardly applicable in the case of a vestigial trait, despite Bigelow and Pargetter's claim that the etiologist is saddled with maintaining that a trait's function now is "whatever it used to be, and was evolved for" (p.196). The etiologist can claim that the trait is here now because of selection of, not selection for, though there was a time during which the trait was being selected for.[39] We have appendices now only because they came along for the ride, as it were, but there was a time when we had them because they being selected for. In the case of vestigial traits, selection, though it did favour the trait at some point, has not done so *in the recent past*, to borrow a phrase from GS.[40]

---

[39] For explication and discussion of the selection of/selection for distinction, see Sober (1984) and § 4.111 below.

[40] This move is also made by Millikan (e.g., 1989b, p. 173). Her approach (RM) is sufficiently similar in this regard to GS so as not to warrant a separate discussion here.

Focussing on a restricted timeframe does allow for the possibility of vestigiality; it allows for the possibility of something that was once selected for, ceasing to have a function by ceasing to be selected for. However, details need to be worked out.

Is being selected for, then subsequently not being selected for, sufficient for something to be vestigial? To answer this question, we must understand better what it is for something to be selected for. Selection operates on relative adaptiveness. If trait X is better at doing *f* than is trait Y, and is less costly, then, in the absence of interfering factors, selection will favour trait X. Trait Y is being selected against, and, in the absence of interfering factors, it will eventually disappear. Does such a scenario license an attribute of vestigiality to trait Y?

The issue is this: in what circumstances can failure to be selected for result in a vestigial trait? After all, it would seem our hypothetical schmidneys above (§ 3.11) are not vestigial. What else is required? If a trait has not been selected for in the recent past, does it lose its functional status?[41]

If the sketch offered at the beginning of this section is correct, one of the requirements for vestigiality is that the feature not benefit the organism that has it. This is not a notion that need be relativized to other extant alternative traits, in contrast to the notion of selection for. Trait Y may benefit organisms that have it, but not benefit them as much as trait X would. If it does benefit them, then, prima facie, it is not vestigial. Is an appeal to selection sufficient to allow this for this possibility? Let us look at the details of various etiological attempts to accommodate vestigiality.

---

[41] I argue, in Chapter 4, that such a consequence is unacceptable, given a certain explanatory project.

GS, as just noted, has an explicit reference to the timeframe in which selection must have been favouring the trait. However, there is another component to his accommodation of vestigiality. He stipulates that, if the trait under consideration is to be considered a functional one, it must now be able to perform its purported function. The environment must be one in which the trait "continues to be selected over", to use Godfrey-Smith's terminology (1994, p. 353). This version of the etiological approach results in it coming close to overlapping with the propensity approach, since it requires that, in addition to being a case of having been selected for (clause v), the trait must also be able to do what it was once selected for doing (clause iii). This is part of what enables Godfrey-Smith to draw the function/vestige distinction. A trait that no longer "can", in some sense, perform a function it once could, may indeed be vestigial. The concept of vestigiality seems to require both that the item is not of benefit (by Godfrey-Smith's lights, that it is not being selected for), and that it can no longer do what functionally (selectionally, for GS) relevant thing it once did.[42]

Now, however, there is an important difficulty. An explication is required of what it is for a trait to be able currently to perform its function. Godfrey-Smith is not explicit here, though he offers a suggestive remark: "...there may be good reason to require that the trait still be able to do now what it was selected for doing, but we should not require that the trait also have the same propensity to succeed under selection that it has had in the past" (1994, p. 359). I take this to mean that the trait must have *some* propensity to succeed under selection. This combines an appeal to selection with what, intuitively,

[42] There is a lingering ambiguity here between token and type. It is not clear from Godfrey-Smith's presentation which is intended, but a type reading is most reasonable.

distinguishes the vestigial from the functional: a vestigial trait is one that not only was once useful and is no longer, it also cannot do what useful thing it once did. However, the etiological approach makes much of being able to understand "was useful" as "was selected for", as this is simply a matter of causal history, and modal commitments are minimal. Note that Godfrey-Smith is attempting to understand the trait under consideration as one *currently* prone to succeed under selection. There is no simple causal/historical analysis of this sort of claim. In fact, to say that a trait has a propensity now to succeed under selection is to be committed to *the very same* "questionable modal commitments about relevant ranges of alternatives and circumstances of selection"(p. 353) that Godfrey-Smith had earlier faulted Bigelow and Pargetter for relying upon.[43]

One avenue partway out of this difficulty is to deny that a vestigial trait need be incapable of performing its former function. This way, we can keep our modal commitments to a minimum. Perhaps our appendices really could process cellulose, but we simply do not ingest grass any more. If this were true, is there any reason to think that the status of the appendix as vestigial would be undermined? Presumably not. There seems no reason to think they *should* be able to process cellulose, no legitimate explanatory project in which to make appeal to such a fact. (See Chapter 4 for expansion on this claim.) A verdict on this ahistorical consideration, this modal worry, might well be otiose in establishing a trait's vestigiality.

---

[43] Furthermore, current propensities do not explain why a trait exists as it does, which is, according to Godfrey-Smith, the sort of explanandum at issue (pp. 347, 353). More on this in Chapter 4.

Not so, however, with the requirement that a trait's failure to perform its erstwhile function not adversely affect the success of the organism of which it is a part. This consideration is independent of whether or not the trait has recently been selected for, even if one construes the past as the *very* recent past, and is a requirement for a trait to be vestigial, as opposed to dysfunctional or non-functional. The success in question is *present* success. A trait is not vestigial unless it *currently* has no role to play.

Let us look at another etiological account of vestiges. Griffiths writes: "[i]n standard presentations, two things are said to be characteristic of vestiges. They exhibit loss or decline of function and they frequently exhibit atrophy" (1992, p. 126). Both these facts, he claims, are of potential epistemic relevance in establishing that the trait in question is a vestige. However, he then goes on to acknowledge that claiming that a trait has atrophied is claiming that changes in its size or shape are *due to* its loss of function. Mere change in size or shape, even if to smaller or simpler, is not, by itself, evidence of atrophy. Atrophy itself is an importantly functional-cum-historical notion—the change in form is somehow *due to* the trait's loss of usefulness. Furthermore, not all traits that cease to have a function also exhibit a change in form. This need not be criterial, as Griffiths recognises. Perhaps there has been insufficient variation to permit the favouring of smaller or simpler variants, or perhaps the trait in question continues performing another function (or acquires a new one) without changing its structure. Similarly, a decline in the incidence of a trait's presence in a population may or may not occur, given its loss of function, since such a decline would require not only that there be tokens of the type of creature that initially had the trait that no longer have the trait (or that have a

variant of it), but also that, if the decline is an instance amenable to "selective explanation", those variant tokens become more prevalent *because* of their superiority The would-be vestigial organ is not such, given this analysis, unless it is being selected *against*.

This is to tie two issues uncomfortably close together. Selection is largely a matter of relative fitness—if function is tied to selection, then function is lost when something else is *better* at doing what the item in question does.[44] Furthermore, this view would seem to entail that function would be lost if there is *nothing* to do better than, i.e., if there are no competing variants.

Griffiths is aware of this problem, and attempts to address it by invoking the notion of a proximal selective explanation. This sort of explanation appeals to the concept of "an evolutionarily significant time period" which is, for a trait T, the amount of time "such that, given the mutation rate at the loci controlling T, and the population size, we would expect sufficient variants for T to have occurred to allow significant regressive evolution if the trait was making no contribution to fitness" (p. 129). Regressive evolution is the atrophying of a trait (p. 126).

PG, like GS, makes significant modal claims. No longer is actual causal history the deciding factor in distinguishing vestigial from functional traits—what matters is what the causal history *would have been*, had certain facts obtained. A proximal selective explanation requires that the trait in question, in order to be vestigial, has either been

[44] See Sober (1984) for more discussion.

selected against in the last evolutionarily significant period, or *would have been*, "had the mutation rate not fallen below expectation" (p. 128).

Because of its modal commitments, Griffiths' view runs into the same sort of difficulties that Godfrey-Smith's does. Etiological theories purport to establish function without appeal to counterfactuals or, at least, without appeal to counterfactuals other than those used in establishing which factors were, *in fact*, causally relevant. One of their virtues is alleged to be the actuality of the events to which they make appeal.

Griffiths has the added problem of being committed to the view that, had mutations occurred in the evolutionarily significant period, they would have been worse, i.e., more costly, than the trait in question.

Both attempts recognise that there is something like an ahistorical element to attributions of vestigiality, along with the historical component. Trying to meet this demand, within the framework of an etiological theory, i.e., without appealing to ahistorical considerations, is difficult. Bigelow and Pargetter's version of the selectionist theory fares better, being committed already to the modal claims about the present. Their view is discussed further in the next section.

§ 3.22 *Systems-Theories*

Now, let us consider the systems-theoretic approach. It handles the problem of non-functional features by means of the initial simple formulation offered in § 2.22 above, although the discussion of dysfunction in § 3.12 demonstrated the need for some careful

refinements or interpretations. Recall that the basic idea is this: if a feature does not have the capacity to contribute causally to the fulfillment of a larger capacity or goal of the system of which it is a part, it has no function in that system. A vestigial trait would be one that once had this, but no longer does. The problem of loss of function of a type of feature *could* be handled by appeal to statistical normality. Once a certain feature ceases to play its role in a sufficient number of cases, it no longer has a function.

However, this has counterintuitive results, as in the cases mentioned above in § 3.12. There, we were forced to conclude that certain "dysfunctional" objects or traits cease to be dysfunctional as the percentage of their incidence increases beyond a certain point, if we rely upon statistical normality in determining functionality. Similarly, certain "functional" traits would cease to be functional if too few systems were to exhibit their would-be functional consequences, and the results are not harmful. If, in the past, some human appendices did whatever useful thing it was that all of them once did, and some did not, the "functional" appendices, the ones contributing to the overall working of their bearers, would cease to be functional after being reduced to a small enough minority. They would be vestigial even if they were of benefit to their bearers. An appeal to statistical normality may seem plausible only when no tokens of the trait under discussion have a function. There is no inclination to think that *anybody's* appendix is functional, so it is not immediately obvious that the appeal to statistical normality is flawed.

One obvious move at this point is to require that no tokens of the type in question be able to perform their (would-be) function, keeping in mind that such a fact not compromise the existence of their bearers. (Such a change must be due to either intrinsic

or extrinsic change in which the organism does not suffer a loss. In intrinsic cases, another organ must have taken over the role. In extrinsic cases, the environment must have changed in such a way as to make the function performed otiose.) This requirement seems too strong, as it would imply that the existence of even one human with a useful appendix would entail that *my* appendix also has a function.

However, in the case of vestigiality, perhaps adherence to a view that requires that it be *traits* that are vestigial is misguided. Just as vestigiality is restricted to particular functions, and can be applicable with regard to one yet not to another, so, too, can vestigiality apply to some tokens, yet not others. On this view, if some appendices do their bearers good, and others do not (yet do them no harm), then the latter are, *ceteris paribus*, vestigial. My appendix does not have a function just because yours does. Nonetheless, given common ancestry, similar setting, and similar morphology, it is not unreasonable to maintain that they are both appendices. Relevant, again, is the type/token distinction. If my appendix were not of the type appendix (however exactly that is cashed out), it would not be vestigial—it would simply be non-functional.

It is important to note that a systems-theory need not be committed to there being *no* relevant historical considerations in the attribution of function. It may depend on the case and on the explanatory project. But, even if no historical considerations are relevant for the attribution of function, clearly some are for an attribution of vestigiality. Vestigiality is a diachronic notion, requiring change, but it is no part of systems-theory to claim that history has no role to play *anywhere*. Given the supervenience of function on morphology and environment, the propensity theorist is able to address questions of

change in function. In the (hypothetical) example of a gradual change in function of the human appendix, the propensity theorist is in a position to distinguish two sub-types of organ, despite *and perhaps even because of,* their common etiology. This sort of response will be open to the systems-theorist as well.

§ 3.3 *Function/Accident*

Another distinction that many think ought to be accommodated by a theory of function is the function/accident distinction. A standard example in the literature is that of the "heartsound"; we want a theory according to which it is a function of the heart to pump blood, but it is not a function of the heart to make heartsounds. Pumping blood is something that hearts *should* do, whereas making sounds is merely something they usually or tend to do. Some aspects of a trait are "accidental" to it, while others are not.[45] Any worthy theory of teleological function will allow us to distinguish between accidental effects and functions.

§ 3.31 *Selectionist Theories*

The function/accident distinction falls nicely out of the etiology theory. As Wright states, "the things X can be said to do by accident are the things it results in which cannot explain how it came to be there" (1973, p.165). It is not the fact that hearts make

[45] Another version of this distinction equivocates between types and tokens. See discussion of Boorse (1976) in the next section.

heartsounds that accounts for them being here—it is the fact that they pump blood. Despite the fact that recent advances in medicine have enabled us to use heartsounds for diagnostic purposes, it still seems strained to suppose that making heartsounds is a function of hearts.[46]

Even so, the function/accident distinction is not one that everyone wants to draw. Boorse (1976) claims that what we want instead is captured well enough by a strong/weak function distinction, and that we are mislead into thinking otherwise by focussing on *the* function of a thing, instead of *a* function of a thing. While it seems that belt buckles have the function of buckling belts, not stopping bullets, consider the belt buckle that not only buckles my belt, but also stops a bullet. According to Boorse, "clearly functions may be performed only once and by accident" (p.80). So my belt buckle has or had the function of stopping a bullet. It is, according to Boorse, by placing too much emphasis on the definite article that we are mislead into thinking that functions must be primary or frequent.

I think that what is missing here, once again, is a careful delineation between type and token. Even Boorse does not want to say that it is a function of belt buckles as a type to deflect bullets, just because one did. Such a loose notion of function robs the concept of the normativity it might be thought to have (see § 3.1)

---

[46] The problem of just when a feature becomes functional can be raised here for the etiological theory. Since making heartsounds contributes now to the likelihood of the continued existence of the bearer of such a heart, those of us who make heartsounds are more likely to survive than those of us who don't (were there such people). Consequently, if making heartsounds is hereditary, hearts that make heartsounds seem either to have this function or be about to have this function, depending upon how long natural selection takes, and the availablity of variants.

The etiological conception is, *prima facie*, much tighter. We are restricted to types, and, furthermore, to the properties of types that are relevant to their etiology. Properties that do not explain the trait's having been selected for, are not functions. Wright, an etiologist, thinks that *the* function and *a* function are interderivable although "the definite-article formulation is paradigmatic" (p.141). He claims that he is "concerned with *a* function of *X* only in so far as it is the sort of thing which would be *the* function of *X* if *X* had no others" (p. 141).[47] It is the colour of the polar bear's coat that provides the bear with camouflage, but it is its thickness that keeps it warm. Both are functions, Wright would argues, since either would have been *the* function, had the other not obtained. I mentioned worries about this move in the previous section, and will address them further in the next chapter.

Nonetheless, this sort of consideration makes good sense of the function/accident distinction. Boorse's belt buckle saves his life "just by accident", because it is not there to serve this purpose. The concept of "accident" is arguably importantly historical, and a selectionist approach involves relevant historical considerations. More on this below, after a look at a systems-theory attempt to handle this problem.

---

[47] This highlights a problem in etiological theory: if a would-be function is that effect which explains the presence of the trait, then, barring overdetermination, there can be only one function. If there is overdetermination, than neither feature accounts for the presence of the trait. *Actual* causal history can track only factuals, not counterfactuals. I shall return to this problem in the next chapter.

There seems to be nothing preventing a similar handling of the distinction in the case of the propensity theory. If the function of the heartbeat is not to make a sound but to pump blood (as it seems correct to say), then there must be some reason for ruling out the former other than any incoherence in the notion of multiple function. Granted, the colour of the polar bear's coat is not an "automatic, unavoidable by-product" of its thickness. That is, it is not a necessary concomitant. (The term "by-product" is prejudicial in this context, implying as it does both a temporal process, and non-functionality.) It certainly seems that the polar bear's coat could be thick but not white and vice versa. It also seems that the heart could pump blood but not make heartsounds, and it could make heartsounds yet not pump blood, provided that "heartsounds" is not characterised as those sounds made when a heart is pumping blood.

There are two ways to make sense of these sorts of modal claims. First, they could be quasi-historical claims about what properties the trait in question could have come to have, depending for their truth values on facts about possible historical variation and possible selection pressures. Second, they could be claims about the necessary connection between the properties considered ahistorically. On the first reading, what matters to the functionality of two properties that were selected in tandem, is which of them was selected for. Etiological considerations could sunder two properties, even if the separation did not occur, whether because of pleiotropy or allometric considerations or

happenstance. Selection *for* is selection for a particular property, even when that property is always accompanied by another.

The second reading of the afore-mentioned modal claims is more congenial to a systems-theoretic perspective. Sense can be made of Bigelow and Pargetter's claim that theirs is a "forward-looking" stance, even if a trait cannot have one property without having another. An option for the ahistorical theorist is to draw a line defined by the goals attributed to the system, or the capacity of interest. If what is to count as a goal or capacity is sufficiently circumscribed, perhaps the function/accident distinction will take care of itself. Our opposable thumb has the function of allowing us to grasp objects, but not of giving us five digits on each hand, because, for instance, the latter effect confers no propensity for survival. As Bigelow and Pargetter write, in the case of the heart, "our reluctance to credit [the sound of] the heartbeat with a function stems from the fact that the sound of the heartbeat is an automatic, unavoidable by-product of the pumping action of the heart. And that pumping action serves other purposes" (1987, p.195). The heartsound is "redundant" in a functional explanation.

However, this line of defence is too strong, if it is taken to forbid the ascription of multiple functions. Why could it not be the case that a trait has a survival-enhancing propensity in virtue of two different properties, even if they are unavoidably connected? That a trait has another function, "serves other purposes", seems no reason, *a priori*, to deny functional status to other of its properties, effects, or capacities, unless one has an actual selectional story in mind. The fact that the heart's pumping action circulates the

blood is, of course, consistent with its making heartsounds. The question is whether or not this latter property is functional.

One of the reasons that Bigelow and Pargetter do not want to hang their analysis on selectionist considerations was explored in a seminal article by Gould and Lewontin. In this paper (1982), the idea of a spandrel was exploited. A spandrel, in the metaphorical sense, is an unavoidable, *non-functional* feature of an item, whose presence is explained by the functionality of some other feature of the item.[48] Spandrels and the structural features that result in them are necessarily concomitant, but this is not what tells against the functionality of spandrels. Rather, it is their failure to be independently useful. There is nothing to which a spandrel contributes. This is exactly the sort of consideration that BP is intended to address.

Let us look at one more systems-theorist, before moving on to address what I take to be an ambiguity in the term "accident". Cummins considers the use of his analytic strategy in the heartbeat/heartsound case:

> Each part of the mammalian circulatory system makes its own distinctive sound, and makes it continuously. These combine to form the "circulatory noise" characteristic of all mammals. The mammalian circulatory system is capable of producing this sound at various volumes and various tempos. The heartbeat is responsible for the throbbing character of the sound, and it is the capacity of the heart to

[48] In Gould and Lewontin's paper, the eponymous example is the spandrels of the cathedral at San Marco. A spandrel is the triangular space between the outer curve of an arch and the right angle formed by the framework it helps to support. The framework and the arches have functions (in the artefactual sense, see § 3.4), yet the attendant spandrels may not. Spandrels may come to have functions, on Gould and Lewontin's account, if they are suitably modified. Again, see § 3.4. I dispute the claim that modification is required for functionality of natural features, but that is because I dispute the claim that attributions of function necessarily explain the presence of an trait. See Chapter 4.

> beat at various rates that explains the capacity of the circulatory system to produce a variously tempoed sound. (1975, p. 763)

As Cummins notes, the above passage is, presumably, true, but does it permit us or force us to ascribe to the heartbeat the function of making the heartsound? Cummins gives a somewhat qualified negative response.[49] However, his reasoning is different from that of Bigelow and Pargetter, and involves an appeal to when an analytic account is of explanatory interest. If we want to explain the capacity of the body to make the noise it does, then heartsounds have a function. This worrisome liberalism of structural functions will be addressed in Chapter 4.

The discussion in this section has taken us far from the function/accident distinction with which we began. At this point, I want to make clear the ambiguity I think implicit in the distinction. In one sense, "accident" is etiological, and in another it is ahistorical. In the former, the issue is one about why a feature is here: is there a non-accidental story to be told about its presence? In the latter, the issue is one of a feature's current relevance: is there an non-accidental story to be told about its efficacy with regard to a larger behaviour of the system of which it is a part? The verdict, in the case of heartsounds, may be the same, but if so, it is so for different reasons. For this reason, I think that the function/accident distinction is best viewed as the function/no-function distinction, and resolution will await a discussion of appropriate explanatory projects, in the next chapter.

---

[49] It is qualified in that he thinks "that there is no black-white distinction here, but a case of more-or-less" (1975, p. 764).

§ 3. 4 *Natural/Artefactual Function*

So far we have been considering examples of function largely from biology, even though our intuitions are often informed by function statements in ordinary language discourse. This discourse is often concerned with humanly designed objects and traits. We make things with the intent that they exhibit certain features. Chairs have backs because that way they are more comfortable to sit in, and the function of those little vents at the back of a Volkswagen van is to direct air to cool the engine. The blanket on my bed in winter is there to keep me warm, and the function of a gas gauge is to tell the observer how much fuel is left in a tank.

How do these function statements fit in with our earlier analyses of function? How well should they fit? Some take such cases to be paradigmatic for teleological talk (e.g., Wright (1976), Falk (1981)). I discuss this issue further in § 4.2, when I examine how closely linked the two relevant explanatory projects are. In the current section, I sketch some responses that have been offered.

§3.41 *Selectionist Theories*

There is some disagreement within the etiological camp with regard to how far this analysis of function can range. Wright (1973) thinks that he can provide a univocal account of both natural functions and the functions of consciously designed items, i.e., artefacts, and that such univocality is a virtue. He claims that "both natural and conscious

functions are functions by virtue of their being the reason the thing with the function 'is there' [subject to some restrictions]" (p.164). Neander (1991b, pp. 462-3) distances herself from this position on the grounds that, while natural and intentional selection are related and are both relevant to attributions of function, they are sufficiently different so as to affect the details. Functions that result from natural selection are "generalizable over types, [whereas] artifact functions may be idiosyncratic".[50]

Neander cites James Bond's one-of-a-kind car as an example. My token blanket on my token bed is another. In all cases, it is central to the etiological account of biological functions that they be bequeathed as a result of selection for, where the "result of" entails the causal relevance of a token's type-ancestors. Neither James Bond's car nor my blanket have the requisite causal history. In addition, conscious agents have a sort of creative foresight that renders appeal to the causal history of a trait otiose. "It is enough, in the case of intentional selection, if the designer believes or hopes that the artifact will have the desired effect and selects it for this purpose" (Neander, 1991b, p. 462). She thinks that her account is more accurate, and that it offers a unification of sorts in that the functions of artefacts are still characterized in terms of their effects, albeit their intended ones.

It would seem that actual causal history cannot be decisive in the case of a *newly created* artefact, if one has a view of selection that requires differential success of actual instances of ancestral tokens of the type of the token in question. Some have thought (e.g., Dennett (1975)) that the "generate and test" model of creativity requires simply the

[50] Millikan (1984) wishes to broaden the extension of "biological category" so as to include some "objects" that might have been considered artefacts (linguistic devices for one) and treat this entire set univocally.

occurrence of a selection process in thought. Perhaps it does, but, notably, it is *perceived* consequences, not actual ones, that affect the production of newly created artefacts.[51] I may produce something because I *think* that it will have a certain effect and I may well have entertained a variety of options before settling on the one I did (though it is not clear to me that I need to have done so). Nonetheless, this seems insufficient to bestow a function on a thing in a sense strong enough to endow the item with normativity.

Imagine that my son had built a contraption from cardboard, tape, and old pop bottles. He spent much time planning, and thinks this contraption will fly him to the moon. Alas, it cannot. Now, is it dysfunctional? There are two lines one could take here. One is that his cardboard artefact is not dysfunctional. There is no sense in which it is malformed or defective. The only sense in which it should do that which it cannot, is the sense in which it fails to meet my son's expectations. The second is that the rocket ship is dysfunctional. It has a function that it *should* perform, but it does not and cannot.

Which line one ought to take will depend upon how one types this token. If it is of the type spaceship, then it is dysfunctional. It is part of being a spaceship that a thing should be able to fly in space. However, if it is of the type thing-someone-wishes-were-a-spaceship, then it is not dysfunctional. Things of this type do not exist because of past successes, therefore, on the etiological line, they do not have the function of flying. They exist because of past successes at being thought to hold promise of future *hoped for* successes. The normativity is a hoped for normativity, not a de facto one. At any rate, it

---

[51] This need not be the case for *reproductions* of artefacts. The market is a hard taskmaster. Yet even it is not up to the standards of natural selection. Witness the number of "talismen" and the like that are sold. Not one of them performs, or is even able to perform, its purported function. (I am assuming that the function of such things is to bring luck, not to make money for their manufacturer—perhaps I am being naive.)

is sufficient simply to note that there is the possibility of a univocal selectionist account of natural and artefactual functions.

However, this univocality is not open to advocates of GS. According to this approach, it is a necessary condition that the item in question be a component of a biologically real system. This precludes artefacts (at least those not on the fuzzy border between biological and non-biological, should there be any).

§ 3.42 *Systems-Theories*

Some systems–theories offer a univocal account, on the grounds (presumably) that the goal of a system, or the larger capacity of interest, can be that which an agent *intends* it to be, whether or not the capacity or goal is realised. Relativizing analyses in this way permits them to range over natural and humanly intended goals.[52] However, this construal is prone to the same worry hinted at above. Some things, despite our best efforts, fail to do what we designed them to do, and it is not clear that they are, simply in virtue of that sort of failure, dysfunctional. Further investigation must await an articulation of the explanatory aims of functional attributions.

At any rate, each approach offers a degree of univocality of account of natural function and artefactual function. *Prima facie*, there ought to be some connection between various uses of the term "function".[53] In the following chapter, I question the degree of univocality required.

---

[52] For example, see Wimsatt (1972, pp.60-6),and Boorse (1976, pp.77ff).
[53] At least, it seems there ought to be for the uses of the term "function" that are relevantly similar. See Wimsatt (1972, pp.3-10), and Wright (1973, pp.139-40) for discussion.

In this chapter, I have considered the performance of the theories presented in Chapter 2 in meeting certain plausible desiderata. As we have seen, no one theory meets all the desiderata, let alone meets them comfortably, unamended, and in all domains. I offer some diagnosis and some prescription in the following chapter on explanation.

Chapter 4

Functional Explanations

A horse walked into a bar. The bartender asked, "Why the long face?"

An Alberta joke, courtesy of Julie Colterjohn

In the previous chapter, I examined two sorts of analyses of teleological function, demonstrating the *prima facie* legitimacy of each sort by showing how some analyses of each type go some way toward fulfilling relevant desiderata. In this chapter, I offer a diagnosis of the difficulties each of them has in meeting *all* the desiderata, or in meeting each desideratum equally well. The crux of the issue is explanation—the explananda appropriate to each sort are different. I argue, roughly, that one family of analyses is best suited to explaining why something is where it is and how it came to be as it is; the other is best suited to explaining how it is that something works.[1] The first section of this chapter is devoted to outlining this position and to noting some complications and subtleties. In the second section, I sketch my position on the relationship between these two sorts of explanatory projects and their attendent notions of function. Finally, I argue that, insofar as any normativity is adduced or required, the selectionist approach must presuppose the existence of a systems–theoretic analysis and that, usually (though not necessarily), a systems–theoretic approach presupposes the existence of a selectionist

[1] This basic distinction was made at least as early as Boorse (1976).

analysis. To put it pithily, Darwin needs Aristotle, and (most of the time) Aristotle needs Darwin. An explication of the import of this analysis for the philosophy of mind is deferred until the next chapter.

§ 4.1 *Two Explanatory Projects*

I maintain that members of the two different families of analyses of teleological function are suited to two different explanatory tasks. Roughly speaking, the selectionist approach is appropriate when what is to be explained is the presence of an item with the form that it has; the systems-theoretic approach is appropriate when what is to be explained is how an item works.

§ 4.11 *Selectionist Functional Explanations*

Recall the central component of the selectionist analyses adumbrated earlier—*viz.*, the appeal to a particular sort of causal history of the item (type or token) in question. Broadly speaking, the item (or its ancestors or earlier versions) must have been selected for having performed the effect that is, as a result, its function. In the case of BP, the item must be such that it has an effect which makes it likely to be selected for in the future. To attribute a function on an etiological selectionist analysis is to make substantial claims

about the past.[2] To attribute a function on a propensity selectionist analysis is to make substantial modal claims concerning the future. In either case, exactly how substantial these claims are depends upon the details of the analysis in question—the more substantial the claims, the more restrictive the notion of function. In all cases, selection is central.

It is, perhaps, commonplace to claim that an etiological selectionist account of function is appropriate when the explanandum is the existence of a thing with the form it has. According to these accounts, a thing's causal history is what gives it a function, but only if that would-be function is implicated in an explanation of the thing's existence. That this is an appropriate sort of explanation is virtually explicit in the characterisations of all the selectionist analyses I have considered.[3]

Before fleshing out and trying to refine this suggestion, I should allow that there may be other sorts of explanations in which the concept of function, given an etiological selectionist construal, legitimately appears. It would be foolish to claim that there is *no* other possible explanandum appropriate to such an account of functions—after all, negative existential claims are notoriously difficult to defend. Instead, I shall argue that another significant sort of explanandum that is also an appropriate venue for invoking teleological function, is better met by a different sort of account, the systems-theoretic

---

[2] It is for skeptical worries about our capacity to know the past that Caro and Borgerhoff Mulder (1987) reject an historical understanding of "adaptation". I think their pessimism surprising, and their conclusion, based on such epistemological concerns, unwarranted.

[3] The exceptions are MB and BP. Due to the ambiguity of the clause that contains "because", MB could be understood in more than one way. Hence, it does not wear its explanatory role on its sleeve. However, a reading of Bedau (1992) makes clear that the "because" of his formulation is intended to be etiological. This is what puts him in the selectionist camp. BP is more troublesome and, as already noted, requires different considerations.

account (see § 4.12). Apart from demonstrating the inapplicability of attributions of selectionist-type functions to this project, I shall take the burden of proof to be on those who would assert that a selectionist account is relevant to more than I have claimed.

At any rate, many writers about teleology, both selectionist and otherwise, acknowledge and defend the view that the first sort of explanandum mentioned above—concerning existence—is one in which appeals to teleology are legitimate.[4] Godfrey-Smith writes that "[etiological] functions are used in explanations of why the functionally characterized thing exists now" (1994, p. 353). Matthen claims that "functional explanation in biology seeks to explain why an organism has a particular characteristic" (1988, p. 14), and that the features of interest in an attribution of teleology to things are those "relevant to explaining their presence with these features" (1997, p. 22). Griffiths writes that "[p]roper functions [i.e., teleological ones] differ from other functions in that they can be cited to explain the presence of a functional item" (1993, p. 411).

Many earlier writers articulate the explanatory schema admirably, and are adamant about the restrictiveness of legitimate appeals to teleological function. Francisco Ayala (1977, p. 498) writes:

> *Teleological explanations account for the existence of a certain feature in a system by demonstrating the feature's contribution to a specific property or state of the system.* Teleological explanations require that the feature or behaviour contribute to the existence or maintenance of a certain state or property of the system. Moreover, and this is the essential component of the concept, the contribution *must be the reason why the feature or behaviour exists at all.*

[4] Some go so far as to claim that this is the *only* legitimate sort of explanandum for any sort of account.

Notice the temporal ambiguity in the word "contribution" in the first sentence. It is not restricted to past performance, so the passage permits a reading that requires current performance of the function in question. This allies Ayala's view with that of Godfrey-Smith. Note also that it is not clear whether it is types or tokens whose existence is to be explained. This distinction is, the reader will recall, central to the function/dysfunction distinction. It is also central to the issue of explanation, as shall be seen.

Hempel, exhibiting some of the same vacillation between, or unclarity about, current and past performance, maintains a similar view: "[t]he kind of phenomenon that a functional analysis is invoked to explain is typically some recurrent activity or some behavior pattern in an individual or a group, such as a psychological mechanism, a neurotic trait, a culture pattern or a social institution. And the principal objective of the analysis is to exhibit the contribution which the behavior pattern makes to the preservation or the development of the individual or the group in which it occurs" (1965, pp. 304-5). Going back a little further, Aristotle maintains that to know the function of a thing is to know what it is for the sake of, and to know what it is for the sake of is to know why it is as it is, in the sense of knowing what caused it. Consider, for example, this passage: "In the sense of end or that for the sake of which a thing is done, e.g., health is the cause of walking about. ('Why is he walking about?' We say: 'To be healthy', and, having said that, we think we have assigned the cause.)" ($194^{b}33$, from "Physics").[5] Providing a teleological explanation, for Aristotle, is providing the reason for which the thing was or is caused to exist.

[5] See, also, e.g., "De generatione Animalium" and "De partibus Animalium."

To modern etiological theorists, every item has some or other causal history, but only histories of a certain sort warrant an attribution of function. Members of a biological type can exist for reasons other than selection history, e.g., because of pleiotropy, random genetic drift, mass mutation, or structural concomitance.[6] When that is the case, appeals to etiological function are not (directly) relevant. The feature is where it is as it is, not because it has been selected for—rather, there were other factors at play.

Though not a possibility I intend to take seriously, one could maintain that something is where it is as it is because God put it there. If one has the Leibnizian view that God selected, of all possible distributions and types of things, the array of actual things because it is the best, then one could claim that selection *does* explain a thing's presence—in fact, it would explain *everything's* presence. However, etiological function, in the modern sense, would not be relevant, unless one holds that supernatural selection is sufficiently analogous to natural selection. However, if one does this, the products of pleiotropy, random genetic drift, mass mutation, or structural concomitance *would* have functions, *contra* the modern etiological view, since God selected those features of the actual world, as well. Such a Leibnizian view would attribute functions ubiquitously.

---

[6] Pleiotropy is one way in which the phenomenon of selection of a genetic component for reasons other than its phenotypic expression can occur. Some genes have multiple phenotypic effects, and other genes are "linked" to genes whose phenotypic effects are selected for. Their continued existence is derivative.

By "structural concomitance", I mean features that are simply the results of other features, not having been directly selected for. Gould and Lewontin (1979) argue that the chin is just such a feature, resulting from the structure of the jaw, which was selected for its role in the ingestion of food. Lewontin (1978) argues that the relative size and shape of the human chin in relation to that of the chins of apes is the result of a greater rate of regression of the alveolar growth field as opposed to that of the dentary growth field (p. 217). Whatever the explanation of this latter phenomenon (presumably selection), the chin itself has no function. The phenomenon of structural concomitance need not be separate from pleitropy, but may be, as it is characterised only at the phenotypic level.

Aristotle could be construed as holding a selectionist view, though it would be selectionist in only an attentuated sense. Aristotle (see *Physics*, Book II) argued against Empedocles' view that the combinations of traits possessed by extant organisms are those which survived selection pressures. He thought that the world did not change in the way implied by Empedocles' theory (and, much later of course, by Darwin's). For Aristotle, there was no evolution, so the question of which mechanisms were responsible for evolution and, ergo, which, on the etiological view, bestowed functions, was otiose. He considered species to be eternal. However, his view that the world is optimally arranged has him in close affinity with etiological theorists. His view, along with the Leibnizian one just discussed, maintains that the explanation of the existence of a trait is its usefulness. This view is compatible with the extreme etiological position that holds that selection explains the existence of all extant traits. However, to call Aristotle's view etiological is, I think, misleading. Although he believes that the presence of items can be explained in terms of what they are for the sake of, his view focusses upon the rational structure of the world, not its causal history. This rational structure is a synchronic constraint for Aristotle and forcing him into the mold of an etiologist would be a disservice to both sides.

Let us return to the modern etiologist's conception. As discussed earlier, Godfrey-Smith offers an elaboration of Wright's (1973) view. He realises, of course, that one can explain why a trait is where it is as it is, without appealing to selection. This is obvious in the case of vestigial traits. The appendix is where it is and how it is for non-selectional reasons. Lack of variation and/or lack of selective pressure explain why we still have appendices. It is just this fact that motivates Godfrey-Smith (1994) to expand

clause (iii) in his F3 from that of F2 (see above, §§ 2.211, 3.21): "among the properties copied between members of *T* is property or property cluster *C*, which can do F" (p. 359), thereby permitting the theory to match our intuitions that a trait such as the having of an appendix simply has no function. An etiological theory that does not require traits to be, in some sense, currently advantageous has the unfortunate consequence of attributing functions to traits to which normativity is inapplicable. It is the addition of "which can do *F*" to Godfrey-Smith's modern history theory of functions that both gets him out of trouble and lands him in it, as we began to see in Chapter 3.

However, for our purposes, and despite our earlier examination of what it means to say that something has been selected, we have not been paying sufficient attention to a distinction that dogged us previously—*viz.*, that between tokens and types. I have been (deliberately!) lax about specifying exactly what it is the existence of which is to be explained by appeals to selection history. Is it the existence of a particular token item? Is it the existence of a type of item? Is it the prevalence or degree of prevalence of instances of a type of item? Is it the form of a token, or of the form that is 'characteristic' (to use as neutral a term as possible) of members of a type? It is obvious that these explananda are distinct, but it is not obvious which one(s) require(s), or can use, ascriptions of function, etiologically understood. In order to work through these options, I now turn to an examination of a dispute between Neander (1988, 1995a, 1995b) and Sober (1984, 1995), with regard to the explanatory role of selection.

These problems, when worked out, show that selectionist considerations explain even less than one might have thought; they explain the existence of a restricted range of

items in a restricted way. This bolsters my claim that the etiological approach is the wrong one for answering many questions in the philosophy of mind; however, my principal objection is that it is not explanations of existence that are our primary concern at all, no matter how little circumscribed. In the following section, I discuss the degree of restrictedness inherent in selectionist explanations of existence, deferring a discussion of the more general problem to § 5.22.

§ 4.111 *Explaining the Existence of What?*

Recall that it is taken to be one of the merits of an etiological selectionist approach that the variables over which statements range refer to types, not tokens.[7] This is, among other things, a requirement for making sense of the function/dysfunction distinction.[8] It seems natural to think, given this sort of analysis, that what is explained by appeal to selection is the existence of the type in question. What this means, assuming that one is not trying to explain the 'existence' of some abstract entity such as a Platonic Form, is that it accounts for the existence of members of the type *qua* members of that type. Even this, however, remains ambiguous, not to mention obscure.[9] Does selection account for (i) the existence of particular individual tokens, (ii) the nature of particular individual tokens, (iii) the 'characteristic' form of a type, (iv) the prevalence of members of that type, or (v) something else entirely?[10]

---

[7] See, e.g., Neander (1993b).
[8] See § 3.1 for discussion.
[9] It is obscure insofar as "*qua*" is obscure, and it is ambiguous in that it is a contender for answering any of the distinct questions that follow immediately below in the text.
[10] I mention the fifth option only to provide a comprehensive map of the conceptual space.

Sober explicitly affirms the fourth disjunct, denies the first two, and may deny the third, as well; Neander argues in favour of all of the first four. She, however, characterises the range of possible explananda differently. Using the example of the opposable thumb, Neander distinguishes these three questions: "[1] the *Individual Development Question ...: how or why does a particular individual develop his or her particular opposable thumb*? ..., [2] the *Persistence Question ...: how or why did the genetic plan for an opposable thumb prosper and persist in the human population, once it arose*? ..., [and 3] the *Creation Question ...: how did the genetic plan for an opposable thumb originate in the human species (or its ancestors)*?" (1995a, pp. 62-3).

I have distinguished four questions rather than three, as I think Neander's taxonomy too vague. I am not vastly happier with my own, though it does have the virtue of helping keep our attention focussed on the type/token distinction. Neander's Individual Development Question is amenable to a token reading—what is it about a particular token that accounts for its developing the traits it does? This interpretation would ally it with our question (ii), though this question would not be completely subsumed, as developmental issues do not exhaust explanations concerning the nature of individuals, as Neander recognises. If one were to be lax enough, one could read the Individual Development Question as allied with our question (i), though this would deflect attention from the features of the individual to its mere existence and would not capture our concern to explain its nature.

The Persistence Question, given a type reading (which is the one Neander intends), is ambiguous between our questions (iii) and (iv).[11] Question (iii) asks why a type has its characteristic form, which is different from (iv), which asks why a certain form is as prevalent as it is among tokens of that relevant type. How to define "characteristic" so that it is independent of questions of the percentage of tokens that exhibit that trait is a vexed issue, involving concerns about type individuation. As we saw earlier, some types are defined historically, some functionally, and some by a combination of the two. In any case, the questions are conceptually distinct.

The Creation Question, given a type reading (again, as I think Neander intends), corresponds roughly to (iii): how did a certain type arise? It is also close to our question (ii): how did certain tokens acquire their status as members of their types?

Both the Persistence Question and the Creation Question can be given token readings. As outlined by Neander, they are questions about the persistence or the origins of a type. However, they could be interpreted as being about the persistence or origins of a token. When they are given token readings, they correspond to question (i) above. At any rate, conceptual space for both sorts of questions is required, though Neander's focus on a type reading is not ill-motivated.

Finally, it should be noted that Neander's questions are phrased in such a way as to make etiology necessarily relevant. This is most obvious in the case of the Individual Development Question: the "how or why" is, by virtue of being a *developmental* issue, a

[11] In the example of Sober's toy, discussed in the text below, this ambiguity is not apparent, as what is characteristic and what is universally prevalent are the same.

diachronic issue. The other questions are also diachronic in that they ask for historical explanations, as evidenced by the use of the word "did".

Neander locates the focus of the controversy over what selection explains as being the last question, the Creation Question. She argues that considerations and puzzles germane to answering this question taint the other two. She maintains that selection *is* part of an answer to the Creation Question, and that it plays a role in explaining the traits of individuals, since answering the Creation Question is part of answering both the Persistence and the Individual Development Questions, though answering the Creation Question does not, unaided, answer the other two. This means that she thinks that selection plays a role in answering both our questions (ii) and (iii).

Sober maintains that selectionist considerations are appropriate in answering only the Persistence Question, which is, on his reading, to answer question (iv)—selection explains the prevalence of certain traits.

There is a principled middle ground between the positions taken by Neander and Sober. Selection can explain more than the prevalence of a trait (question (iv)), despite the fact that, as discussed in Chapter 3, it involves the comparative notion of fitness. It also plays a role, I shall argue, in answering our questions (ii) and (iii). However, it does not explain of *a given individual* why it has the traits it does. Nonetheless, it explains why there exist individuals with the traits in question. It can explain why there exist item(s) *with that particular form*, without explaining why *a particular individual* has it. In fact, it must explain this. If it cannot play a role in answering questions (ii) and (iii), then Paley's challenge is not met, and the theory of evolution by natural selection,

contrary to what many have thought, is no part of a response to the argument from design.[12] It also plays a role, albeit a very indirect one, in answering question (i).[13]

At any rate, the issue hangs on how one views selection. Setting to one side naturalistically untenable views that give selection the powers of foresight and planning, the options are few. Neander (1995b) maintains that either selection is a purely negative force, or it is something with "a more creative role to play" (p. 422). According to her, if selection were a purely negative force, all it would explain is the prevalence, or frequency, of tokens of a trait. She concludes that it cannot be a purely negative force for, if it were, it would be *unable* to explain the "phenotypic or genotypic properties of particular individuals" (ibid, p. 422). Sober concurs with Neander's characterisation of the explanatory limits of natural selection if it were simply a negative force, but does not find this problematic. He argues that natural selection explains population level facts, not facts about particular individuals. For this latter sort of explanation, details about genetic encoding, meiosis, and other mechanisms of inheritance are all that is required.

Sober provides a compelling and intuitive analogy in his explication of the nature of selection (1984, p. 99).[14] He illustrates the difference between selection *of* an object

---

[12] Recall from Chapter 1 that Paley was a prominent advocate of the argument from design to the existence of God. Meeting Paley's challenge is a matter of explaining how it is that complex items can exist without having been created by God.

[13] Despite the fact that I appear to be siding with Neander, I consider my position a "middle ground" between her position and Sober's, since my "defence" is deflationary of the importance of the notion of selection. Furthermore, I agree with Sober that it is purely a negative force. See below in the text.

[14] Sober credits Wright (1973) with making the same point. Wright's distinction is between those instances of selection which are instances of "consequence-selection" (p. 163), and those which are selection at random or for no reason. The former match up with Sober's notion of selection for, and the latter with his notion of selection of, though Wright is not so concerned with the details of the different forces of evolution, preferring to focus on artefacts. He does, however and as earlier noted, think his analysis has the benefit of univocality over artefactual and natural features.

and selection *for* a property by means of appeal to a toy. The toy consists of a tube in which one can put marbles of different colours and sizes. There are barriers that divide the tube horizontally when the tube is standing on end. Each barrier has holes in it which are of uniform size within each barrier, but which are progressively smaller in each barrier, moving from top to bottom. As a result, marbles placed in the top of the tube are filtered, when the tube is shaken, according to size. The smallest make it past all the barriers, the next smallest past all but the smallest barrier, and so on. All and only the smallest marbles make it to the bottom. All and only the smallest marbles are green, so all and only the green marbles make it to the bottom. However, their colour is irrelevant to their "success"—there is selection *of* green marbles, but selection *for* being small.

This is a simple example, but it illustrates nicely the general point. Natural selection, though it acts on or selects objects, selects them in virtue of some of their properties and not others. As I shall discuss immediately below, this analogy does not provide us with a sufficiently complete characterisation of selection, but let us, for now, assume that it does. What phenomena of the scenario can we explain? Look at the whole set of marbles at the bottom of the tube. We can explain *why they are all small* by adverting to the selection pressures of the size of the holes in the barriers arranged as they are. It is this sort of fact—a population fact—that Sober thinks is clearly the domain of etiological explanations that appeal to selection. It corresponds to (iv) above—we are explaining the prevalence of a trait. It may also answer (iii), if "characteristic" is taken to mean statistically normal or typical, but with respect only to size, not to colour.

Furthermore, it answers Neander's Persistence Question (with allowances made for the lack of a genetic plan in the case of marbles).

Now choose a single marble from the set of those that reached the bottom of the tube. The workings of the toy are transparent to us, but we are not in a position to explain a great deal about this token marble. It was both green and small before it went through the barriers—the process of selection was not responsible for making it either of those. What we can explain is why it "survived" as long as it did—i.e., we can explain how it made it to the bottom of the tube. This is not to explain why it is small, but instead to explain why it is where it is. Being there did not give it the property of being small; rather, being small accounts for it being there. This, too, answers the Persistence Question, albeit at the level of a token, not type.

Already, it is apparent that there is a feature of our target explanandum that is missing from our example. There is neither reproduction nor mutation in the marble toy. Everything that goes in the top stays the same size and colour. Nothing that happens in the tube affects any of the properties of the tokens, with one notable exception—the property of their location in the tube. Furthermore, nothing that happens in the tube, except at the last barrier, affects the likelihood of all and only the small marbles making it to the bottom. Neander considers these disanalogies relevant. Evolution by natural selection is, she says, "a cumulative selection process" (e.g., 1995a, p. 75). What happens at each step can affect what happens at the next. In Sober's toy, none of the barriers but the last one determine which tokens get to the end or are 'selected'. In essence, this is an instance of single-step selection, even though there are a number of

barriers in the toy. The toy is not a "cumulative selection device, *because the series of sieves is inessential to the final selection*" (Neander, 1995a, p. 73). In cases of cumulative selection, natural selection helps answer the Creation Question, argues Neander, because natural selection affects the probability of future genotypes arising. The likelihood of a particular genotype arising through recombination or through mutation is a matter, in part, of the genotypes that exist, and which genotypes exist is a matter, in part, of natural selection.

There is something right about this way of thinking. Recall that one of the attractions of natural selection is that it is alleged to explain the existence of complex organisms. *Contra* the creationist, the modern evolutionary biologist maintains that explaining both the intricacy and the adaptedness of extant items does not require postulating a thoughtful deity who creates. Such "creative" results are to be explained by appeal to natural selection. As mentioned earlier, natural selection is thought to be part of a legitimate, naturalistic response to Paley's challenge. That is, natural selection is thought to explain the existence of complex and adapted entities—their existence, their complexity, and their fit with the environment.

Natural selection is to explain such phenomena in roughly the following way: some small changes, whether due to reproductive recombination or to mutation, will be favoured. Those small changes are, as a result, disproportionately strongly represented in the subsequent generation, given heritability. Other, new, small changes may or are more likely to happen, given the new pool of traits.[15] To put it intuitively, the chances of

[15] It is, of course, genotypes that are inherited, though it is, most often, phenotypes upon which selection acts. The distinction is not relevant to my discussion here.

getting a watch with a second hand are much greater if one starts with a watch without a second hand than if one starts with a brick.[16] Paley did not want to know why all the things he found on the heath were watches. He wanted to know why he found a watch.[17]

Answering this sort of question *is* answering a Creation Question and plays a role in answering all of questions (i) through (iv) above. In a sense, we can explain the existence of a token by appeal to selection, provided that, had selection not operated as it had, the token's ancestors would not have existed. This addresses question (i). Selection plays a role in explaining why the token has the traits it does, in the case of those traits which are inherited, since, had selection acted differently, that token's ancestors would have had different traits.[18] This addresses question (ii), though, naturally, the mechanisms of inheritance must also be invoked. Question (iii) is trickier, since it requires that issues of the individuation of types be settled. Even so, if selection has played a role in individuating at least some types, as surely it has in the case of, for instance, species, then it can play a role in answering question (iii). However, I think that in most cases, question (iii) is either question (iv) or a question about complexity. I return to the issue of complexity below. As for question (iv), it is not in dispute that

---

[16] This is reminiscent of Simon's (1969) discussion of Tempus and Hora. Both are watch builders, but one builds things part by part, so that, if he is interrupted, he will not need to start from scratch again. Building that proceeds in stages is much more efficient than building that requires each item be made step-by-step without interruption. Hence, it is not so surprising that complex structures exist, if their history includes discrete stages.

[17] I use this example, despite the fact that I think that artefacts and natural phenomena may require different sorts of explanations. Paley's case is made rhetorically stronger by its appeal to an artefact, but the general point is not obscured. It could be as easily made, though not with such rhetorical power, by invoking a pine cone or a dog.

[18] I assume that sense can be made of the notion of an individual having had different ancestors. Perhaps this is not possible (see Kripke, 1972); then the question would concern why it is that the token that exists is *that* one and not another.

selection often plays a role in answering it, i.e., in explaining the prevalence of traits. In the case of Sober's toy, the presence of *all* the tokens in the last compartment is explained by their capacity to pass the last barrier, their having been selected.

On this view, *contra* Sober, selection *is* part of an answer to the Creation Question. However, it is not the notion of selection *for* that will be doing all the work. It was crucial to our understanding of function that we understood the difference between selection of and selection for. Allowing that selection helps answer the Creation Question requires, almost paradoxically, that we ignore this distinction.

Suppose the green marbles and the second smallest marbles, the yellow ones, can mate, but will do so only if no other coloured marbles are present.[19] (This is, clearly, no longer Sober's toy!) Some of their offspring are lime green with a size mid-way between that of their parents. Others breed true to one parent or the other, either in size or colour or both. However, there are, for whatever reason, no yellow offspring that are smaller than their yellow parent. Finally, suppose that the holes in the last barrier are big enough to let through anything that is smaller than the yellow ones. The set of marbles that make it to the last section is comprised of both green and lime green marbles, all of which are small enough to fit through the holes. Selection does explain, in an indirect way, both the size and the colour of the marbles that make it to the end, even though their colour is irrelevant to their location. It plays a role in explaining why, say, fifty percent of the marbles in the last compartment are lime, even though being lime was never relevant to having been selected. If there had been no selection for being small, there would be

[19] For heuristic reasons, I choose a phenotypic example. The reason I stipulate that no other marbles can be present is that I want to limit the possibilities of recombination and focus on one compartment in the toy.

yellow marbles at the last stage and, therefore, the percentage of those that are lime green and green would be lower.

The moral I take from this hypothetical case is that Neander is right that selection plays a role in answering more than just the Persistence Question. However, the degree to which selection *for* plays a role in answering other questions is straight-forwardedly empirical—perhaps a great many current traits, even complex ones, are more like being lime green than like being small.

Despite this somewhat deflationary reading of the role of selection in answering Neander's and my questions, I believe that selection for does other important work— it explains why so many extant things are as complex as they are and why they are as well-suited to the environment as they are. These are the issues that question (iii) above could be construed as asking. Selection makes incremental cumulative change less improbable, and incremental cumulative change makes complexity less improbable. Small changes can add up. Selection helps explain the complexity and "adaptedness" of Paley's watch in just this way. What it cannot do is tell us anything about how the watch works.

I return to this issue in § 4.12 below, but allow me to sketch the idea further now. Appeals to selection offer us what Lipton and Thompson (1988) call "filter explanations".[20] They claim that "Darwinian explanations are filter explanations" (p. 222). They write:

[20] In their paper, Lipton and Thompson are concerned to refute the claim that the theory of adaptation by natural selection is tautological, which is not our concern here. Nonetheless, their characterisation of explanations that invoke natural selection is helpful.

> Filter explanations are informative because, no matter how ignorant we may be about the criterion [for passing through the filter], the explanation tells us that:
>
> (1) Somewhere, sometime, there was an initial aggregate of X's which were both Y and not Y. [I.e., some were Y and some were not.]
> (2) There is a process which distinguishes X's that are Y from X's that are not Y and selects the former.
> (3) The present aggregate is the result of the application of the process in (2) to the initial aggregate in (1).
>
> This information excludes at least three possibilities: that the present aggregate is a chance occurrence, or that there never were anything but Y individuals or that all the not-Y individuals mutated into Y's.
>
> (sic, 1988, p. 221)

According to Lipton and Thompson, a filter explanation is a kind of recursive, as opposed to circular, explanation. It narrows down the relevant options.[21] It points us toward asking about the criterion that acted as a filter: "... for questioners seeking a more specific causal account of the phenomenon, recursive explanations serve a directive function" (p. 226). Recall Rosenberg's reconstruction of a Darwinian enquiry from Chapter 1.[22] Paraphrasing, the project he outlines involves a series of steps: first, we look to see which species have endured and for how long; second, we take relative length of endurance to be indicative of degree of fitness ; third, we look to see whether the fitness levels hypothesized in the second step are borne out by evidence that members of those species exhibit general fitness enhancing traits; and, fourth, we investigate the particular

---

[21] Lipton and Thompson are concerned to refute the view that "the theory of adaptation by natural selection is a circular explanation and has no empirical content" (1988, p. 215.) Their project is *not* that of ascertaining whether selection plays a role in answering the questions we have raised, so should not be understood as taking a stand on these issues.

[22] See p. 4 above.

way in which members of a given species achieve those general characteristics.[23] Filter explanations serve, at best, up to stage three, insofar as they identify generic fitness-making characteristics. However, they do not address stage four in that they are mute about how an organism achieves those characteristics. The "directive function" of a filter explanation points us toward step four. We need a theory about the organism that explains how it manages to pass through the filter. This theory must be a synchronic theory.

Suppose that these worries about the exact nature of the questions that are addressed by appeals to selection, and the extent to which selection *for* plays a role in addressing them, are misguided. Let us be liberal and allow that etiological selectionist accounts can play a role in answering all four of our questions, and in explaining the degree of complexity and adaptedness that obtains in the world. Nonetheless, answers to all of the questions we have just discussed will not provide us with answers to the question of what gives a particular trait its *normative status*. Recall that ascribing to a thing a teleological function is to attribute to it a normative dimension. There is something it is *supposed to* be doing, and, therefore, it may be *mal*functioning. It is this feature that makes an appeal to teleology promising, since there is a prima facie analogy between misrepresentation and malfunction. However, this normativity is, I argue in § 4.3, an ahistorical, synchronic matter. Hence, selectionist accounts of function to do not adequately accommodate it.

---

[23] In the case of Sober's toy, the distinctness of the steps is masked by the toy's simplicity. First, "species" of marbles are delineated; second, the marbles at the end are assigned the highest fitness in virtue of their "endurance"; third, their generic ability to pass through all the holes is noted; and fourth, the property responsible for their generic ability, i.e., their smallness, is adduced.

§ 4.112 *"Forward-looking" Selectionist Accounts*

Having outlined what it is that etiological selectionist accounts may explain, I turn now to consideration of the sorts of explananda to which non-etiological selectionist accounts of function can contribute. Bigelow and Pargetter are the only exemplars. What, as selectionists, can they explain? They are "future etiologists", on this understanding of their project. What they explain is why a certain feature will (likely) be represented in future generations. Basically, they explain why a thing will still be here tomorrow, with the features it has now. Instead of asking what tokens of a thing did in the past that accounts for the existence of current tokens of that type, they are asking what tokens of a type are doing that *will* account for, or make likely, the existence of future tokens of that type.[24]

There is the possibility of giving BP a token reading, though this is implausible both exigetically, given the examples Bigelow and Pargetter provide, and conceptually, if the function/dysfunction distinction is to be maintained. (See § 3.1 for discussion.)

At this point, I am glossing over most of the subtleties adduced in the previous section, as I want merely to note that the selectionist reading of Bigelow and Pargetter construes them as offering the same sort of explanations available to the etiological selectionist, except that they are predicting the future as opposed to accounting for the present. Both the etiological selectionist and the selectionist theorist using a propensity

[24] Similarly, *mutatis mutandis*, for the other questions discussed.

approach are concerned with the effects of a filter. The difference between them is that one is backward-looking and the other forward-looking.

However, there is an underlying tension in BP, when it is given a selectionist reading. Bare prediction is *not* a kind of explanation. Predicting that it will rain tomorrow is not explaining why or how it is that it rains tomorrow, even if the prediction is correct.[25] Without engaging in an extended discussion of explanation, clearly bare prediction is not sufficient.[26] We are looking for some reason to expect the predicted state or event and to unify it with our current view of the world.[27]

Of course, BP is not simply a matter of *bare* prediction. The predicted future success is explained in terms of a specific realised survival-enhancing propensity. The variables in the BP schema are not left blank, after all. For example, it is because the world will 'filter' out animals that do not have efficiently circulated blood and because the heart has a propensity to pump blood, that the heart confers a survival-enhancing propensity on its bearers. Selectionist propensity explanations are not contentless. However, the content they have is *ahistorical.* It is a propensity *currently held* that explains the truth or legitimacy of the prediction. It is because BP is concerned with (future) selection that the theory fits into the selectionist camp. However, because their

---

[25] The Deductive-Nomological model of explanation maintains that explanation and prediction (though not "bare prediction") are intimately connected. This is not the venue in which to attack the model directly, though I do think it is flawed. See note 51 below.

[26] Not only is bare prediction not sufficient for explanation, it is not necessary either. We could explain why it rained yesterday, even if we had not, or even could not have, predicted the event. This sort of after-the-fact explanation is not at issue in the context of a "forward-looking" theory such as BP.

[27] I shall discuss what I take to be the constraints placed on realist theories of intentionality by realist theories of intentional explanation in § 5.22. I shall argue that, insofar as we are interested in causal explanation, we require a synchronic account of the intentionality of some states.

explanatory project essentially involves present considerations, I think it best viewed as belonging in the systems-theoretic family. The content of their explanation is ahistorical. For this reason, I shall consider only those theories belonging to the etiological family of the selectionist camp as being *essentially* concerned with selection. BP, despite its selectionist face, belongs in the structural camp.

§ 4.12 *Systems-theoretic Functional Explanations*

As I have already noted, there is another explanatory project in which an appeal to functions is appropriate. I shall argue that it is distinct from, though not wholly unconnected to, the selectionist one. Instead of endeavouring to explain the presence of an item, or the prevalence of tokens of a type, or even the complexity of an item, one could wish to explain how an organism (or other entity) performs a certain task and, in the course of offering that explanation, make appeal to the effects of certain components, properties, or parts of the item.

It is precisely this sort of task for which a systems-theoretic construal of function is *a propos*. Offering a functional analysis along the lines of Cummins (1975) would meet the demands of just such an explanandum. Putting this project in terms of a Darwinian enquiry, we are looking for answers at step four. What is it about an organism that allows it to perform a certain task? We want to know how it is organized and what properties it has that explain its larger capacity.

The same sort of equivocation between type and token that was noted with respect to selectionist explanations is possible with a systems-theoretic explanation of how an organism or item achieves the relevant characteristics. However, if the function/dysfunction distinction is to be made, a type reading is required in offering either kind of explanation. This is the most natural in any case, since a systems-theoretic explanation of a token would provide us with the grounds for a more general explanation that ranges over the type.[28]

Of the three structural versions of systems theories,[29] BP restricts the larger capacity in question to that of being selected in the future, while RC and WW are both open-ended enough to allow that the task in question need not be directly linked to selection, either past or future. WW is more restrictive than RC in that it requires that an ascription of function be relativized to a "purpose", where purpose is spelled out partly in terms of complexity. Wimsatt (1972) thinks that all the biological items to which functional ascriptions are appropriate have *in fact* been selected for, but that this is not a constitutive fact about their functionality. It is, if true, an empirical fact that such complexity does not arise in nature *ex nihilo*, so to speak.[30]

A worry presents itself. Does a systems-theoretic approach attribute functions too liberally? It would seem that it might, for two reasons. The first reason is that the nature of the larger capacity in question may be, especially in the case of RC and WW,

---

[28] The difficulty with this view lies in assigning tokens to a type even when they cannot perform. I tried to mitigate this problem in Chapter 2, as the reader may recall. It remains a serious problem, but it is not restricted to the systems-theoretic view.

[29] Recall that our discussion of systems theories will focus on those taking the structural approach, as opposed to the goal-directed one. See § 2.3 for discussion.

[30] See Chapter 6 for a discussion of the ramifications of Swampman, the hypothetical creature whose origins are agreed to be extraordinary.

insufficiently circumscribed. The second is that the nature of what is to constitute a system may also be insufficiently circumscribed. There may be nothing to prevent the ascription of functions to arbitrary collections of causally connected items, assigning them "functions" as a result of their contribution to the production of their joint or overall effect. Such a result does violence to our intuitions about what things have functions, especially if, as most commentators agree, attributing a function is attributing normativity.

The key to meeting both these worries lies in appealing to the complexity both of the system under scrutiny and of the nature of its interaction with the environment. Providing a comprehensive analysis of the nature of complexity is beyond the scope of this work, so I shall simply offer a brief characterisation of the concept before addressing how an appeal to it helps prevent the inappropriate liberalism in the systems-theoretic approach.

First, let us consider the complexity of an item. Intuitively, it is a matter of two things: the heterogeneity of its constitutent components and the elaborateness and diversity of the connections between them. An item for which an understanding of its behaviour or properties requires a decomposition of the item into many different types of constituent parts, with some parts being parts of other parts, is more complex than an item with fewer types of parts with less of a mereological hierarchy. Similarly, an item whose parts display a variety of types of interaction, with some interactions resulting in others, is

more complex than one whose parts display just one type of interaction. A Lego toy truck is more complex than a Lego "brick wall".[31]

Second, let us consider the complexity of the nature of an item's interaction with the environment. Again intuitively, it is a matter of the variety of behaviours that the item can exhibit in response to the variety of influences offered by the environment. An item whose behavioural repertoire includes walking and eating[32] is more complex than one whose repertoire consists of nothing but sitting or standing. An environment that is heterogeneous is more complex than one that is homogeneous; the more kinds of things in the environment, the more complex it is. Thus, the complexity of the interaction of an organism and its environment is a function of the degree of their respective complexities of the sort just mentioned and the degree to which they impinge upon each other.[33]

This discussion of the notion of complexity harks back to the earlier consideration of Cummins' analytical strategy and his criteria for its correct application.[34] Recall that the analytical strategy is contrasted with the subsumption strategy. The former is appropriate when an organism is decomposable into a number of sub-systems, and proceeds by examining each sub-system for its role in producing a capacity of the

---

[31] Wimsatt (1972) distinguishes between the descriptive and the interactional complexity of objects. Descriptive complexity is a matter of the degree to which the parts that result from decomposition from different theoretical perspectives fail to map one-to-one. Interactional complexity concerns the degree to which the parts decomposed from one perspective interact with parts decomposed from another. (See esp. pp. 70-2.) His discussion, though technical, yields verdicts on complexity commensurate with those of pre-theoretic intuition.

[32] I am trying to offer examples that are as theoretically neutral as possible. It is often the case that behaviours are characterised in terms of their function, e.g., as flight from predation, as pursuit of a mate. Here I am relying on an intuitive understanding of an item's doing something as opposed to having something happen to it. See Millikan (1993c) for discussion.

[33] In the limiting case of total homogeneity of one of behaviour or environment, there is no interactional complexity, regardless of how hetergeneous the other side of the equation may be.

[34] See § 2.221 above.

organism. The organism is viewed as the "containing system". Functional ascriptions are always relative to both a containing system and to an environment. It makes no sense to assign a function to an item that is not part of a larger system, and it is not possible to provide an analysis in the absence of a reference to an environment. On Cummins' view, functions can often be assigned to a sub-sub-system, with the relevant sub-system now being considered the containing system. One has reached the bottom, so to speak, of such an iterative process when one is no longer able to explain the capacities of items at that level by appeal to their component parts and their arrangement. It is then that the subsumption strategy is appropriate, and functional considerations drop out. This is compatible with (though the details remain to be worked out!) the plausible view that functional ascriptions do not play a role in physics. If it is sensible to view nature as organized in levels corresponding, in broad strokes, to the domains studied by the various sciences, and to consider physics as the discipline that deals with those items at the bottom or fundamental level, then there is the beginning of an explanation of the correlation between the legitimacy of functional talk with non-basic sciences. Complex systems occur only at levels above basic physics and only complex systems warrant attributions of function. When considering criteria for the applicability of the analytic strategy, Cummins himself writes, "[t]he explanatory interest of an analytical account [is partly a function of] the relative complexity of the organization of component parts/processes that is attributed to the system" (1975, p. 764).

If anything like these considerations are legitimate restrictions on the applicability of structural accounts in systems-theory, then the alleged liberalism of those accounts is

mitigated. The individuation of systems is not arbitrary: not just any assortment of items constitutes a system. The degree of complexity exhibited within a system and exhibited in its interaction with the environment, though multi-factored and sometimes difficult to ascertain, is not observer-relative.[35]

§ 4.2 *The Relationship between the Selectional and Systems-theoretic*

Given that I have granted that the function/dysfunction distinction is crucial to any notion of teleological function, I must address a more serious objection to the structural approach than that it fails to maintain univocality between natural and artefactual instances of function or that it attributes function too widely. The objection is that this sort of explanation is not properly teleological because the element of normativity is missing.[36] I turn to that issue in the section following this one. First, however, I want to foreshadow some of that material by summarising where the discussion so far has taken us on the issue of the general relationship between the different senses of function at work in the structural and etiological accounts.

Many commentators, including myself, grant the legitimacy of both the project of explaining a thing's current nature by appeal to facts about its selection history and the project of explaining how it is that something currently works.[37] As a direct

---

[35] A discussion of problem cases is deferred to § 6.1.

[36] The issue of normativity is not separate from the problem of the alleged over-liberalness. Part of what drives the intuitions of many that ascriptions of teleological function must be restricted is precisely that they carry (and *should* carry!) normative force.

[37] For my purposes here, it does not matter how expansive one is about the range of questions that selection plays a role in answering, provided one leaves out the question about how it is that something currently works.

consequence, many commentators, including myself, grant the legitimacy of both families of analyses.[38] Those who consider both legitimate do not, of course, consider them to be in competition with one another. Where significant disagreement lies is in understanding the relationship between the two. My method of viewing them as appropriate at different steps of a Darwinian enquiry confirms the legitimacy of both sorts of explanation by, in part, demonstrating the smooth fit of each family of accounts of function into one sort of explanation and not the other. It also goes further— it links them. A Darwinian enquiry constitutes a series of questions. Answering questions at step four often requires having answered questions at earlier stages. If the analogy with Darwinian enquiry is well-taken, then systems-theoretic, ahistorical accounts assume that a selectionist, etiological account is available. Step four assumes that step three has been made. I argue in the following section that a selectionist account must assume that a structural one is available, both for conceptual coherence and to retain its normativity.

Against this position, some writers (e.g., Godfrey-Smith (1993, 1994) and Neander (1991a)) have claimed that there is a near consensus as to how to construe teleological functional claims—namely, etiologically.[39] More strongly, though they acknowledge the legitimacy of an ahistorical (roughly structural) account, they deny its connection to the etiological approach (e.g., Godfrey-Smith (1993) and Mitchell (1995, p. 41)). Godfrey-Smith writes "... what reason is there to use the word "function" for

[38] For explicit examples, see Godfrey-Smith (1994), Kitcher (1993a), and Millikan (1989b).
[39] Their contention that the etiological theory is near-consensus is supported both by the references they cite and by that fact that, at the Pacific APA in 1993, all three participants, Godfrey-Smith, Mitchell, and Neander, in a symposium on teleological function acknowledged being etiologists. Godfrey-Smith made a quip to the effect that nobody could be found to represent the 'opposing' view.

both Wright [etiological] and Cummins [structural] functions? What do the concepts have in common that justifies this usage? My reply is: there is no strong reason for using the same word ... let no philosopher join what science has put asunder" (1993, pp. 206-7).

Kitcher (1993) is a prominent exception.[40] Let me explain how, by rehearsing the line of argument he gives there. He claims that he wants to reconcile two disparate domains of functional discourse by showing that they share a common feature. He argues that common to both biological and non-biological instances of functional ascription is the assumption of design or of something *like* design.[41] He writes that "[t]he function of an entity *S* is *what S is designed to do*" (1993, p. 379). The non-biological entities he has in mind are artefacts. Artefacts are, quite literally, designed, even if some of their features may be fortuitous. The intentions of the designer are explicit in terms of the intended overall workings of the item, and often explicit with regard to most of its parts. However, some may be unforeseen or even unobserved. His example is an accidentally dropped screw that acts as a needed connection.[42] Biological entities are also "designed", in an attenuated sense, by having been selected for. Natural selection is the designer. Kitcher argues that, analogously to the case of artefacts, not all aspects of biological

---

[40] Millikan (1989b) is also an exception, but in an attenuated sense. She thinks (p. 176) that there is a connection between Cummins' functions and etiological "proper" functions only in a particular domain, *viz.*, the study of life cycles. She thinks this because of the cyclical nature of life cycles and of what she considers to be the subject matter of biopsychology. See § 5.21.

[41] Matthen (1997) has a related view, in that he thinks that functions " ... are attributed to natural things by virtue of analogy with instruments designed for use by or actually used by an agent for a purpose." (p. 31). He calls this "the product analogy" and objects to Kitcher's construal of natural selection as an instance of design.

[42] His example is of a token artefact, but the same point could be made using an example of a type.

entities need be "observed" by natural selection—i.e., they need not have been selected for. His example is of a wing pattern in a moth that offers a sub-optimal degree of camouflage benefit; it is being selected against. However, some random event wipes out the superior variant. Even though the pattern has not been selected for, it has, says Kitcher, a function. In both the natural and artefactual cases, it is because the part in question plays a role, despite the absence of actually having been selected for, that it has a function.

Kitcher is drawing our attention to the parallel features of biological and non-biological function. His motivation for appealing to artefactual examples is, I think, not so much to show that accounts of artefactual and biological function must be or even can be univocal, but rather to show that thinking of natural phenomena *as* artefacts informs much of our pre-theoretical thinking about functions. This is no doubt true, though another step is needed to show that a connection or unity is desirable.

At any rate, there is an important aspect or consequence of his view. On his account, some biological functions have been selected for and some have not. These latter have their functional status ahistorically—they are structural functions. In considering that artefacts and biological phenomena are both designed, we view nature from what he calls the "environment-centred perspective" (1993, p. 381). If we view nature from this perspective, we can ascertain which selection pressures are facing an organism and evaluate how the organism's components meet these demands, whether or not they have been selected for. With this move, Kitcher does what most commentators

find objectionable—granting the legitimacy of a structural approach to biological, hence teleological, function.

While I certainly agree with Kitcher that there is a legitimate sense of structural teleological function, I do not agree that it is grounded in the assumption of design. It is grounded in the legitimate need to explain the complex behaviour of complex organisms. Design, whether natural or artificial, may perhaps explain the existence of such phenomena, but design is not mandatory.[43] Undesigned items that exhibit the same sort of internal and behavioural complexity are improbable, but their workings require explaining.

§ 4.3 *The Source of Normativity*

One may wonder about the connection between system-theoretic accounts, with their attendant structural sense of function, and selectionist ones, with their etiological sense of function, *specifically with regard to normativity.* I take the options here to be five: (i) that there is no unifying thread or connection—the would-be normativity of either sense must be defended without appeal to the other,[44] (ii) that attributions of structural function rely upon or assume attributions of etiological function for whatever normative force they

[43] Allen and Bekoff (1995) object to Kitcher's view on the grounds that "design" implies alteration, not just selection, and that, therefore, design and selection cannot legitimately be conceptually linked so closely.

[44] See, e.g., Godfrey-Smith (1993) and Amundson and Lauder (1994), who maintain that the only legitimate sense of ahistorical is not normative. Godfrey-Smith (1994) does acknowledge that there is a connection between the non-normative ahistorical and the etiological.

have,[45] (iii) that attributions of etiological function rely upon or assume attributions of structural function for whatever normative force they have, (iv) that the two notions are symbiotic, so to speak, each relying upon or assuming the other for its normative force, and (v) that the notions are unified by means of some third factor or consideration which bequeaths normativity on both.[46] I shall argue for the third option, while acknowledging that there is something importantly right about the fifth option, in some cases.

At a less detailed level, the generic options with regard to normativity are straightforward: (i) neither sense of function is normative, (ii) the etiological sense of function is normative, but the structural is not, (iii) the structural sense of function is normative, but the etiological is not, and (iv) both senses are normative. Despite some disagreement on the nature of the relationship between the structural and etiological, the current consensus view in philosophy of biology and of mind is, as I noted above, (ii)—the etiological sense of function is genuinely normative, but the structural one is not. It is the etiological one that is alleged to capture the teleological nature of biological function.

Despite the differences between the world views of Aristotle and of Darwin, there is an aspect of Aristotle's view, as mentioned earlier, that offers support to the notion that it is the etiological, not the structural, that is normative. Recall that, for Aristotle, to know the function of a thing is to know what it is for the sake of, and to know what it is for the sake of is to know why it is as it is. The natural world, according to Aristotle, exhibits a rational order; "the invariable plan of nature in distributing the organs is to give each to such animal as can make use of it; nature acting in this matter as any prudent man

---

[45] Walsh and Ariew (1996) may hold this view, though they have a category, "evolutionary function", that crosscuts the ahistorical/etiological distinction.

[46] Kitcher (1993) argues that this third factor is the assumption of design.

would" (687^a12). Ascertaining a thing's function allows one to situate it in the rational order; it explains its existence.

By re-interpreting the metaphysically spooky claim that the world exhibits a rational order, the etiologist may well find this sort of view congenial. The natural world does have a "rational order" in that most biological things are the products of selection. Most organs do have a *telos*, in an attentuated sense—they exist because they were selected for. They exist because they are for the sake of performing that effect which is designated as their function. Darwin, it would seem, has kept what is right about Aristotle and jettisoned the rest.

Even so, I want to resist the claim that normativity resides only in etiological function, maintaining instead that it resides in both, and that its presence in etiological ascriptions is derivative. To do this, I shall explore further the relationship between structural and etiological function, arguing for option (iii) (that the etiological gets its normativity from the structural). First, I discuss an important consideration which is independent of the nature of that relationship. In short, it is by exploring what and how an appeal to etiology explains that we will begin to illuminate it.

The modern, Darwinian etiologist faces a difficulty that Aristotle did not realise obtains. To Aristotle, species do not come into or go out of existence—species do not even change. *A fortiori*, he thought that things can neither cease to have a function they once had, nor acquire a new one. We know differently. Consider, for example, the human appendix. It was once selected for its capacity to digest cellulose. It is now considered a functionless organ; it once had a function, but is now vestigial. There is

nothing that your appendix is *supposed to* be doing—it does nothing useful for you, but is not, even so, malfunctioning. Having once been selected for is clearly not sufficient for ascriptions of normativity.

The etiological view should not founder on so obvious a rock. Let me remind you of our discussion in § 3.2. There are three ways the etiologist may attempt to navigate around it. One is to restrict the relevant selection history to the recent past.[47] After all, human appendices have not been selected for digesting cellulose for a very long time. If a principled notion of "recent past" can be adumbrated,[48] then the etiologist would not be saddled with attributing functions to vestigial traits. The problem with this defence is that, if there fails to be sufficient variation, and thus selection, a trait will, counterintuitively, fail to be functional, no matter how useful it is. Selection *for* is a relative notion—in order to be selected for, something must be more successful than its competitors. If there are no competitors, there can be no selection for. If there has been no competition with human hearts in the last while, then, on this view, hearts have no function. They have not been selected for, even though, presumably, they *would* have been selected for, had other competing variants arisen (since mutation is more often detrimental than not). Of course, this sort of modal claim is not permitted on an etiological construal—it is only actual, not counterfactual, causal history that explains the present distribution of traits.[49] Some etiologists[50] are willing to bite this bullet—hearts

---

[47] Godfrey-Smith (1994) does this, but he incorporates the following two ways as well.

[48] Griffiths' (1992) notion of "an evolutionarily significant time period" may be of use here, but will not help with the problem I next discuss in the text.

[49] See § 4.111 for a discussion of how restricted the explanandum of a selectionist account should be. Even if, as I argue there, the explanandum may be legitimately broadened to include more than the mere prevalence of a trait, it should be able to handle at least this. Hence my discussion in this section is not undermined.

don't have the etiological function of pumping blood, because (assuming a dearth of variants) they have not been selected for doing this in the recent past.

A second route, one which would allow the etiologist to recognise vestiges yet attribute a function to the heart even if it has not been recently selected for, is to appeal to the fitness conferred on creatures in which a trait performs the relevant task. Hearts do us a great service by pumping blood. If they did not pump our blood, our fitness would be, to state the obvious, severely and adversely affected. This is doubtlessly true, but it does not explain the existence of current hearts. The fact that, if hearts did not pump blood, we would not live, does not show why hearts exist.[51] Some etiologists respond, sensibly enough, by appealing to *past* contributions to fitness.[52] The pumping of blood by ancestral hearts contributed to the fitness of those ancestors, making them more likely to pass on their genes. There is something right about this. The fitness of past tokens of a trait is at least a partial cause of the existence of tokens of that type in subsequent generations.[53] However, fitness, if it is not to collapse tautologically into the comparative notion of "selection for", does not explain the current *prevalence* of anything. Fitness is an aptness to be successful under selection, not such success itself.[54] It is *relative* fitness

---

[50] Godfrey-Smith seems to flirt with this consequence (1994, p.357). It is what taking what he calls "the hard line" would recommend.

[51] This is related to criticisms of appeals to function earlier this century. Adherence to the D-N model of explanation lead some people to think that, if the existence of hearts cannot be deduced from their capacity to pump blood and their function of pumping blood, then something has gone awry. Cf. Hempel (1965). Note that this is the explanandum considered appropriate for the etiological account— it is the D-N model that is rejected.

[52] Cf. Griffiths (1992). Walsh and Ariew (1996) seem to want to have their cake and eat it, too. They claim that a broader notion of evolutionary function subsumes both past selection and future selection. I think this is results in an unnecessary blurring of two distinct explanatory projects.

[53] For discussion, see § 4.111.

[54] See Mills and Beatty (1979).

that could explain, at least partially, the prevalence of hearts, by explaining why such traits were more successful than their competitors. On this view, which is, I think, on the right track, hearts have an etiological function, even though that function does not explain why *all of us* have hearts.[55]

The third avenue open to the etiologist who, unlike Aristotle, must handle vestigial traits and those whose function has changed, is to require that the trait still be able to perform the thing for which it was selected in order for that thing to be its function. (This requirement would have seemed otiose to Aristotle.) Thus, the heart has the function of pumping blood only if hearts still can pump blood.[56] Let me note one *prima facie* difficulty, before expanding on a second problem that will lead us into a discussion of the normativity of structural function. The *prima facie* difficulty is in cashing out the modal force of "can" in "can do the function in question". Is it true that our appendices *cannot* digest cellulose? Would it matter if it were false? Suppose (counterfactually, let us assume) that if humans were to eat grass, our appendices actually would aid us to digest it. In the absence of a more restrictive notion of "can", it would seem that this would render our appendices functional after all.

My more serious worry is that this requirement melds the ahistorical, structural function with the etiological in an unacceptable way. If there is a requirement with regard to current ability (assuming that can be satisfactorily cashed out), then our explanandum has changed. The fact that (at least) some ancestral hearts did in fact pump blood is causally relevant to the fact that cardates now exist. Their existence is not explained by

---

[55] This leads me to think that the "hardline" mentioned but rejected by Godfrey-Smith is the right one to take if one's only explanatory project is the prevalence of the trait in question.
[56] Godfrey-Smith (1994, p. 359).

the hearts that *currently* can, or even do, pump blood. Even if no current hearts could pump blood, the etiological story explains their existence by attributing to them an etiological function.

The second route, appealing to past contributions to fitness, does get us to the right sort of consideration in answering the question "why are there hearts?". The motivation for adding the condition that current tokens be able to perform what past ones did that contributed to their fitness runs deeper, I think, than just a recognition that vestigial traits must be accommodated. Attributing normativity to a current token strictly on the basis of its etiology is to say that it now ought to be doing something that its ancestors did, since doing that thing was useful to the ancestors. Pause for a moment to consider how odd this is. Imagine telling your teen-aged daughter that she ought to be more demure since being demure was beneficial to you.[57] Perhaps the fact that you are demure even accounts for her proclivity to be demure. However, if it is not in her current interests, then it is hard to see what normative force it could have.

To doubt this is not to doubt that you should have been demure. Note where the normativity lies—back in your history, not in your daughter's present. In fact, what normativity there is can be seen only if we step back in time and then take an ahistorical, structural view. You should have been demure because it served your interests then. This is not an attribution of current function, but an attribution of past current function.

[57] Perhaps it was even beneficial to your inclusive fitness, thereby accounting, in part, for your daughter's existence. Leaving aside the question of how heavily we should weight considerations of inclusive fitness in deciding how to act, it's not clear that her being demure would increase *her* inclusive fitness anyhow.

What gives etiological functions their normative cachet? It is, I think, the ease with which we incorporate past structural function into the etiological story. (Parents often think their children should act as they did!) This bias reveals a lingering Aristotelianism which is not so much in need of excising as it is in need of recognition. We know that the world changes far more radically than Aristotle thought it did. When considering normative attributions, we need to consider whether or not things have changed. We can explain why some items are the way they are, even though other things have changed, without being committed to the idea that such items *should* be the way they are.

Recognising this fact forces us to look at the *present* environment when speaking normatively. An attribution of dysfunction is inappropriate if the item in question is not in the sort of environment in which at least some tokens of its type could prosper. Recall our earlier discussion (e.g., § 3.1) of the intrinsic/extrinsic continuum that is to be considered in making functional attributions. Briefly, the idea is this: items in environments that are sufficiently inimical to the successful performance of the item's function are not malfunctioning when they fail to perform their function—the *environment* should be different. If the normativity is to fall to the item itself, as opposed to the environment, the environment must be, in some sense, appropriate for the item, yet the item be unable, for more intrinsic reasons, to perform its function, regardless. The intrinsic/extrinsic continuum is the continuum along which we can assign "blame" for failures to perform, for lack of success of the item.

Granted, it is far from obvious how we are to delineate the "appropriate" environment and hence situate any given example along the continuum. Etiological accounts that opt for a strictly historical demarcation of appropriate environment face the problems of accommodating changes in the environment and of allowing for vestigial traits. Structural theories, though in a position to take into account some historical facts when considering how to individuate environments, require that there be some possible "successes" of tokens of an item in an environment in order for that environment to be appropriate. Otherwise, there is no capacity of the item to submit to a functional analysis.

A further discussion of this difficulty is deferred to Chapter 6. In the present chapter, I have offered a diagnosis for the difficulties each of the approaches to teleological function, systems-theoretic and selectionist, has in meeting *all* the desiderata adduced in Chapter 3. I have argued that the cause of this is that their explananda are different, though not incommensurate. I have also argued that an ahistorical approach, specifically the structural one, is required for attributions of present normativity, and that an etiological approach presuppose the past applicability of a structural account. In the next Chapter, I return to issues in the philosophy of mind.

Chapter 5

Teleological Functionalism and the Philosophy of Mind

> ... if it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying .... if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world.
>
> Fodor (1989, p. 77, embedded ellipsis his)

In the previous chapter, I described two explanatory projects and argued that they require different senses of teleological function. I then described the relationship between these two projects and, hence, the relationship between the two notions of function. It is now time to return to the philosophy of mind to see how this discussion is relevant to issues there.

In Chapter 1, I introduced the general puzzle of the relationship between the contents of our mental states and their role in the production of our behaviour. One undertaking in the philosophy of mind has been that of trying to explain how it is that what we believe and desire can affect our behaviour. A natural first step in this undertaking is to spell out the individuation conditions for intentional content. One family of attempts to do this falls under the rubric of causal/information-theoretic accounts.[1] As I outlined in § 1.2, this sort of account soon runs into what has been

[1] Millikan is one theorist who makes appeal to teleological function but who does not advocate a causal/information-theoretic approach. I discuss her views in section § 5.21 below.

dubbed "the disjunction problem". The range of features or objects a state represents is counter-intuitively large and indeterminate, if any cause or information-theoretic source is considered germane. One ends up with an unwieldy and unpalatable disjunction. For realists, solving this problem requires rendering content determinate, while leaving its causal efficacy intact. In answering the disjunction problem, we (we realists, that is) do not want to have to deny the causal efficacy of content. This entails, I shall argue, that a selectionist construal of teleological function is inadequate for the task at hand.

My criticism of this approach does not depend on worries about lingering indeterminacy, or on the problems of "new", evolutionarily non-salient beliefs. Instead, I focus on the way in which this approach leaves content causally extraneous. Prior to doing this, I first explore a different tack, to distinguish my concern from others, and to show that certain common worries are misplaced.

§ 5.1 *What Is Not Wrong with Teleological Functionalism*

There have been a number of criticisms of teleological functionalism raised over the years, especially as it applies to issues in the philosophy of mind. In the course of laying out and evaluating the different accounts, I have discussed many of them. In the following subsection, I look at Fodor's concerns, showing that they can be met.

§ 5.11 *Fodor's worries*

Fodor argues against the use of teleological functionalism in individuating the contents of mental states (e.g., 1990b). Furthermore, although he construes teleological functionalism etiologically, he states that his arguments will hold for non-etiological accounts as well. In this section, I outline and criticize Fodor's arguments against teleological functionalism's handling of the disjunction problem and conclude that, although he sometimes misrepresents the etiological position, some of his criticisms find their mark. I then address his contention that he has effectively dealt with non-etiological accounts as well. I conclude that he has not. I leave what I think to be the genuine problem with etiological accounts for the next section.

Fodor claims that the usual way to attempt to handle the disjunction problem from the perspective of causal/information-theoretic theory is to distinguish between two sorts of situations, type one and type two, which correspond to the division between proper causings and improper causings. In type one situations, all causings would be proper, whereas in type two, none would be. In the former, "lawful covariation determines meaning" (1990b, p. 60), but not in the latter. In a type one situation, the content of the state that the frog tokens when presented with a fly *ipso facto* has the content <fly>. In type one situations, there is no error. In contrast, type two situations permit of error. In such a situation, our frog may well sometimes token "fly" when it is presented with something other than a fly (e.g., a beebee), yet the content of its mental state would not thereby be <fly-or-beebee>. Of course, our frog may be lucky and get things right even

when in a type two situation; type two situations are not those in which it is impossible that accurate mental states occur, but rather are those in which they are not guaranteed. By privileging type one circumstances, a principled distinction between type one and type two situations would allow for error while retaining the primacy of the external world/mental state relation in the fixing of content, as is postulated by causal/information-theoretic accounts.

Having set up the problem this way, Fodor goes on to canvass naturalistic explications of this distinction which, he says, "have proved to be a little thin on the ground" (1990b, p. 61). He discusses two sorts of approaches: a learning-based account (e.g., Dretske (1988)) and a teleological/functional account. He dismisses the former, roughly on the grounds that one must identify learning situations with type one situations and that this will not work; if error is to be possible post-learning (i.e., in type two situations), then it must have been possible during learning, given what little we know about learning. If it were possible during learning, then it is only good fortune, or the intentions of the teacher, that prevented erroneous tokenings. Good fortune is hardly robust enough to ground this distinction, and appealing to the intentions of the teacher is circular.

In the case of teleological accounts, the general scheme, as Fodor sees it, is to identify type one situations with Normal situations, where "Normal" is a technical term and carries normative or evaluative force.[2] Fodor has in mind a selectionist etiological

[2] As the reader will recall, the term "Normal" comes from Millikan (1984). It is capitalized to distinguish it from "statistically common", or "usual". A situation is Normal if it is the sort of situation in which the object of our attention has historically been successful. Thus, the Normal condition for a sperm is such that it penetrates an ovum, even though, statistically, very few sperm are ever in such a situation. Fodor claims to be following this convention, although there

schema such as RM or GS. A thing is behaving Normally if it is behaving in the way in which its ancestors behaved that caused those ancestors to be selected. For example, a heart is behaving Normally when it is pumping blood, since the pumping of blood is what its ancestors were selected for doing. Pumping blood is the heart's "proper function".[3] Normal conditions are those conditions that obtain when an item performs its proper function.

In the case of mental items, the story goes, they too are performing Normally when they are doing whatever it was that their ancestors did that caused them to be selected. In the case of beliefs, Normal conditions (Fodor's type one situations) are arguably those in which it is *true* beliefs that are generated.[4]

§5.111 *Fodor's "Digression" on Millikan*

At this point, Fodor turns his sights on Millikan, a leading proponent of teleological functionalism. However, on his interpretation, her views make a somewhat different use of the type one/type two distinction than the one that Fodor initially suggests and that I

---

are times his grasp of the concept seems a bit tenuous (e.g., 1990b, p. 67). See Chapters 2 and 3 for more discussion of the selectionist approach, and Millikan's variety of it.

[3] "Proper function" is another term from Millikan (1984). As should be clear from the context, it refers to that effect for which the trait in question was selected. Again, see Chapters 2 and 3 for more discussion.

[4] I am being deliberately vague about what sorts of items are candidates for Normal behaviour; the most natural reading is that it is mental *mechanisms* that have proper functions but, as we shall see, Millikan holds that it is also mental *states*.

Also, it is far from clear that beliefs would be selected in virtue of their being true, or that, *mutatis mutandis*, belief-fixing mechanisms would be selected in virtue of their production of true beliefs. For example, as Fodor notes (1991, p.294), false beliefs may be advantageous, if Freud is right about repression.

have outline aboved.[5] In her theory (according to both Millikan and Fodor), a type one/type two distinction serves as part of a ground to the notion of proper functions—Normal conditions are those that are cited in a Normal explanation of the presence of an item, by means of their role in the performance of an item's proper function. It is this latter, the proper function, that, in the case of mental states, determines content for Millikan, according to Fodor. When a situation is Normal, it is, according to Fodor, the proper functioning of a mental state determines its content.

Fodor's attack consists of two separate claims, although he tends to run them together. First, he denies that mental states can have proper functions. Second, he denies that reference to proper functions is sufficient to individuate intentional states. I shall try to keep these worries separate.[6]

First, Fodor argues that the etiological story of teleological functionalism does not require that mental states themselves have a proper function. It is the *mechanisms* of, say, belief formation that are selected, not the belief states themselves. Granting a selectionist story[7] of the existence of mechanisms of belief formation does not, argues Fodor, commit one to thinking that mental states have functions.

I think Fodor is a little too quick here. It is true that the heritability of traits plays a central role in any selectionist story that purports to establish their function. Variation, heritability, and differential reproduction are essential features of a selectionist account.

---

[5] I think that Fodor is right to contrast Millikan with other writers who make appeal to teleological function, but on the grounds that her conception of explanation differs. I return to Millikan's views in § 5.21 below.

[6] I realize that the two concerns are connected; if the first one is successful, the second one is moot.

[7] I have serious misgivings about the hyper-adaptationist assumptions that underlie this view, but I shall not rehearse them here.

Belief contents or states themselves are not heritable,[8] so the relevance of selectionist considerations is in need of careful explication. The etiologist must give us reason to think that the considerations that establish a mechanism's function follow through to the output of such a mechanism.

Such a reason might go something like this: a given mental mechanism is selected only by means of its outputs, since its selective value is determined by its effect on the organism's behaviour.[9] The mechanism helps form the belief, but it is the belief state itself that helps regulate behaviour. After all, it is the heart that beats, but its "output", pumpings of blood, are what evolutionary theorists think gave it a selective advantage.

Millikan (1986) offers a good example. She notes that it is a proper function of the mechanism that allows chameleons to match their colour and pattern to that of the background on which they rest, to provide camouflage, thereby offering some protection from predators. It is also, she argues, a proper function of any given colour-and-pattern state to provide camouflage at that time from possible predators; "the proper functions of the pattern are to make the chameleon invisible and to prevent it from being eaten—functions that it derives from the proper functions of the mechanism that produced it" (1990b, p. 57). Fodor is right in claiming that the view "that the mechanisms that make/use cognitive states have functions does not *entail* that cognitive states themselves

---

[8] I am ignoring the issue of innate "beliefs", such as those demonstrated by infants' reactions to a visual cliff. There may be many innate beliefs (or proto-beliefs) that structure our perceptions and instinctive reactions, but most of our beliefs simply are not innate. Even Fodor, who has thought that a great many concepts are innate, has not maintained that, e.g., believing that my desk is cluttered is innate.

[9] This is a gross oversimplification—not only is a thing's selective value a relative measure, dependent on the other contenders in its environment, it is also a measure that must take into consideration a thing's cost. A wonderfully beneficial trait will have little selective value if the costs of maintaining the trait are large. See Chapter 3.

do" (1990b, p. 66, italics mine). Even so, he has not thereby shown that such states *cannot* have functions.

Fodor's second worry is a more serious threat to Millikan's position (as he understands it). Even if mental states should have proper functions, reference to them is insufficient to individuate their content. Fodor puts the point this way:

> It's a sort of distributive fallacy to argue that, if having beliefs is functional, then there must be something that is the distinguishing function of each belief. The function of the human sperm cell is to fertilize the human ovum; what, then, is the distinguishing function of *this* sperm cell? The hair on your head functions to prevent the radiation of your body heat; what, then, is the distinguishing function of *this* hair (or, for that matter, of *red* hair)? (1990b, p. 66)

Rather a lot is going on here. As I mentioned above, Fodor is conflating his two worries, as this passage illustrates. He seems to be re-iterating his contention that individual belief states do not have proper functions, but he is also saying something more—namely, that knowing something's proper function is not sufficient to distinguish it from other tokens of the same type.

However, to get this crucial point from Fodor's argument, one must navigate carefully. If this passage is taken as an argument against attributing proper functions to states, several things lurk in the fog. For instance, the situations he discusses are not sufficiently analogous. The distinction between human sperm and a particular human sperm is not the same as the distinction between a mechanism and its output. The former is one of type–to–token, and the etiologist has no more problem than anyone else in

saying that the function of a token is the same as the function of the type. The function of this sperm is the same as the function of any other sperm, since they are all of the same type.

When Fodor asks what the *distinguishing* function of a token is, his discussion allows for an ambiguity.[10] On one reading, he is asking, in the terms of our above example, what makes this thing a sperm? On another reading, he is asking what makes this sperm *this* sperm, as opposed to another one? The etiologist can appeal to selectionist elements in answering the former question,[11] but not in answering the latter. However, this may not indicate a flaw in the etiologist's position; the task of individuating tokens of the same type is simply not the purview of selectionist/etiological theory. Spatio-temporal facts and arcane considerations in identity theory are the sorts of things that are germane. Besides, it is often taken as a virtue of the etiological approach that it appeals to types, not tokens (e.g., Neander, 1991). Given that evolution is a feature of populations, not individuals, a theory that restricts itself to types, not tokens, is *prima facie*, the etiologist argues, to be preferred.

Nonetheless, if the proper function of a mental state determines its content, and proper functions of states are derived from the proper functions of the mechanisms that fix them, then there must be an explication of the relationship between type of mechanism and type of output. If the relationship is one-to-one, there must be a great

---

[10] If etiologists were genuinely guilty of some sort of distributive fallacy, then the word "distinguishing" should be omitted from the passage in Fodor just quoted. Also, his mention of red hair is off-track, if our discussion is of humans. The etiologist (and anyone else, I suppose) would say that the *redness* of some human hair has, alas, no function.

[11] I am not claiming that taxonomy in biology is simple; elements of history, morphology, and function all play roles. My point is merely that it may be appropriate to invoke etiological considerations.

number of types of mechanisms, if anything like the variety of belief attributions we now make is to be vindicated. The mechanism that leads me to form the belief that something dangerous is nearby must be different from the mechanism that leads me to form the belief that something edible is before me, and so on, for every belief type. This sort of one-to-one relationship seems plausible in the case of non-intentional phenomena—the heart produces heartbeats, for instance. All heartbeat tokens are of the same type. No further individuation is necessary for physiology or medicine. Heartbeats, though, are not *about* anything at all. A one-to-one relationship of mechanisms to output is not so plausible in the case of the mental—admissible mental mechanisms seem to be things like belief-fixers and desire-formers. Not all my beliefs are of the same type, except in the trivial sense of all being beliefs. Unlike heartbeats, further individuation—*intentional* individuation, individuation in terms of content—is required.

If the relationship between mental mechanism and output is not one-to-one, as is more likely, then the route from individuated mechanism to individuated belief is, to put it mildly, not clear. Presumably, my belief that my cat is sitting on the fence is formed by the same mental mechanism (viz., my belief-fixer mechanism) that allows me to form the belief that my dog is sleeping under the table. Even if these belief states have proper functions, those functions do not, it seems, distinguish them. Their proper function, derived from the proper function of the mechanism, is, let us suppose, to be accurate representations of the world, thereby guiding my behaviour in life-preserving ways. Their proper function is to be true.

Yet, if this is so, what makes one of my beliefs a belief about my cat, and another a belief about my dog? How could anyone have thought that an appeal to proper function would suffice as a response to this question? Answer: nobody has. Not even Millikan.[12]

I think that the sensible way to read Millikan is to highlight her appeal to "*the right circumstances*" (1986, p. 64, italics hers) under which the proper functions of states are performed. It is these right circumstances, in tandem with a mental state's proper function, that are alleged to determine content—not a state's proper function in isolation.

This reading of Millikan brings us back to Fodor's original concern with distinguishing type one and type two situations. It also shows that Millikan is not so much a digression as she is part of the ongoing debate. If she is wrong (and I think she is), she is wrong in a different way than Fodor makes out.

§ 5.112 *The Type One/Type Two Distinction and Opaque Contexts*

Recall that Fodor's initial concern was to delineate type one and type two situations. He returns to this issue by offering another version of teleological functionalism in which the focus is on the environments in which mechanisms historically performed. Those circumstances in which a mechanism was selected for are to be considered Normal, whereas those in which it was not, are abNormal. The former are type one situations and the latter type two. On this view, information about type one situations provides us with

[12] Just for the record, I am not surprised that Fodor has interpreted Millikan the way that he has. She often seems to be saying just what he claims she is.

grounds for assigning content. This, I think, is more in keeping with Millikan than Fodor's reconstruction would allow.

At any rate, recall the earlier story about the frog and the fly. The selectionist approach under consideration would claim that it was because of the fact that the frog's mental state, upon seeing a fly, had the content that it did that the mechanism that helped bring about that state was selected. This content was <fly>; if the content had been, say, <predator>, then frog bearers of that mechanism would almost certainly have died out.

Fodor claims that this approach, an etiological selectionist one, cannot solve the disjunction problem but rather leaves content indeterminate. In the frog case, having a mental state with the content <fly>, and having a mental state with the content <fly-or-beebee> would have been equally selected for. Since, in the environment of our frog's ancestors, all flies were flies and all flies-or-beebees were flies, either would have played the same role in the frog's snapping at flies. "Darwin cares how many flies you eat, but not what description you eat them under" (1990b, p. 73).

The difficulty, according to Fodor, is that the intensionality of intentional contexts is not reflected in an etiological account; appeals to causal history seem to allow for the substitution of co-referential terms, *salva veritate*. If it is true that the eating of flies was causally relevant to the survival of frogs, and all the flies eaten by frogs were of the kind fly-or-bee-bee (which, of course, they were), then it is true that the eating of flies-or-bee-bees was causally relevant to the survival of frogs. In the case of appeals to selection, Fodor writes "the context: *was selected for representing things as F* is transparent to the substitution of predicates reliably co-extensive with F" (1990b, p. 73, italics his).

I think this is wrong. Causal history is not transparent. Things are selected *in virtue of* certain of their properties.[13] In Fodor's terms, my claim is that Darwin cares about *properties*, not *objects*. Of *course* Darwin does not care about different descriptions of objects. Crudely put, natural selection cares about certain of the properties things have, not what the things "are" or how they are described.[14]

Fodor's discussion at this point in his attack on appeals to teleology is illuminating of his overall approach and of its problems. He considers, but rejects, an appeal to counterfactuals as a way to make content determinate for the teleologist (e.g., "what *would* happen if the locally reliable coextension between flies and flies-or-beebees had been broken" (1990b, p. 75)). He argues that there are three problems with this approach. First, it has been explicitly rejected by Dennett (1987) and Millikan (1984). Second, it is difficult to decide, in a non-question-begging way, just what the relevant counterfactuals are. Third, this approach diverges from Darwin at the crucial juncture—counterfactual circumstances play no role in actual selection history.

Fodor's first reason, that appeals to counterfactuals have been rejected by the likes of Millikan and Dennett, is no reason, in and of itself, to reject the view. I think this worry is subsumable under worry number three—etiological selectionist accounts such as those of Millikan and Dennett have no business countenancing counterfactuals, or so it would seem. It is considered to be a virtue of the etiological selectionist approach that it

---

[13] Considerations of this sort lead me to think that, if a frog has mental states in anything like the way we do, then, on an etiological construal, the content of its mental state upon perceiving a fly is <food>.

[14] I do not mean to downplay the general problem of ascertaining the legitimacy or "projectability" of would-be properties. (See, e.g., Goodman (1954).) Would that I could solve that problem here. My aim here is to re-iterate and show the relevance of the claim that selection operates on certain features of organisms or traits, and not on others.

deals in *actual* causal history, rendering counterfactual issues seemingly otiose. This is part of what is supposed to render an appeal to teleology scientifically respectable. Fodor is right—counterfactuals themselves play no role in actual selection history. The etiologist who appeals simply to the objects germane to a causal/evolutionary history cannot rely on counterfactuals to handle the disjunction problem.

This objection does not rule out two things: i) teleological accounts that are not etiological, and ii) etiological accounts that make explicit appeal to properties. In the first case, Fodor thinks that "any naturalistic story about teleology is going to rest on some sort of appeal to evolutionary history" (1990b, ftn.17, p. 85) but, as I am in the process of arguing, this need not be so. Systems-theoretic approaches to teleological function are both legitimate and required. In the second case, the use of counterfactuals can be a means of discerning etiologically relevant properties. It is not the counterfactuals *themselves* that are selected, of course, but rather the causally relevant properties that counterfactuals help us to discern to be such, or are true of.

So let us leave that objection to one side as a worry for etiologists of a particular bent—Fodor has scored a hit. However, it may be permissible to appeal to counterfactuals *if* one's teleology does not rest on etiology, or if one's etiological teleology deals in properties. This leaves us with Fodor's second objection, viz., that there is a problem in determining just which counterfactuals to entertain in solving the disjunction problem.

§5.113 *Content and Counterfactuals*

My task in this section on Fodor is negative: I aim only to show that something is left standing after Fodor has taken his shots. At this point, still standing is the possibility of an ahistorical account of teleological functionalism, and some as-yet-untouched method of appealing to counterfactuals.

Given Chapters 2 and 3 above, I hope that the notion of an ahistorical account of teleological functionalism does not strike the reader as completely implausible. Fodor attempts to characterise such a view, but in a particularly misleading way. He combines his worry about how to appeal to counterfactuals with a claim that the following method is not historical:

> ...one might just give up on the reduction of content to selection history and try for a *nonhistorical* theory of content; one in which content is determined not by the selectional pressure that *actually* governed the evolution of a psychological state but by the selectional pressure that would apply if certain counterfactuals were to apply. E.g.: Either [sic] fly-snaps and fly-or-bee-bee snaps are equally advantageous in *this* world. But the intentional objects of frog snaps are flies and not-flies-or-bee-bees because fly-snaps would be selected in nearby worlds where there are flies *whether or not there are bee-bees there* but fly-or-bee [sic] snaps would not be selected in nearby worlds where there are bee-bees *unless there are also flies there*. (1990b, pp. 76-7, italics his).

Of course, his is not the sort of ahistorical view I have in mind, since it is couched in terms of objects, not properties, and is "backward-looking", as opposed to appealing to

capacities, which are importantly "forward-looking".[15] This view is some sort of selectionist view, but with an implausible method of ascertaining the relevant selection history.

Fodor's prognosis for an "ahistorical" theory of the sort he outlines is grim, but not for the reasons one might think. He seemingly abandons his worry about ascertaining which counterfactuals ought to be entertained, and he goes so far as to grant that such counterfactuals are plausibly exactly what matter in ascribing function; this grounds his castigation, in the same article, of Dennett and Millikan for their perceived inability to accommodate such counterfactuals. However, he denies the relevance of an appeal to function in determining content, and not because such an appeal leaves content indeterminately disjunctive, as was the case, according to him, with straight-forwardedly selectionist etiological accounts.

What ends up being the crucial issue is not the disjunction problem at all, but rather the worry that the *immediate* causes of mental states are not always (or even often) their objects. Fodor's "glaring counterexample [to teleological functionalism] is the occurrence of representation in *thought*" (1990b, p. 80, italics his). Some thoughts are not caused by their objects, but instead are caused by other thoughts; some of our thoughts cause other of our thoughts, and these latter are not thereby *about* the former. So causal/information-theoretic accounts are in need of a story that connects contents of thoughts with something other than their immediate causes. What that story could be is not clear, but what *is* clear is that this is not the disjunction problem, as normally

---

[15] Bigelow and Pargetter make use of the "backward-looking/forward-looking" distinction in their analysis of functions as propensities. See Chapters 2 and 3 for discussion of the propensity theory, in both its selectionist and systems-theoretic guises.

construed. Rather, it is some version of the depth problem; where, on the causal/information-theoretic chain, does one stop? This is a genuine problem, but not one peculiar to teleological approaches.

To conclude this section, let me rehearse what is not wrong with teleological functionalism. If (as I have been arguing, and Fodor allows) teleological functionalism need not rely on selection history, it need not rely on implausibly strong adaptationist assumptions; and, if it need not rely on selection history, it can also make use of counterfactuals, rendering content determinate, at least in the standard cases. In fact, even if it appeals to selection history, there is arguably an ahistorical underpinning for this process. Such an underpinning, if cast in terms of specific sorts of properties, can make sense of an appeal to counterfactuals. Can it help with the problem of representational symbols that are not caused by their objects? Very proximal causes are the wrong places to look for content for two reasons. One, error is rare if not impossible, at extreme closeness. Two, our successful interaction with the environment cannot be explained without reference to that environment. If we never think about what is happening external to us, it would be a mystery if our thoughts aided us in dealing with our environment. Furthermore, perhaps, with sufficient care, the notion of "the right conditions" can be spelled out so as to permit the individuation of contents in other than quasi-perceptual states. Even so, as just noted, these problems are not peculiar to teleological approaches to the disjunction problem. Their roots lie in information/indicator theories themselves.

§ 5.2 *Explanation in the Philosophy of Mind*

In this section, I explore the nature of psychological explanation, with the goal of motivating the afore-mentioned restriction of an adequate account of the content of mental states to one that is synchronic, as opposed to diachronic. The criteria that one thinks must be met for an account to be adequate are determined, in part, by the sort of project in which one thinks an appeal to intentional states is legitimately embedded. What are we doing when we appeal to intentional states? When we offer a psychological explanation of a thing's behaviour that includes reference to the contents of its intentional states, of what relevance is the content? Answers to these questions constrain what is to count as an adequate theory of intentionality. My view is that both realism and pre-theoretic intuition require that, in giving an intentional explanation, one is giving a causal explanation. Furthermore, in giving a causal explanation, one is asserting the relevance of the intentionality of the states invoked. However, before sketching my view about what a realist must say about a good deal of intentional psychology, I turn to Millikan's contrasting view of what should form the core of psychology.

§ 5.21 *Explanation in Millikan's "Biopsychology"*

Millikan has a particular view about the nature of the sort of explanations in which appeal to teleological functions would be appropriate. Her analysis of teleological functions is, the reader will recall, a selectionist etiological one. (See "RM" in the appendices.) I have

argued above that a natural domain for such etiological concepts is in explanations of why something exists with roughly the form or properties it has. A closer look at Millikan is warranted, because this sort of explanation is not (at least not straight-forwardly) the sort she takes her project to provide, and she also has strong demands on what sort of explanations may be considered legitimate in psychology and prohibitions against the use of teleological functions in causal explanations.

Millikan writes that "the core of the science of psychology should be a study of teleofunction" (1993b, p. 174).[16] She dubs this core "biopsychology", claiming that it concerns "the functions of the mechanisms that regulate those life processes, those links in the life chain from generation to generation, that are completed through the mediation of behaviours" (1993b, p. 175). Millikan reminds us of something we noted earlier — many items with biological functions fail to perform their functions, and this is so for reasons ranging from dysfunction to unfortunate circumstances. The performance of a function of a thing need not be statistically common, let alone invariant. Millikan suggests that wide-spread failure to perform functions should be expected in biopsychology, given that successful functioning requires "mediation by the environment" (1993b, p. 176).[17] There is little reason to expect the reliable performance of many biological functions; there is even less reason to expect the reliable performance of those that require the 'co-operation' of the environment external to the organism, such

[16] "Teleofunction" is Millikan's neologism for etiological function. Formerly, she used "proper function" in the same way. I shall avoid the later terminology because it seems to preclude any but an etiological construal of function from being teleological.

[17] Millikan writes that cognitive failure is "ubiquitous". I do not think she intends this to mean that failure could be universal. In § 6.2, I discuss the requirement that there must be some instances of success in order for there to be functions.

as cognitive functions. This fact, thinks Millikan, has serious repercussions for the nature of explanation.

Millikan stipulates that biopsychology is *not* concerned "to examine or speculate on details of the complex structure inside the black box, to check on the dispositions of the components, nor to examine how the little dispositions inside add up to the complex dispositions of the whole" (1993b, p. 177). Though she acknowledges that such projects have concerned and continue to concern some psychologists, she argues that they ought not to be part of the core of the discipline.

She offers two examples. First, there is the area of industrial and media psychology. Here, the primary concern is with the average dispositions of people, not what their dispositions *should* be. What organizational structure is most conducive to productivity? How best can one sell a car? The second example is characterised by its methodology—it is one in which animal experiments that involve inflicting "extreme deprivations" on the subject are performed.[18] Millikan is not clear about what questions are being addressed by these sorts of experiments, though she claims that they are "done under the flag of behaviorism" (1993b, p. 178). Presumably, their salient feature is that they do not speak directly to how things *ought* to behave. Industrial and media psychology merely describe average behaviour (then, perhaps, prescribe ways to elicit desired behaviour). In destructive experiments, the investigation concerns what happens when things are mutilated or placed in hostile, alien environments.

[18] Millikan, with great rhetorical effect, invokes the Nazi experiments on Jewish prisoners. Exactly what the rational import of this comparison is intended to be is not clear.

Millikan tells us that such projects view their subject matter in an inappropriate way. In alleged contrast, the "biopsychologist's study has little to do with averages over chunks of living matter" (1993b, p. 178). This crude characterisation of the comparison class leaves something to be desired— which would-be psychologists study such things? Even industrial psychologists and Nazis would hesitate to describe their enterprises in such disparaging terms. Gathering statistical data can be a means of learning about causal and other mechanisms. Using invasive and disruptive experimental techniques is entirely compatible with viewing one's subject matter as an intact organism that is, for the moment, outside its usual environment. To find out how something operates normally (in Millikan's sense of "normally"), it may be necessary to tamper with it.[19] Surely the use of Mill's method of difference is not biologically suspect. I think Millikan has an oversimplified view of what motivates destructive experimentation.

Perhaps it would be more helpful to look at Millikan's positive account of her approach. She writes that the "chunk of living matter" isolated at a single moment in time on a table in a lab "represents cross sections from a numerous set of loosely coordinated intertwined parallel processes, each having strands of its own, each developing through its own inner logic and at its own pace in rough harmony and interchange with the others ... The biopsychologist's study concerns only the central unbroken strands of this fabric, and each fibre in these only so far as it has spun itself out in a principled historically precedented way." (1993b, pp. 178-9). Less fancifully, she writes that "psychology is the study of processes" (1993b, p. 179), and that those engaged

[19] The moral ramifications of some animal experiments are beside the point in this discussion.

in such a study ought not to expect the relevant processes to be contained within the organism's skin. "To study an animal's behavioral systems without at the same time studying the normal integration of these into the environment, without studying the loops through the environment it is the function of these systems to initiate, would be exactly like studying the digestive system without considering what normally passes through it." (1993b, p. 180).

This is a salient point, though not perhaps as radical or novel a one as Millikan implies. Nonetheless, it is crucial and serves to divide current theorists in the philosophy of psychology into two camps. The issue, to use terminology different from Millikan's, is that of individualism, or internalism versus externalism in the philosophy of mind.[20] There are those who think that psychological subjects can be studied and mental states individuated without reference to the environment, whereas there are those who think that reference to the environment is crucial. Millikan clearly falls in the latter camp, as the metaphor of loops and strands is intended to indicate.

Within the externalist camp, there are further divisions. Having presented a case for externalism,[21] Millikan immediately characterises it as requiring an historical, selectionist component. This, of course, is in keeping with her commitment to a selectionist etiological analysis of function—functions are determined by a thing's selection history. Requiring this historical component is not an illegitimate move on her part, given her arguments elsewhere. However, it is important to keep the notions of

---

[20] Fodor (1980) and Stich (1983) both adopt (different versions of) methodological solipsism. This term comes from Putnam (1975), who is critical of the notion. See § 6.12 for some discussion of the Twin Earth argument for externalism.

[21] I am unsure whether to consider Millikan's presentation an argument for externalism or more of a manifesto.

externalism and etiology conceptually separate. One could be an externalist, yet not place the same demands on history that Millikan does.

From this understanding of the business of biopsychology, Millikan draws some substantive conclusions about its explanatory goals. She writes "[t]hese reflections on the nature of intentional psychology entail that, as a biological science, it does not aspire to be predictive" (1993b, p. 182). I think that the crux of this issue is whether or not biopsychology makes essential reference to causal mechanisms, but let us first investigate the two reasons that Millikan explicitly offers for making this claim: "The first is diversity among individual constitutions" (1993b, p. 183). From the fact that there are sufficient innate differences between humans to make it the case that "different newborns inserted into identical environments would not behave at all alike, unless under the most general and vacuous of descriptions" (1993b, p. 183), Millikan concludes that prediction is simply not in the cards. She bolsters this claim by noting that it is likely that many of our cognitive processes may be partially stochastic, and hence that which action we perform in a given case "is not governed by well-defined psychological principles at all" (1993b, p. 183).

Millikan's notion of prediction must be extremely rigourous. Requiring that there be no intrinsic differences between tokens of the same type of objects of study would, presumably, put most of science out of the business of prediction. The task of casting our descriptions at an appropriate level of generality is not easy, but Millikan has not demonstrated that it is a task that cannot be done in the case of human psychology.

The second reason that Millikan offers for the claim that psychology does not aspire to be predictive[22] is the afore-mentioned prevalence of cognitive failure. Since cognitive functions "loop" through the environment, and the environment is a changing and sometimes uncooperative place, often cognition fails. "Because this is so obvious, it has been equally apparent to all that there could not possibly be any reliable laws of organism–distal environment interaction, certainly not for the case of humans" (1993b, p. 183). I think Millikan overstates her case. Aside from not being apparent to me, it has not been apparent to, for example, Fodor, who writes "it's hard to doubt that at least *some* psychological regularities are lawlike (for example: that the Moon looks largest when it's on the horizon ..." (1994, p. 3). It is true that the occurrence of failure to fulfill function is relevant to prediction, but it is not clear how it *prevents* prediction. If failure to fulfill function were impossible, the specificity and degree of certainty of our predictions would be greater, but this impossibility is not a prerequisite of prediction.

Millikan writes "[c]learly, from the fact that drawing rational inferences may be a norm for the human cognitive systems, it does not follow that any reliable predictions about inference patterns can be made" (1993b, p. 184). This may be so, but it also clearly does not follow that such predictions are not made, let alone cannot or should not be made.[23]

---

[22] At this point, she moves to what appears to be a modal claim— "psychology *cannot* be required to predict individual behaviours ..." (1993b, p. 183, italics mine). I think she intends this to be read as a prescription, not as one about possibility.

[23] I turn below (§ 5.22) to my concern that Millikan is ignoring important explanatory, as opposed to predictive, work done by intentional psychology.

Millikan attempts to explain away the seeming predictive success of folk psychology not so much by denying these successes,[24] but by offering an alternative explanation of them. In making quotidian predictions, we are not employing a theory that appeals to causal mechanisms such as beliefs and desires. Rather, thinks Millikan, we use "the method of brute correlation" (1993b, p. 185). Our knowledge of past regularities grounds our present predictions; our predictions are coarse-grained enough not to require knowledge of any causal mechanisms.

To buttress her claim that intentional psychology is not predictive, and to delimit further what she takes the explanatory task of psychology to be, Millikan denies that psychology appeals to causal relations at all. In other words, she denies the legitimacy of the systems-theoretic attribution of function and its attendant decomposing/capacity-analysis explanation of how something performs that function. There is little room, on her view, for such concerns. As noted in § 4.2, Millikan considers "function" importantly ambiguous and allows that a Cummins' style functional analysis will attribute functions independently of history. However, these functions are not relevant to the explanandum she considers to be the primary domain of biopsychology, with one exception. Those Cummins' functions "defined relative to the life cycle of the species" (1989b, p. 176) have a restricted role to play in biopsychology, since they, in contrast to all other non-selectionist functions, *do* explain the presence of the item that is functionally

[24] Denying these successes is the line taken by such theorists as Churchland (1986), and Stich (1983).

characterised. They are what explain the ontogenetic developmental history of items, which goes some way to explaining their presence.[25]

Since, on Millikan's view, true ascriptions of teleological function make essential reference to historical facts, and "having a certain history is not, of course, an attribute that has 'causal powers'" (1993b, p. 186), it follows that psychological states or processes, individuated in terms of their teleological functions, are causally inert; "that a thing has a teleofunction is a causally impotent fact about it" (1993b, p. 186).

Millikan takes two routes to defend this position: a negative attack on the view that causal laws are preserved by appeal to *ceteris paribus* clauses, and a positive account of an alternative. Millikan ties causality to causal laws,[26] and causal laws to prediction. Against the view that causal laws can be framed because a thing will perform its function *ceteris paribus*, she argues that (i) this is to ignore defective tokens, and (ii) performance could be very rare, entailing that *ceteris* is very rarely *paribus* (from 1993b, pp. 183-4). The existence of defective tokens and the rarity of successful function even in the case of non-defective tokens make prediction an unreliable business. This is evidence, according to Millikan, of the dearth of causal laws, which in turn is reason to think that causal explanation is not the goal of psychology. I will not address her concerns directly. Instead, in the following section, I offer an argument for the conclusion that the coherence

---

[25] Millikan may need to consider ontogenetically defined functions to be genuinely teleological, even when they have not been selected for. I do not think this is a problem, but she might not agree. Obviously, I am more of a pluralist about the nature of teleological functions than she.

[26] Millikan relies upon a notion of causal law as a ground for causal relevance and efficacy. My criticisms of her approach do not require me to question her on this, but it should not be assumed that I agree.

of her own enterprise requires that psychology be in the business of offering causal explanation.

For now, let us leave aside Millikan's arguments against intentional psychology offering causal explanations and focus on how it is that she think psychology does explain. Following Pettit (1986), she terms the relevant explanations "normalizing explanations". "Intentional-attitude explanations of behaviors proceed ... by subsumption of behaviors under biological norms rather than laws and/or by noting departures from these norms and perhaps causes of these departures" (1993b, p. 187). Thus, in the simplest case, we explain "the occurrence of a phenomenon by reference merely to something whose function it was to produce that phenomenon" (1993b, p. 188). This is a case of the sort of selectionist explanation explicated in Chapter 4 above. We can explain the existence of a trait, or the occurrence of an event, in terms of its being the result of a certain sort of causal history. In more complex cases, we "tell or implicitly refer to the place an event has in a series or interdependent pattern of functions, or tell where and perhaps why malfunction occurred within such a series or pattern" (1993b, p. 188), That is, we describe *how* one thing functions to bring about another and, sometimes, explain malfunction in terms of a failure within this complex series or systems. Thus, Millikan acknowledges that complex normalizing explanations make implicit reference to causal explanations. By making reference to components and their success or failure to perform their functions, a causal story is being assumed. Millikan does not take these latter, despite appearances, to be variants on systems-theoretic

explanation. I think that her discussion is reminiscent of the view I advocated in Chapter 4 on the relationship between the structural and the etiological. Millikan would not agree.

The relationship that Millikan says obtains is this: "[b]y making implicit reference to such causal explanations [i.e., to those that explain historical success], normalizing explanation may thus circumscribe quite specific physical explanations ..." (p. 190). In giving a normalizing explanation, one is noting that a certain behaviour is relevant to explaining the existence of a thing *and* alluding to the capacities that ground this behaviour. In doing this, asserts Millikan, one is narrowing the range of sorts of non-functionally characterised behaviours or properties that one thinks relevant. One is offering a sort of filter explanation; this entails that knowledge of the nature of the filter and the organism will circumscribe the range of ways in which the capacity to pass through the filter may be realised.

Perhaps surprisingly, this view is amenable to my own. Allow me to explain in the context of presenting what I take to be the nature of realist intentional explanation.

§ 5.22 *What Realism About Intentionality Requires*

In this section, I argue that an ahistorical or synchronic account of the individuation of intentional states is required for realism about intentionality and intentional explanation. I contend that this is so on the grounds that realism requires granting that the states in question have causal efficacy, and a historical or diachronic account of their individuation

either precludes this entirely (as Millikan maintains) or renders it unnecessarily and unacceptably mysterious.

§ 5.221 *Intentional Explanation is Causal Explanation*

A full-fledged defence of the view that realism about intentional psychology requires individuating content in such a way as to allow for its causal efficacy is in one important respect beyond the scope of the current thesis. This is because such a defence must in part rely upon the integration of this view with a completed theory. Even so, it is possible, at this early stage, to marshall some essential, preliminary support by appealing to pre-theoretic intuition, introspection, and what Kim calls 'Alexander's Dictum', the claim that "[*t*]*o be real is to have causal powers*" (1993, p. 201).

Pre-theoretic intuition is a difficult thing for philosophers to invoke with any degree of confidence in two aspects: (i) which judgements such intuition supports, and (ii) how heavily such "evidence" should be weighted, if at all. Reading and writing philosophy is almost guaranteed to alter one's initial intuitions to a certain extent, at least if the philosophy is any good. Philosophers have been exposed to too many competing theories and considerations for their intuitions to be "pre-theoretic". Even without this worry, it is unclear what is to be gained by invoking such intuitions. That a belief is common is not much of an epistemic recommendation.

Despite these reservations, I think that pre-theoretic intuition holds that our behaviour is often causally explained by appealing to the content of our intentional states

and that this intuitive judgement is *prima facie* relevant to our work in the philosophy of mind.

When we explain that Fred is going to the kitchen because he wants a beer and believes that one is there, we certainly seem to be offering a causal account of Fred's behaviour.[27] Had his belief or desire been different, had they had different contents, he would have behaved differently. We are not simply offering a rationalisation of his behaviour, or using belief and desire attributions for instrumental purposes. To deny that his beliefs and desires play a causal role in his behaviour is counter-intuitive.

Of course, many things are counter-intuitive but nevertheless true. Perhaps the intuition that we are offering causal explanations when we appeal to intentional states is one such thing. However, one needs good reason to deny this intuition. If two theories accommodate the phenomena equally, but one violates pre-theoretic intuition, the other is to be preferred. If we cannot accommodate the phenomena without violating pre-theoretic intuition, then violate it we must, assuming that all reasonable options have been investigated.

The results of introspection are another dangerous (nowadays) source of support for philosophical positions. Even so, it remains true, in my own case at least, that what I do seems to be affected by what I believe and what I want. Deliberating affects my subsequent behaviour (though perhaps not as often as it should). Changes in my beliefs can affect my behaviour and sometimes even my desires. Although I realise that there are problems of self-opacity, it seems to me that my telling you what it is that I believe does

[27] I suppose that one way to ascertain whether I am right about this being a judgement of pre-theoretic intuition would be to poll a random sample of non-philosophers. In lieu of that, I can only assume that this seems plausible to the reader.

more than just rationalise my behaviour or allow you (or me) to predict my future behaviour— it helps explain what caused it.

I do not wish to rest my case on either pre-theoretic intuition or introspection. There is another consideration in favour of my contention, though it may appear to be weak support since it amounts to the contrapositive of my position.[28] It is the view that, if intentional states are not causally efficacious, then there is little reason to regard them as genuine existents. The wide-spread belief that this is true would explain why epiphenomalism is so disturbing to so many realists.[29] As I noted above, Kim captures this idea by invoking Alexander's dictum: "*To be real is to have causal powers*" (1993, p. 201).

Consider the following passage by Alexander cited by Kim: "[Epiphenomalism] supposes something to exist in nature which has nothing to do, no purpose to serve, a species of *noblesse* which depends on the work of its inferiors, but is kept for show and might as well, and undoubtedly would in time be demolished" (1927, p. 8).

Interestingly, such an argument for the causal efficacy of extant items has an evolutionary twist to it. One would expect that epiphenoma would be selected against. This is to assume, of course, that there is a cost to having a trait that does no work and that there is sufficient variation to allow selection to operate.[30] This is what is being

---

[28] It is not, however, too weak, since it ties in with the reasonable view that cases of Darwinian enquiry assume that a step four can be made. More on this below.

[29] Fodor (1989) attempts to cure this prevalent "epiphobia".

[30] This assumption is problematic when it takes on the aura of religious dogma. To believe that it always obtains is to be committed to a sort of hyper-adaptationism that is blind to the effects of pleiotropy, random genetic drift and the like. See Chapter 4 for discussion.

assumed in a Darwinian enquiry.[31] Success (longevity) is likely due to something (fitness), which can be explained in generic terms, and these generic considerations can be explained in more detail, by explaining the causal interaction between organism and environment. To explain the capacity of the organism to interact successfully with the environment, an analysis of the capacities of its parts is required.

§ 5.222 *Causal Explanation and Temporal Considerations*

If realism about intentional explanation requires that it be causal explanation, then a synchronic account of function is required, if function is to help individuate content. This is so because we will be at step four of a "Darwinian enquiry". Recall the following characterisation of the task of Darwinian enquiry: "first, enumerate current and past species that have endured; second, attribute comparative fitness-levels on the basis of species endurance; third, explain the fitness by identifying generic fitness-making characteristics; and fourth, describe in detail the way organisms achieve these characteristics" (Rosenberg, 1996, p. 18, verb tense changed, summarising Kitcher, 1993b).

On my understanding of this view, a Darwinian enquiry is started with the faith that one will, in principle at least, be able to take step four. If there is no explanation of how the organism exhibits its fitness-enhancing traits, then its fitness is mysterious.

---

[31] I discussed this enquiry in § 1.1 and will rehearse some of that discussion immediately below in § 5.222.

Intentional properties are no different from other properties to which appeal is made in such an enquiry.

My view of intentional psychology is not incompatible with Millikan's.[32] In fact, as I noted at the end of § 5.21, the causal explanations to which she alludes are part of our quest when we engage in step four. If our cognitive capacities are part of an explanation of our success, as surely they are, we want to know how it is that we exhibit them. It is here that a synchronic account is required, on pain of regress. If we cannot ground such efficacy in the case of intentional properties, we have no reason to be realists about intentionality, given both "Alexander's Dictum" and the logic of etiological explanations. Similarly, if we cannot ground causal efficacy in the case of less controversial, but selected for, properties, then we have no reason to think we are right about the etiological story we have offered. Darwin needs Aristotle, Millikan notwithstanding.[33]

---

[32] Millikan's insistence that causal/information-theoretic considerations do not belong in a theory of the individuation of mental states creates problems. See Pietroski (1992) for an example in which the mere correlation that Millikan thinks sufficient results in a highly counter-intuitive content attribution. Creatures are assigned mental states with a content that they have no means of discriminating, rendering mysterious the work to be done at step four. (However, see § 5.113 for a *prima facie* problem about causal/information-theoretic accounts, *viz.*, that of determining the content of thoughts caused not by their objects but by other thoughts.)

[33] Setting aside the context of both Alexander's dictum and a Darwinian enquiry, it is still clear that a synchronic account of mental capacities is required. Cummins (1983) distinguishes property theories and transition theories. The latter theories attempt to provide an account of change in the states of a system, while the former attempt to explain how a system manifests the properties it does. Property theories do not explain changes, they explain how it is that a system instantiates a property. Although Cummins likely would not approve of the use of his work for this purpose, I think it helps explicate the concerns I think need to be addressed at step four.

Chapter 6

Evaluating Structural Theories of Teleological Function

Anything is possible if you don't know what you're talking about.

Overheard in a Vancouver restaurant

I have argued that an etiological construal of teleological function, though appropriate for some explanatory work, will not ground the normativity inherent in the notion. I have maintained that an ahistorical account is required, once the conceptual space is adequately framed, and have indicated that a structural approach could fit the bill. In this chapter, I defend against charges of *prima facie* implausibility the conclusion that the structural account of teleological function is a promising avenue for work in the philosophy of mind. This defence consists in both defending the verdicts rendered by the structural account and demonstrating some implausibilities in the etiological approach. Next, I note some of the limitations of appeals to structural function. Finally, I offer a brief summary of the dissertation as a whole.

§ 6.1 *Why Structural Accounts are not Wonky*

In this section, I discuss a series of troublesome cases, with the aim of establishing that a structural account of teleological functionalism can accommodate them without doing

violence to some common strong intuitions. Not surprisingly, part of this discussion includes arguing that an etiological account *will* do violence to those intuitions.

The problem cases I address in this chapter are often invoked in an attempt to show that structural teleological functionalism casts its net too widely. The general objection, as we saw in Chapter 4, is that a structural analysis requires that we ascribe teleological functions in situations in which we should not. In the first sub-section immediately below (§ 6.11), I discuss three general cases: the solar system, pendula (and some other artefacts) and crystals. In the second (§ 6.12), I discuss two sorts of cases: swamp entities and the 'problem' of ascribing to our noses the function of supporting eyeglasses. I argue that structural teleological functionalism does not attribute functions in the cases discussed in § 6.11, which is as it should be.[1] I argue that it *does*, under certain conditions, attribute functions in the cases in § 6.12, which is also as it should be.

§ 6.11 *Solar Systems, Pendula and Crystals*

As already noted, physics and biology differ in a number of ways; all else being equal, an adequate theory of teleological function ought to accommodate, and perhaps even help to explain, this fact. One of the more salient differences between physics and biology lies in their respective attitudes with regard to the respectability and applicability of normativity. Discussions of what something *should* be doing, in physics, are largely out of place.[2] It is

---

[1] In the course of my exposition of how the structural approach is not too liberal, I also defend my solution from charges that it renders the approach the opposite— i.e., too restrictive.

[2] There may be some branches of physics that do use normative discourse freely. I have in mind those areas of applied physics in which the subject matter is artefacts. For example, it is commonplace to say that the carburetor of a gas-powered engines is *supposed to* mix air and fuel.

surely not appropriate to say that a sample of liquid we took to be an acid is defective when it fails to turn a piece of litmus paper pink, or that a planet is malfunctioning when its orbit is not elliptical. Instead, we say that we mispredicted the relevant outcome due to our own lack of understanding. When normative discourse occurs in physics, it is used to express our own surprise. It *should* have been the case that the sample turned the litmus paper pink, the planet's orbit *should* have been elliptical, but only in the sense that these were what our understanding of the circumstances had us expecting. There need be nothing *wrong* with the sample or the planet—what is wrong is our comprehension of their natures and situations. When we know more, such 'errors' will disappear.

As noted earlier, this is not the case in biology, let alone in psychology. Even without embarking on an extended discussion of how to delimit the realms of the various sciences, it is not controversial to note that normative notions are often applicable in biology, but rarely, if ever, in physics. My concern in this section is to address the objection that a structural theory will ascribe teleological functions, and hence normativity, to things it clearly should not.

I have said that a structural analysis is appropriate when dealing with systems that exhibit a degree of complexity which is relevant to their successful complex interactions with the environment in which they are found. Doubtless, this characterisation could use some refinement. However, despite its vagueness, it offers a preliminary delineation of the set of interest, helping to mark off a class of systems that are amenable to functional analysis.

However, this normativity is indirect— it is derived from our intentions, not original in the artefacts themselves.

However, consider the planet Venus. Does a structural approach force us to ascribe a function to it, relative to the solar system? Venus is certainly a part of the solar system and plays a role in the solar system's manifesting the 'behaviour' it does. For example, without Venus, the orbits of the other planets would be different and the solar system (assuming it could even maintain its cohesion without Venus) would exhibit a different pattern. Is the function of Venus to give the solar system its current configuration? If this is its function, its *teleological* function, then there is something it *should* be doing. On this view, if Venus did not perform this role, then it would be malfunctioning. This result sounds absurd. Is the structural theorist forced to give assent to it?

A sophisticated etiological account such as GS can rule out the claim that Venus has a function in our solar system on the grounds that it is not a component of a biologically real system; one such as RM can do so on the grounds that it is not a member of a reproductively established family. The solar system may be a real system on some construals, but it is not biological.[3] Venus is not a member of a reproductively established family—it has no ancestors.[4]

I shall return to the etiological account in a moment. For now, let us note that these options are not available on the structural approach. Does this mean the structural theory would assign Venus a function? There are, of course, two routes open—either it

---

[3] Defining what makes a system biological is not easy. Godfrey-Smith (1994) opts for the circular suggestion that what is biological is what figures in biological science. However, that might be sufficient for his purposes—there are few, if any, theorists who want to claim that the solar system is a biological phenomenon.

[4] See § 2.211 for a discussion of what is required for something to be a member of a reproductively established family.

does or it does not. If it does, then, in order to maintain its plausibility, the theory must offer an explanation of why most people's intuitions suggest that it should not. If it does not, then an account of what precludes such an ascription is required. I maintain that the approach does not entail that Venus has a function.

The reason that the structural approach will not assign Venus a function is two-fold. The behaviour of Venus relative to the solar system is not sufficiently complex and, furthermore, the behaviour of the solar system relative to the universe is not sufficiently complex. The way in which the solar system maintains its internal integrity and is affected by the impinging forces of the environment is not amenable to a functional analysis. Though the solar system consists of parts (Venus being one of them) it does not decompose into sub-systems. That is, its parts do not interact in such a way so as to form sub-systems that contribute to the workings of the whole. The solar system, despite the large size of its parts, is, in terms of levels of organization, more like fundamental items in physics than like items in biology. Furthermore, the solar system does not interact with its environment in a sufficiently complex way. Its repertoire of behavioural responses to external impingements is limited. The solar system is essentially passive.

Bedau (1992b) has some remarks that will be helpful in beginning to flesh out this notion of complex interaction, although it should be kept in mind that he would not be sympathetic to my view on teleological function.[5] The context of Bedau's remarks is in opposition to systems approaches to teleology. His stalking horse is goal-directed

[5] Recall that Bedau is the author of schema MB, discussed earlier. It is an etiological theory, but it makes essential reference to value in order to handle some of the difficulties encountered by the etiological approach. (One of Bedau's examples, crystals, will be discussed shortly.) MB was rejected for its incompatibility with naturalism.

theories, though I think he would intend that many of his criticisms may be applied equally to structural accounts, as the over liberal attribution of functions is a problem for a systems-theorist of any stripe. His focus is on theorists who adopt the goal-directed approach and who attempt to define "system" in such a way so as to avoid assigning teleology to things such as pendula or marbles in bowls.[6] He characterises one such attempt[7] this way: in order to be a goal-directed system "[t]he causal connection between the deflecting variables and the restoring variables must not be simple and direct" (1992b, p. 42). Deflecting variables are those that move parts of a goal-directed systems from their goal states, and restoring variables are those that move them back. Bedau writes that, though the phrase "simple and direct" may not be completely clear, it is clear enough that the causal connections in simple equilibrium systems, such as the marble and bowl case, are simple and direct. Bedau's objection is that we can design items in which the causal connections would clearly not be simple and direct, but which we should not consider goal-directed. He has in mind such things as "a pendulum suspended from a balloon with a steel bob hanging from a rubber band and swinging above an electro-magnet powered by a battery ... couple this pendulum with another by suspending the first from the bob of the second ... isolate the whole apparatus inside a gravity-free chamber ... power the electromagnet by a generator running off a waterfall far outside the chamber ... and concoct a mechanism that makes the temperature in the chamber (and,

---

[6] The problem arises for theories in which the "goal state" is one of equilibrium. Without further restrictions, a pendulum has the goal state of being motionless at the lowest point in its arc (or, if frictionless, of moving back and forth cyclically) and a marble in a bowl has the goal state of being in the bottom of the bowl.

[7] Bedau cites Van Gulick (1980) as an example.

thus, the length of the rubber band) a function of, say, the current number of vehicles on the Golden State Bridge." (1992b, p. 42).

My point in bringing all this up is to turn Bedau's point against him.[8] The requirement that the causal connections between deflecting and restoring variables not be "simple and direct" should be seen as analogous to the requirement that the interaction between system and environment not be simple and direct in order for functional attributions to be appropriate. If sense can be made of causal connection not being simple and direct, as Bedau grants, then his alleged counterexample does not tell against the structural approach. His complicated pendulum system is not a goal-directed system, but it *is* one in which attributions of function are appropriate on the structural view. However, they are so only insofar as the pendulum is an artefact. If an item is the result of intentional construction, then its parts can be ascribed functions on the basis of their contribution, or would-be contribution, to the intended overall effect. Malfunction is possible either if the overall effect is not achieved or if the parts do not make their intended contribution.[9]

On the structural view, the reason such a system would not have functions, aside from such derivative ones if it is an artefact, is that, despite its internal complexity, the system responds to only one input from its environment—the number of cars on the Golden Gate Bridge. It also has a very restricted range of output—variations in the length of the arc of the pendulum. Such a contraption is not sufficiently complex with regard to its interactions with the environment to warrant an analysis with respect to it.

---

[8] This is somewhat unfair, as Bedau does not treat systems approaches that are not of the goal-directed family. He may have a different objection to the view I am advocating.

[9] See § 3.4 for discussion.

At any rate, the requirement of complexity along two dimensions handles the solar system example in both dimensions. It is not internally complex and its interactions with the environment are not complex.

Is there anything whose interactions with the environment are complex, whose behavioural repertoire is large, yet whose internal structure is not complex? I think not. If we encountered a very adaptable and interactive new creature and discovered that it was structureless internally, we would be more than a little surprised. We would think either that is was being controlled remotely, so we were not really looking at the *whole* system, or that it was an exception worthy of challenging our hierarchical picture of the natural world.

At any rate, before leaving the solar system (so to speak), let us return to the etiological account. There is an attenuated sense in which the solar system has been selected; not all configurations of planets, rocks, and stars will survive the entropic forces in the universe. Some configurations are selected in virtue of the properties of some of their parts. One would expect that this would not be enough to give these parts functions according to the selectionist account. However, a simple account that does not rely upon stipulations that only biological phenomena or members of reproductively established families are candidates for functional ascription must resort to another line of defence. I can think of two. One, heritable variation in fitness is required for selection to occur. This is related to the stipulation that members be of a reproductively established family.

Two, the scenario envisioned is only an instance of what Neander calls single-step selection, and hence does not result in functions.[10]

Functions, the etiologist argues, are not bestowed when an item is selected for only once. Several heats of competition must be run before function is awarded. These several heats are, in the biological case, generations. *Exactly* how many generations of selection are required, the etiologist may complain, is one of those "philosophical" questions, in the pejorative sense of the term. There is no need for a sharp line dividing function from soon-to-be-function; it is sufficient to note that significantly more than one heat is required.

The problem I want to raise with this second line of defence is not that it renders function attribution vague or that it is ill-motivated. The salience of selection pressures in explaining the trait in question will depend upon how significant a role selection has played, and that is a factor, in part, of how long selection has been operating. Rather, the problem is that what is doing the work is the reproductively established family requirement. There is no reason to think that only one heat has been played in the universe, unless heats are measured in terms of generations. Our solar system is constantly bombarded by the entropic forces in the universe. It has withstood many "tests". It has been "selected" many times. So, it appears that being biological is criterial for etiological theories, except in the case of artefacts. However, now is the point at which to push for a more detailed answer to the question of what makes something

[10] Neander is contrasting this with cumulative selection. See § 4.111 for discussion.

biological. I suggest that an important feature is the level of organizational and interactive complexity.

A phenomenon that raises a similar difficulty for the etiological approach is crystals.[11] Clay crystals, though not biological, exhibit features that the solar system does not. Clay crystals reproduce and their traits are heritable. Varieties of crystals that are well-suited to their environment proliferate; those that are not do not. Unless non-biological, non-artefactual items are ruled out *ad hoc*, the etiological approach is faced with attributing functions to those traits of the crystals responsible for their success. Even RM, with its restriction of the phenomena to those that are part of a reproductively established family, will attribute functions to features of crystals.

The structural approach is not faced with such a consequence for two reasons: neither the internal structure nor the "behaviour" of crystals are of the right sort of complexity for a functional analysis.

Crystals, though fascinating, are not complex enough to warrant a decomposition into sub-systems. Explaining their larger capacities to, say, survive and reproduce does not require an analysis of the capacities of their mereological parts. Their parts are groups of atoms that form unit cells which are repeated, making up the crystal itself. A crystal's macro-properties are determined by its micro-properties, to be sure, but there is no significant "nesting" of parts— at the unit cell level, the parts are fairly homogeneous and the interactions between them are of limited variety. Similarly for sheets of crystals. What is essentially aggregation of simple parts does not result in complexity. (Recall the

[11] See Bedau (1992a) for an extended discussion of this example. His solution, as already noted, is to add value as an irreducible feature of the world and one that does not apply to crystals.

discussion of complexity in § 4.12.) Crystals are, on an hierarchical picture of nature, relatively flat. A science of crystals would not be far removed from physics and chemistry.[12]

Similarly, crystals do not have a behavioural repertoire to deploy in their contact with the environment. They maintain their individual integrity and their lattice structure by means of bonds at the atomic level, but they do not interact with the environment in any but a passive sense. Thus, the structural approach will not attribute functions to them.

§ 6.12 *Swamp Creatures and Pangloss' Glasses*

There is another family of objections to the structural approach. As with most of those of the previous section, they are sometimes invoked in an attempt to show that structural teleological functionalism is too broad. However, in contrast to the previous section, these cases are not, despite how they appear to some, problems at all. Structural teleological functionalism will (with restrictions still to be discussed) attribute functions in these scenarios, but that is not a problem. Quite the opposite—*not* attributing functions would be a problem.

---

[12] Cairns-Smith (1982) argues that it is the relative simplicity of crystals that makes them likely candidates for precursors of organic molecules. On his view, anti-Darwinian criticisms that appeal to the improbability of organic molecules arising in a world that did not already contain life can be met by postulating that clay crystals, which are not organic, played an essential role in early evolution. His view is that it is the simplicity of crystals that means that their existence does not need to be explained by invoking a creator or pointing to a potentially vicious regress of prior natural selection. I do not know whether or not his theory is correct, and mention it only for the support it provides to my claim that crystals are not complex systems.

Swamp creatures, so the the story goes, are entities that have come into existence by decidedly unusual and improbable means.[13] They are the stuff of science fiction or outrageous philosophical thought experiments, making their appearance as a result of random molecular activity, perhaps occasioned by an equally random electrical discharge. The details do not really matter, as long as such creatures do not have any more history than that. They are not designed by intentional agents and they are not copied from anything. They are not the descendants of anything; *a fortiori*, they are not the descendants of intentional agents.

The point of invoking swamp creatures is to lay bare intuitions with regard to the consequences of a thing's having virtually no history. Exactly what swamp creatures are taken to show depends very much on one's other views. Davidson invoked his Swampman in the course of re-iterating his contention that, if something lacks the appropriate etiology, it lacks intentional mental states. I think that Swampman causes problems for etiological accounts of content.[14]

In Davidson's story, he (Davidson) is annihilated, and Swampman fortuitously and independently appears. Every token molecule of Swampman's body is of the same physical type as every analogous token molecule of Davidson's, and they are arranged in the same way. Swampman and Davidson are doppelgängers. At this point in the story, there are a number of things one could think would happen next, but the consensus in the philosophical literature is noteworthy—Swampman behaves in a way that makes him indistinguishable from Davidson. None of Davidson's friends notice a difference,

---

[13] Davidson (1987) is the locus classicus. See also Brown (1993).
[14] See, e.g., Neander (1996) and Antony (1996).

Davidson's wife is not shocked to find Swampman in her bed, and Swampman publishes articles on radical translation.

Let us try to fill in some more detail. Davidson is in the marsh. He is annihilated but, as luck would have it, Swampman appears. Next, Swampman goes to Davidson's house. Does he drive? Of course. Davidson's Tercel is right there and he left the keys in it. Does Swampman head to Davidson's by the most direct route? Yes, although he does stop to buy lottery tickets on the way, just as Davidson had planned to do. The Tercel is dangerously low on gas. Does he stop to buy some? Yes, at a self-serve, no less. When he gets to Davidson's, does he chat with Davidson's wife? Recall—Davidson's wife cannot distinguish Swampman from Davidson. Presumably (unless the Davidsons had a truly horrible marriage), his arrival results in some moderate degree of interaction. Swampman pulls it off, even managing to get her name right. He goes into the kitchen and fixes himself a sandwich, managing to avoid the peanut butter to which he is (as poor Davidson was) allergic. Later, night falls, he is tired, and he falls asleep in Davidson's bed, not in Davidson's bathtub or shed.

If Swampman can do all this without thinking, as most commentators maintain he can, what good is thinking? Why, when we do such things (or when poor Davidson did), do we think it appropriate to advert to the contents of our (or Davidson's) thoughts? On the etiological view, Swampman may eventually come to have thoughts. With enough of the right sort of history, he will become a "thinker". Would this be an honorific? Why? He would not behave any differently. There would seem to be no reason to think that his

phenomenological life would be different.[15] If we deny that Swampman has thoughts, are we committed straight-forwardly to denying that having thoughts is causally relevant to behaviour and feelings?

Allow me to try to draw out two general motivations for denying Swampman any contentful thoughts. One motivation is allied with a consideration of externalist concerns in general; the other is a consideration of the role of etiology in particular in typing tokens.

Putnam (1975) presents another hypothetical scenario that some take to show that the content of a mental state is determined, in part, by the social and environmental setting in which the thinker is situated. The famous example goes roughly this way: Oscar, who lives here on Earth, has thoughts about water. That is, he has thoughts about H_2O. His molecular duplicate, Twin Oscar, lives on Twin Earth. Twin Oscar has thoughts about Twin Water. Given the different chemical compositions of water on Earth and Twin Earth, it is the case that, when Twin Oscar has "water" thoughts, he has thoughts about the chemical compound XYZ, not about H_2O. Oscar and Twin Oscar are, *ex hypothesi*, doppelgängers, but the contents of their thoughts, the story goes, differ.

I dredge up this scenario not to solve the problems that surround it, but to make more intuitive the concerns raised in the Swampman case. If one finds it reasonable to think that the content of Twin Oscar's thoughts are determined by things external to him, one will be sympathetic to other views which similarly rely upon historical and situational

[15] Although I think that appeals to the subjective nature of experience are relevant, I am unable to address that issue here—there are problems enough. However, see Seager (1997) and Searle (1992) for further discussion.

considerations for their force. I wish to deny that social factors are paramount, though I am willing to allow that situational considerations are crucial.

Most commentators of the teleological persuasion who object to attributing thoughts to Swampman do so on the grounds that there is no saying what he *should* be doing.[16] On the etiological view, he does not yet have the sort of history that endows some of his states with normativity. If something had materialized that was just like Swampman but with a "heart" that could not pump blood, it would be, this line of argument maintains, inappropriate to claim that the "heart" is dysfunctional. It would be inappropriate to call it a heart at all, since that is to classify it as a token of a functional type. This is precisely what the etiologist maintains we cannot do, in the absence of the right sort of history. The Swampman with the beating "heart" does not have a heart, any more so than does the one with the non-beating "heart".

Consider what is interesting about Swampman, aside from his unlikelihood. What is interesting is his successfulness—that is what needs explaining. If Swampman prospers in this world, it seems likely that his cognitive capacities have something to do with it.[17] That he prospers is a noteworthy fact about him. Presumably, most spontaneous swamp entities do not do nearly as well as Davidson's doppelgänger. A swamp incarnation constituted of "copies" of my left elbow, a coffee cup, and my dog's tail could scarcely be expected to be around for long. There is an astonishing point of

---

[16] See, e.g., Millikan (1996) and Neander (1996).

[17] Millikan (1996) objects that, because Swampman does not have the right history, he is not human. Because he is not human, he cannot have a human psychology and does not fall under the domain of human psychological theorizing. I agree with this, but fail to see how it is relevant. Being human may require a certain etiology, but it does not follow that being *cognitive* does. It is not constitutive of cognitive psychology that it apply only to humans— the requirement that Swampman be *human* is unmotivated.

agreement among most swamp-commentators—Swampman goes on to live the good life, writing articles on radical translation, sleeping with Davidson's wife, and generally befuddling his critics. There are only a few avenues of explanation here. One is that Swampman is not just fortuitous in occurrence, but fortuitous in continued existence. It is sheer luck that he does not drive off the road. It is sheer luck that the words that come out of his mouth enable him to buy gas, greet Davidson's wife, and entertain his friends.

Another option is that, since his neurology and physiology is identical to Davidson's, it is not luck at all that explains his continued existence. Rather, given the fact of his existence, the rest is not surprising at all. His brain is in the exact same configuration that Davidson's was the moment Davdison was annihilated, *except that it has no contentful states.* Swampman is not lucky. There is an explanation for his success, but it is not an intentional explanation. This option makes the contentfulness of certain mental states causally irrelevant to the behaviour and success of their bearer. Swampman behaves as he does, even though he does not have a single thought. As I argued in § 5.22, this renders intentionality unreal.

A third option is that Swampman does have contentful mental states. We explain his behaviour in the same way we would have explained Davidson's because it is caused in the same way as Davidson's. He puts gas in the car because he thinks it needs it. He goes to the fridge because he believes that is where the cheese is. However, probable as it may seem, this option is not available to the etiologist. Swampman does not have intentional states until he has existed for a while and has been subject to selection pressures.

The fact that the etiological approach initially withholds intentional ascriptions from Swampman constitutes a serious difficulty for the applicability of the approach to the individuation of mental states. Both Swampman's success and the advent of his intentionality are rendered mysterious. Content, if it is to be causally relevant, ought to be a property of Swampman that is "visible" from a structural perspective. (See Chapter 5 for further discussion.)

Dretske (1996) raises an interesting case that would seem to support the etiological intuition that a spontaneously generated item would not have any functions. He imagines that a molecule for molecule twin of his Tercel appears spontaneously. The only difference is that the item that, in Dretske's Tercel, would be the gas gauge, is not connected to the fuel tank. Despite the fact that it cannot indicate the amount of fuel in the tank, it is not malfunctioning, argues Dretske. There is nothing it is *supposed* to be doing, and therefore there is no such thing as failing to fulfill function.

I think that this is right, but I do not think that it shows what Dretske intends it to show. Tercels are artefacts, and artefacts get their functions from their designers' intentions. Twin Tercel is not an artefact.

Even so, suppose that Dretske adopts the Twin Tercel. Once he does, the structural approach will attribute functions to the Tercel and its parts insofar as it *and Dretske* form a system. If there were no such thing as drivers, the parts of Tercels would not have the functions we now attribute to them, even though they are artefacts.

Interestingly, on the structural approach, it might appear that certain artefacts cannot be given a functional analysis and that, hence, they do not have functions.

Consider a club. It is not complex. There is no analysing its components into sub-systems and investigating the way in which these sub-systems contribute to the working of the whole. It is not sufficiently complex for a functional analysis to be required. A club is, after all, just a stick, even if it is a stick that has been modified by its user. So, the systems-approach would appear to force us to say that clubs have no function.

However, this appearance is misleading. The system to be considered is not the club in isolation, but rather the club in concert *with its user*. The club is a component in the capacity of the club-and-user to damage, say, perceived enemies. The club, unless part of a larger system, has no function. This is also true of the user. The user has no function if the user is not a component of a larger system. People, *qua* people, are functionless. People, in the context of a social system such as a family or business, may well have functions. Relative to such systems, they can have the functions of, for example, being a caregiver, an accountant, or a bricklayer. If a person were not positioned in the relevant system, these functional claims would be meaningless.

I am not convinced that our accounts of artefactual and natural function must be, in every case, univocal. Some artefactual functions exist purely because of our hopes and aspirations. When we are wrong about what role a thing can play, what things it can do, there is no possibility of an actual functional analysis of its containing system performing the desired or hypothesised activity.[18] Nonetheless, even in these cases, the systems-theoretic approach is instructive. The function to attribute, even if it has never been and

[18] It is possible that something else plays the requisite role, so that the desired activity does come to pass anyhow. However, in that case, it will not be due to the thing we had thought or intended to play a role.

cannot be performed, is the one the designer or user intended or thought could be performed. (For more discussion, see § 3.4.)[19]

This brings us to another favourite intuition pump that involves artefacts: Dr Pangloss' understanding of the relationship between noses and eyeglasses. In Voltaire's classic parody of Leibniz, Pangloss explains to Candide that, given the principle of sufficient reason, everything happens or exists for a reason. Furthermore, since this is the best of all possible worlds, everything that happens or exists *should* happen or exist. Adopting this view results in Candide's humourous attempts to interpret the series of calamaties that befall him as being for the best. It also results in his attributing the function of supporting eyeglasses to noses. This sounds absurd, just as Voltaire intended. Does a structural conception of teleological function force us to the same absurd conclusion?

It would seem that it does, at least if certain circumstances hold.[20] On the structural approach, if being able to wear eyeglasses were a component in our successful interaction with the environment, then those of us whose noses cannot support eyeglasses malfunction in this regard. After all, the capacity of the nose to support eyeglasses is part of the explanation of the successful interactions of the complex system that is constituted by eyeglasses and eyeglass wearer. Surely, though, it is just as absurd for us to attribute

---

[19] Recall our earlier expository device, Sober's selection toy. Although it illuminates the distinction between selection of and selection for, it does not result in functional ascriptions. The marbles that make it to the end of the tube do not have the function of being small— it seems ridiculous to say that they *should* be small. A structural approach would not require us to say this. Whatever functions are attributable to the parts of the toy are a matter of the designer's intentions, and the habits and desires of the users of the toy.

[20] I do not wish to argue *that* the structural approach results in this functional attribution. I simply want to argue that, *if* it does, the structural approach is not thereby shown to be flawed.

this function to noses as it was for Pangloss. However, the reason this sounds absurd is that an etiological conception is being assumed. Recall that our explanandum is not, *contra* Leibniz and Voltaire (and Millikan and Godfrey-Smith), 'why is this here?' Rather, it is to explain the success of a type of thing. If functions figured in explaining only why things exist, then supporting eyeglasses could not be the function of noses. Noses would exist even if they could not support eyeglasses. If, instead, functions can explain how things succeed, then the idea is no longer absurd. It is still, of course, implausible, but that is because this capacity in fact plays no significant role in our success.

§ 6.2 *The Limits of Appeals to Structural Teleological Function*

I argued in § 5.22 that structural, not etiological, teleological functionalism has us looking for the right sort of answer to questions about the individuation of intentional mental states, *viz.*, a synchronic one. Not surprisingly, such a functionalism by itself is insufficient to render content determinate and to ground and explain its causal efficacy. In this section, I explore the limits of appeals to structural teleological functionalism and suggest some ways in which it needs to be supplemented. The difficulties I discuss include (i) lingering problems of indeterminacy, (ii) the need for a principled way to delimit the appropriate environment against which to analyse functions, (iii) questions about the extent to which the view will attribute "idealised" contents, and (iv) the need for further metaphysics to explicate the notion of levels of nature and to ground the causal efficacy of the properties of the special sciences.

The first problem, that of lingering indeterminacy, is also encountered by the etiological approach to teleological function. Both selection and successful behaviour may fail to be sensitive to the degree of specificity or detail that we think obtains in the content of some of our mental states. Although I disagree with Fodor that selection does not care about properties (see § 5.113), I think that the properties that are relevant to selection will be restricted to those that concern discriminating circumstances involving food, sex, and danger. In the case of the structural approach I have been advocating, the content attributed to a state on the basis of its role in the successful behaviour of a complex system may be more general than introspection suggests.[21] It seems that there is often a variety of types of mental states that can play the same role; hence, what distinguishes them cannot be that role. A belief that a black dog is approaching might play the same role in my behaviour as a belief that a black dog with brown eyes is approaching. More work is required in assigning tokens of mental states to types and in accounting for the degree of detail of some contents.

The second problem that faces structural accounts of teleological function has been raised before. It is the need for a principled way to characterise the environment appropriate to the system in question. When encountering a system suffering some degree of unsuccessful interaction with its environment, in order to know whether it is malfunctioning, one needs to know whether its lack of success is due to intrinsic or

[21] Millikan (1984) rails against considering the results of such introspection, calling it "meaning rationalism". Although it is possible that she is right to focus on more conventional, empirically available sorts of evidence and that we are searching for a way out of the intentional circle, I do not think that our inner mental life can be ignored. Much of the recent work on consciousness is relevant here. See Seager (1997) for an overview.

extrinsic factors. One needs to know where on that continuum this interaction falls.[22] In other words, is there something wrong with the system or is it in the wrong place? Etiological accounts appeal to the historical circumstances in which the system was successful to fix the environment. This option can be modified somewhat and adopted by the structural approach. The environment against which to analyse functions is that in which the type of system is successful. Recall that we are trying to explain how a system is able to engage in complex behaviour. This explanandum requires that some successful interaction between environment and system be possible, or else there is nothing to explain.[23]

The problem of delimiting the environment, though serious, is neither quite so pressing nor quite so difficult as some commentators have suggested. If our explanandum is not the presence of a trait, then there is no need to know the historical environment of the system. The normativity offered by the structural approach, which is the only kind available, is a sort of conditional normativity. It helps to tell us what the parts of the system should be doing if the system is to succeed.[24] It should do x, *if it is in environment y*. The requirements for successful interaction, and hence the normativity, may vary with environments.

This brings us to the third consideration. Does the structural approach to teleological function result in overly idealised functional attributions? That is, does it entail that we assign functions to parts that those parts *cannot* perform? After all, there is

---

[22] See Chapters 2 and 3.
[23] What is still required is a method of categorising tokens as of the same type. Both the etiological and the structural approach face this problem. I discuss its solution in, e.g., § 3.1.
[24] Recall that I argue, in Chapter 4, that the normativity the etiological approach accrues is by its association with an ahistorical, presumably structural, account.

a sense in which my heart should pump blood even if I am floating in outer space and it cannot. This does not seem to be a case of malfunctioning on the part of my heart— *no* tokens of that type can function in this type of situation.

This brings us back to the type/token distinction that we found relevant before. The item in my chest is of the type heart for historical, morphological, and functional reasons.[25] My heart does not differ significantly from other successful hearts. My failure to be successful in outer space falls at the extrinsic end of the intrinsic/extrinsic continuum. The structural approach does not attribute to traits normativity with regard to activities that no token can perform in the circumstances. In such cases, there is no successful behaviour to explain.

Neander (1991a) presents an argument against the statistical definition of biological function that may seem to tell against my conclusion here. However, it does not. She argues, convincingly, that it is a mistake to characterise biological function in terms of what a significant proportion (e.g., most) of the tokens of a trait do:

> ... biological norms cannot be reduced to statistical norms, as we can see by noting that dysfunction can become widespread within a population through epidemics or major environmental disasters. A statistical definition of biological norms implies that when a trait standardly fails to perform its function, its function ceases to be its function; so that if enough of us are stricken with disease (roughly, are dysfunctional) we cease to be diseased, which is nonsense." (1991a, p. 182).

I agree that the statistical approach is doomed. However, I do not think that my requirement that some tokens be functional, that there be successful behaviour to explain,

---

[25] Recall our earlier discussion in Chapters 2 and 3. Historical considerations will not suffice, if the morphology is different enough. Similarly, functional equivalence would be enough, if I had an artificial heart or if I were a swamp creature.

is misguided. In the examples Neander offers, there is either widespread organismic change (an epidemic) or widespread situational change (an environmental disaster). In the latter sort of case, it is the environment that should be different. If it were, successful interaction of the sort that now obtains would be possible. That Neander thinks that the normativity adheres to the environment is indicated by her choice of the term "disaster". In the former sort of case, *we* should be different. The structural approach will attribute functions to our non-functioning parts insofar as they remain of the same type. Historical considerations will not suffice for this. It is possible that disease alters token organs to such an extent that they are only nominally, at most, of the same type they were when they functioned. If that were the case, the normativity would amount to maintaining that we should have different organs.[26] I think that ubiquitous failure to function would effectively preclude functional ascriptions.

The important point to note here is that requiring that there be *some* success is not the same as defining function in statistical terms. Neander's epidemic and environmental disaster scenarios do not cause problems for the structural approach.

Let us turn to the fourth difficulty adduced above for the structural approach. I have been relying upon an intuitive notion of complexity drawing in part on the idea of hierarchical levels of natural phenomena. More work in metaphysics and in epistemology is required to provide the explication and support this sketch requires. How do the phenomena that are postulated in biology relate to those in physics? Are they equally "real"? Relatedly, how can the properties in the special sciences be causally efficacious

[26] See § 3.1 for further discussion.

without interfering with the privileged status of physics? In the context of the current project, all I am able to do is invoke a compatibilist view of the relationship of the phenomena of the special sciences to those of physics. My attempt to cast light on a problem in the philosophy of mind raises as many questions as it answers. However, that need not be a sign that I am on the wrong track.

§ 6.3 *Concluding Remarks*

I began this dissertation by discussing the problem of individuating mental content. Recall from Chapter 1 that the causal/information-theoretic approach encounters a difficulty in rendering content determinate. If the content of a mental state is determined by its cause, or by the information it bears, then there is no tokening a mental state with inaccurate or non-veridical content. Its content would be constituted by all that causes it or by all the information it bears. As noted earlier, an adequate account of intentionality cannot type tokens solely in terms of the things that give rise to them, on pain of attributing massively disjunctive contents and precluding error. For example, it sometimes happens that the state I token upon seeing a raccoon on a dark night has the content <dog>, not <raccoon-on-a-dark-night> or < raccoon-on-a-dark-night or dog>.

It is here that an appeal to teleological function is made. The basic idea, in the philosophy of mind, is that the analogy between malfunction and misrepresentation will help solve the disjunction problem by invoking a suitably naturalised notion of

normativity. A state's content need not be, say, what caused it, but rather what *should* have caused it.

In Chapter 2, I looked at various analyses of teleological function taken primarily from its most natural home— biology. I argued that there are two legitimate families of ways of characterising generic teleological function. Selectionist theories, the current favourites, attribute functions on the basis of selection history; a thing's function is that effect or behaviour for which it has been selected. Systems-theoretic, or ahistorical, accounts attribute function on the basis of an analysis of components with regard to the workings of a whole; a thing's function is that effect or behaviour which contributes to the performance of the whole of which that thing is a part. I argued further that both these families of approaches can be divided into two sub-groups. The selectionist family is comprised of a set of etiological views and of the selectionist propensity approach. The systems-theoretic family is comprised of a set of structural accounts and a set of goal-directed views.

I subjected these views to some scrutiny in Chapter 3, where I examined how well each sort met the desiderata one might reasonably expect to be met by accounts of teleological function. These desiderata consisted in being able to accommodate both natural and artefactual functions and to make distinctions between the following pairs: functional and dysfunction items, functional and vestigial items, and functional and accidental items. Not all theories were equally successful.

I argued in Chapter 4 that this difference in meeting purported desiderata is explicable—the different notions are suited to two different, though related, explanatory projects.

In Chapter 5, I argued that teleological functionalism does not fall prey to one of its most vociferous critics— Fodor. I then argued that etiological teleological functions require an ahistorical story to underpin them. I did this by calling into question Millikan's restriction of the explanatory project of psychology and by showing that even her project assumes crucially a synchronic component.

In this chapter, I showed that there are resources to handle objections that structural teleological functionalism attributes functions inappropriately. The most common objection is that the view would attribute functions too liberally. I argue that it does not. Finally, I discussed some limitations of appeals to structural teleological functionalism and outlined areas for future research.

BIBLIOGRAPHY

Adams, Frederick R. (1979) "A Goal–State Theory of Function Attributes", *Canadian Journal of Philosophy*. Vol. 9, pp. 493-518.

Alexander, Samuel (1927) *Space, Time, and Deity*. Vol. 2, London: Macmillan Press.

Allen, Colin, and Mark Bekoff (1995) "Function, Natural Design, and Animal Behavior: Philosophical and Ethological Considerations", *Perspectives in Ethology, Volume 11: Behavioral Design*. N.S. Thompson (ed.), New York: Plenum Press, pp. 1-46.

Amundson, Ron, and George V. Lauder (1994) "Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology", *Biology and Philosophy*. Vol. 9, pp. 443-69.

Antony, Louise (1996) "Equal Rights for Swamp-persons", *Mind and Language*. Vol. 11, pp. 70-5.

Aristotle "De Generatione Animalium", *The Complete Works of Aristotle: The Revised Oxford Translation*. Jonathan Barnes (ed.) (1984), Princeton: Princeton University Press, pp. 1111-1219.

Aristotle "De Partibus Animalium", *The Complete Works of Aristotle: The Revised Oxford Translation*. Jonathan Barnes (ed.) (1984), Princeton: Princeton University Press, pp. 994-1086.

Aristotle "Physics" *The Complete Works of Aristotle: The Revised Oxford Translation*. Jonathan Barnes (ed.) (1984), Princeton: Princeton University Press, pp. 315-446.

Armstrong, David M. (1968) *A Materialist Theory of the Mind*. London: Routledge & Kegan Paul.

Ayala, Francisco J. (1977) "Philosophical Issues", *Evolution*. Theodosius Dobzhansky (ed.), San Francisco: W.H. Freeman and Company, pp. 474-516.

Beckner, Morton (1959) *The Biological Way of Thought*. New York: Columbia University Press.

Bedau, Mark (1991) "Can Biological Teleology Be Naturalized?", *Journal of Philosophy*. Vol. 88, pp. 647-55.

Bedau, Mark (1992a) "Where's the Good in Teleology?", *Philosophy and Phenomenological Research*. Vol. 52, pp. 781-806.

Bedau, Mark (1992b) "Goal-Directed Systems and the Good", *Monist*. Vol. 75, pp. 34-51.

Bedau, Mark (1993) "Naturalism and Teleology", *Naturalism: A Critical Appraisal*. Steven J. Wagner and Richard Warner (eds), Notre Dame, Indiana: University of Notre Dame Press, pp. 23-51.

Bergson, Henri (1913) *L'Evolution Creatice*. Paris: Librairie F. Alcan.

Bigelow, John, and Robert Pargetter (1987) "Functions", *Journal of Philosophy*. Vol. 84, pp. 181-96.

Block, Ned (1994) "Functionalism", in *A Companion to the Philosophy of Mind*. Simon Guttenplan (ed.), Oxford: Blackwell Publishers.

Boorse, Christopher (1976) "Wright on Functions", *Philosophical Review*. Vol. 85, pp. 70-86.

Braithwaite, Richard Bevan (1953) *Scientific Explanation*. Cambridge: Cambridge University Press.

Broad, C.D. (1925) *The Mind and Its Place in Nature*. London: K. Paul, Trench, Trubner, & Co., Ltd.

Brown, Deborah J. (1993) "Swampman of La Mancha", *Canadian Journal of Philosophy*. Vol. 23, pp. 327-348.

Burian, Richard M. (1983) "Adaptation", in *Dimensions of Darwinism: Themes and Counterthemes in 20th Century Evolutionary Thought*. M. Grene (ed.), Cambridge: Cambridge University Press, pp. 287-314.

Buss, David M. (1989) "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures", *Behavioral and Brain Sciences*. Vol. 12, pp. 1-49.

Canfield, John (1964) "Teleological Explanations in Biology", *British Journal for the Philosophy of Science*. Vol. 14.

Cairns–Smith, A.G. (1982) *Genetic Takeover and the Mineral Origins of Life*. Cambridge: Cambridge University Press.

Caro, T.M., and M. Borgerhoff Mulder (1987) "The Problem of Adaptation in the Study of Human Behaviour", *Ethology and Sociobiology*. Vol. 8, pp. 61-72.

Cartwright, Nancy (1983) *How the Laws of Physics Lie*. New York: Oxford University Press.

Churchland , Patricia Smith (1986) *Neurophilosophy*. Cambridge: MIT Press.

Churchland, Paul M., and Patricia Smith Churchland (1983) "Stalking the Wild Epistemic Engine", *Nous*. Vol.17, pp. 5-18.

Cosmides, Leda, and John Tooby (1987) "From Evolution to Behaviour: Evolutionary Psychology as the Missing Link", in *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge: MIT Press, pp. 277-306.

Cummins, Robert (1975) "Functional Analysis", *Journal of Philosophy*. Vol. 72, pp. 741-65.

Cummins, Robert (1983) *The Nature of Psychological Explanation*. Cambridge: MIT Press.

Davidson, Donald (1970) "Mental Events", *Essays on Actions and Events*. (1980) Oxford: Oxford University Press, pp. 207-25.

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

Dawkins, Richard (1986) *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design*. New York: Norton.

Dennett, Daniel (1975) "Why the Law of Effect Will Not Go Away", *Brainstorms: Philosophical Essays on Mind and Psychology*. (1978) Cambridge: MIT Press, pp. 71-89.

Dennett, Daniel (1983) "Intentional Systems in Cognitive Ethology: The 'Panglossian' Paradigm Defended", *Behavioral and Brain Sciences*. Vol. 6, pp. 343-90.

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

Dennett, Daniel (1991) *Consciousness Explained*. Boston: Little, Brown and Co.

Dewey, John (1938) *Logic: The Theory of Inquiry*. New York: Henry Holt and Company.

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

Dretske, Fred (1986) "Misrepresentation", *Belief: Form, Content, and Function*. Radu J. Bogdan (ed.), Oxford: Oxford University Press.

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

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

Driesch, Hans (1908) *The Science and Philosophy of the Organism: The Gifford Lectures*. London: A. & C. Black Ltd.

Faber, Roger J. (1986) *Clockwork Garden: On the Mechanistic Reduction of Living Things*. Amherst: The University of Massachusetts Press.

Falk, Arthur (1981) "Purpose, Feedback, and Evolution", *Philosophy of Science*. Vol. 48, pp. 198-217.

Fodor, Jerry (1974) "Special Sciences, or, the Disunity of Science as a Working Hypothesis", *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*. (1981) Cambridge: MIT Press, pp. 127-45.

Fodor, Jerry (1980) "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*. (1981) Cambridge: MIT Press, pp. 225-53.

Fodor, Jerry (1981) *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge: MIT Press.

Fodor, Jerry (1986) "Why Paramecia Don't Have Mental Representations", *Midwest Studies in Philosophy*. Vol. 10, pp. 3-24.

Fodor, Jerry (1989) "Making Mind Matter More", *Philosophical Topics*. Vol. 17, pp. 59-79.

Fodor, Jerry (1990a) *A Theory of Content and Other Essays*. Cambridge: MIT Press.

Fodor, Jerry (1990b) "A Theory of Content, 1: The Problem", *A Theory of Content and Other Essays*. (1990a) Cambridge: MIT Press, pp. 51-87.

Fodor, Jerry (1990c) "Psychosemantics or: Where Do Truth Conditions Come From?", *Mind and Cognition: A Reader*. William G. Lycan (ed.), Oxford: Basil Blackwell, pp. 312-337.

Fodor, Jerry (1990d) "Information and Representation", *Information, Language, and Cognition*. Philip P. Hanson (ed.), Vancouver: University of British Columbia Press, pp. 175-90.

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

Godfrey-Smith, Peter (1989) "Misinformation", *Canadian Journal of Philosophy*. Vol. 19, pp. 533-50.

Godfrey-Smith, Peter (1993) "Functions: Consensus Without Unity", *Pacific Philosophical Quarterly*. Vol. 74, pp. 196-208.

Godfrey-Smith, Peter (1994) "A Modern History Theory of Functions", *Noûs*. Vol. 28, pp. 344-62.

Godfrey-Smith, Peter (1994b) "A Continuum of Semantic Optimism", *Mental Representation: A Reader*. Stephen Stich and Ted A. Warfield (eds), Oxford: Blackwell, pp. 259-77.

Godfrey-Smith, Peter (1995) *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Goodman, Nelson (1954) *Fact, Fiction, and Forecast*. London: The Athlone Press.

Gould, Stephen Jay (1980) *The Panda's Thumb*. New York: W.W. Norton.

Gould, Stephen Jay (1983) *Hen's Teeth and Horse's Toes*. Harmondsworth: Penguin.

Gould, Stephen Jay, and Richard C. Lewontin (1979) "The Spandrels of San Marco and the Panglossian Paradigm: a Critique of the Adaptationist Programme", *Proceedings of the Royal Society*. London, B205, pp. 581-98.

Gould, Stephen Jay, and Elisabeth S. Vrba (1982) "Exaptation— a Missing Term in the Science of Form", *Paleobiology*. Vol. 8, pp. 4-15.

Griffiths, Paul (1992) "Adaptive Explanation and the Concept of a Vestige", *Trees of Life*. Paul Griffiths (ed.), Amsterdam: Kluwer Academic Publishers, pp. 111-31.

Griffiths, Paul (1993) "Functional Analysis and Proper Functions", *British Journal for the Philosophy of Science*. Vol. 44, pp. 409-22.

Hall, Richard J. (1990) "Does Representational Content Arise from Biological Function?", *Philosophy of Science Association*. Vol. 1, pp 193-9.

Hanson, N.R. (1958) *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hempel, Carl G. (1965) "The Logic of Functional Analysis", *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, pp. 297-330.

Hull, David (1974) *The Philosophy of Biological Science*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kauffman, Stuart A. (1971) "Articulation of Parts Explanation in Biology and the Rational Search for Them", *Boston Studies in the Philosophy of Science*. Vol. 8, pp. 257-72.

Kazmi, Ali Akhtar (1990) "Comment", *Information, Language, and Cognition*. Philip P. Hanson (ed.), Vancouver: University of British Columbia Press, pp. 191-7.

Kim, Jaegwon (1991) "Dretske on How Reasons Explain Behavior", *Dretske and His Critics*. Brian McLaughlin (ed.), Oxford: Blackwell Books, pp. 52-72.

Kim, Jaegwon (1993) "The Nonreductivist's Troubles with Mental Causation", *Mental Causation*. John Heil and Alfred Mele (eds), Oxford: Oxford University Press, pp. 189-210.

Kitcher, Philip S. (1985) *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge: MIT Press.

Kitcher, Philip S. (1993a) "Function and Design", *Midwest Studies in Philosophy*. Vol. 18, pp. 379-397.

Kitcher, Philip S. (1993b) *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Kripke, Saul (1972) "Naming and Necessity", *Semantics of Natural Language*. Donald Davidson and Gilbert Harman (eds), Dordecht, Holland: Reidel, pp. 252-355.

Lettvin, J.Y., H. Maturana, W.S. McCulloch, and W.H. Pitts (1959) "What the Frog's Eye Tells the Frog's Brain", *Proceedings of the Institute of Radio Engineers*. Vol. 11, pp. 230-55.

Levy, Edwin (1988) "Networks and Teleology", *Philosophy and Biology: Canadian Journal of Philosophy*. Supplementary Vol. 14, pp. 159-86.

Levins, Richard, and Richard C. Lewontin (1985) *The Dialectical Biologist*. Cambridge: Harvard University Press.

Lewontin, Richard C. (1978) "Adaptation", *Scientific American*. Vol. 239, pp. 213-30.

Lipton, Peter, and N.S. Thompson (1988) "Comparative Psychology and the Recursive Structure of Filter Explanations", *International Journal of Comparative Psychology*. Vol. 1, pp. 215-29.

Matthen, Mohan (1988) "Biological Functions and Perceptual Content", *Journal of Philosophy*. Vol. 85, pp. 5-27.

Matthen, Mohan (1997) "Teleology and the Product Analogy", *Australasian Journal of Philosophy*. Vol. 75, pp. 21-35.

Mayr, Ernst (1976) *Evolution and the Diversity of Life*. Cambridge: The Belknap Press of Harvard University Press.

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

Millikan, Ruth Garrett (1989a) "In Defense of Proper Functions", *Philosophy of Science*. Vol. 56, pp. 288-302.

Millikan, Ruth Garrett (1989b) "An Ambiguity in the Notion 'Function'", *Biology and Philosophy*. Vol. 4, pp. 172-6.

Millikan, Ruth Garrett (1993a) *White Queen Psychology and Other Essays for Alice*. Cambridge: MIT Press.

Millikan, Ruth Garrett (1993b) "Explanation in Biopsychology", *White Queen Psychology and Other Essays for Alice*. (1993a) Cambridge: MIT Press, pp. 171-92.

Millikan, Ruth Garrett (1993c) "What is Behavior? A Philosophical Essay on Ethology and Individualism in Psychology, Part 1", *White Queen Psychology and Other Essays for Alice*. (1993a) Cambridge: MIT Press, pp. 135-50.

Millikan, Ruth Garrett (1996) "On Swampkinds", *Mind and Language*. Vol. 11, pp. 103-17.

Mitchell, Sandra (1989) "The Causal Background of Functional Explanation", *International Studies in the Philosophy of Science*. Vol. 3, pp. 213-29.

Mitchell, Sandra (1993) "Dispositions and Etiologies? A Comment on Bigelow and Pargetter", *Journal of Philosophy*. Vol. 93, pp. 249-59.

Mitchell, Sandra (1995) "Function, Fitness, and Disposition", *Biology and Philosophy*. Vol. 10, pp. 39-54.

Nagel, Ernest (1961) *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace, and World, Inc.

Nagel, Ernest (1979) *Teleology Revisited and Other Essays in the Philosophy of Science*. New York: Columbia University Press.

Neander, Karen (1988) "What Does Natural Selection Explain? Correction to Sober", *Philosophy of Science*. Vol. 55, pp. 422-26.

Neander, Karen (1991a) "Functions as Selected Effects: The Conceptual Analyst's Defense", *Philosophy of Science*. Vol. 58, pp. 168-84.

Neander, Karen (1991b) "The Teleological Notion of 'Function'", *Australasian Journal of Philosophy*. Vol. 69, pp. 454-68.

Neander, Karen (1995a) "Pruning the Tree of Life", *British Journal for the Philosophy of Science*. Vol. 46, pp. 59-80.

Neander, Karen (1995b) "Explaining Complex Adaptations: A Reply to Sober's 'Reply to Neander'", *British Journal for the Philosophy of Science*. Vol. 46, pp. 583-87.

Neander, Karen (1996) "Swampman Meets Swampcow", *Mind and Language*. Vol. 1, pp. 118-29.

Paley, William (1802) *Natural Theology: Or, Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature*. London: R. Fauldner.

Papineau, David (1987) *Reality and Representation*. Oxford: Basil Blackwell Ltd.

Pettit, Philip (1986) "Broad-Minded Explanation and Psychology", in *Subject, Thought and Content*. Philip Pettit and John McDowell (eds), Oxford: Oxford University Press, pp. 17-58.

Pietroski, Paul M. (1992) "Intentionality and Teleological Error", *Pacific Philosophical Quarterly*. Vol. 73, pp. 267-82.

Pietroski, Paul M., and George Rey (1995) "When Other Things Aren't Equal: Saving *Ceteris Paribus* Laws from Vacuity", *British Journal for the Philosophy of Science*. Vol. 46, pp 81-110.

Pinker, Steven, and Paul Bloom (1990) "Natural Language and Natural Selection", *Behavioral and Brain Sciences*. Vol. 13, pp. 707-84.

Prior, Elizabeth W. (1985) "What Is Wrong with Etiological Accounts of Biological Function?", *Pacific Philosophical Quarterly*. Vol. 66, pp. 310-28.

Putnam, Hilary (1975) "The Meaning of 'Meaning'", in *Language, Mind, and Knowledge*. Vol. 7, Minnesota Studies in the Philosophy of Science. Keith Gunderson (ed.), pp. 131-93.

Rescher, Nicholas (ed.) (1986) *Current Issues in Teleology*. Lanham, Maryland: University Press of America.

Rorty, Richard (1979) *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Rosenberg, Alexander (1985) *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Rosenberg, Alexander (1996) "A Field Guide to Recent Species of Naturalism", *British Journal for the Philosophy of Science*. Vol. 47, pp. 1-29.

Rosenblueth, Arturo, Norbert Weiner, and Julian Bigelow (1943) "Behaviour, Purpose and Teleology", *Philosophy of Science*. Vol. 10, pp. 18-24.

Salmon, Wesley (1989) "Four Decades of Scientific Explanation", *Scientific Explanation*. P. Kitcher and W. Salmon (eds), Minneapolis: Minnesota University Press, pp. 3-219.

Schaffner, Kenneth (1993) *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Scheffler, Israel (1959) "Thoughts on Teleology", *British Journal for the Philosophy of Science*. Vol. 9, pp. 265-84.

Seager, William (1997) "Critical Notice of Fred Dretske 'Naturalizing the Mind'" *Canadian Journal of Philosophy*. Vol 27, pp. 83-109.

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

Shapiro, Larry A. (1992) "Darwin and Disjunction: Optimal Foraging Theory and Univocal Assignments of Content", *Proceedings of the Philosophy of Science Association*. Vol. 1, pp. 469-80.

Simon, Herbert (1969) *The Sciences of the Artificial*. Cambridge: MIT Press.

Sober, Elliott (1982) "Why Logically Equivalent Predicates May Pick Out Different Properties", *American Philosophical Quarterly*. Vol. 19, pp. 182-90.

Sober, Elliott (1984) *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge: MIT Press.

Sober, Elliott (1995) "Natural Selection and Distributive Explanation: A Reply to Neander", *British Journal for the Philosophy of Science*. Vol. 46, pp. 384-97.

Sommerhoff, G. (1950) *Analytical Biology*. Oxford: Oxford University Press.

Sterelny, Kim (1990) *The Representational Theory of Mind*. Oxford: Basil Blackwell.

Sterelny, Kim (1992) "Evolutionary Explanations of Human Behaviour", *Australasian Journal of Philosophy*. Vol. 70, pp. 156-73.

Stich, Stephen (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge: MIT Press.

Stich, Steven (1994) "What is a Theory of Mental Representation?", *The Mind-Body Problem: A Guide to the Current Debate*. Richard Warner and Tadeusz Szubka (eds), Oxford: Basil Blackwell, pp. 171-191.

Tooby, John, and Leda Cosmides (1989) "Evolutionary Psychology and the Generation of Culture, Part I", *Ethology and Sociobiology*. Vol 10, pp. 29-49.

Tye, Michael (1992) "Naturalism and the Mental", *Mind*. Vol. 101, pp. 421-41.

van Fraassen, Bas (1980) *The Scientific Image*. Oxford: Oxford University Press.

Van Gulick, Robert (1980) "Functionalism, Information, and Content", *Mind and Cognition*. (1990) William C. Lycan (ed.) Cambridge: Basil Blackwel, pp. 107-129.

Van Gulick, Robert (1992) "Nonreductive Materialism and the Nature of Intertheoretic Constraint", in *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*. Ansgar Beckermann, Hans Flohr, and Jaegwon Kim (eds.) Berlin: Walter de Gruyter & Co., pp. 157-79.

Voltaire (1759) *Candide*. John Butt (trans., 1942), Harmondsworth: Penguin.

Walsh, Denis M., and André Ariew (1996) "A Taxonomy of Functions", *Canadian Journal of Philosophy*. Vol. 26, pp. 493-514.

Whyte, J.T. (1990) "Success Semantics", *Analysis*. Vol. 50, pp. 149-57.

Williams, George C. (1966) *Adaptation and Natural Selection*. Princeton: Princeton University Press.

Wimsatt, William C. (1972) "Teleology and the Logical Structure of Function Statements", *Studies in the History and Philosophy of Science*. Vol. 3, pp. 1-80.

Wimsatt, William C. (1974) "Complexity and Organization", *Boston Studies in the Philosophy of Science*. Vol. 20, pp. 67-86.

Wimsatt, William C. (1976) "Reductionism and Levels of Organization", in *Consciousness and the Brain: A Scientific and Philosophical Inquiry*. Gordon G.Globus, Grover Maxwell, and Irwin Savodnik (eds.) New York: Plenum Press, pp. 205-67.

Wright, Larry (1973) "Functions", *Philosophical Review*. Vol. 82, pp. 139-68.

Wright, Larry (1976) *Teleological Explanations: An Etiological Analysis of Goals and Functions*. Berkeley: University of California Press.

Appendix A

Key to Abbreviations for Theories of Function and their Attendant Schemata, Sorted Alphabetically

BP (Bigelow and Pargetter)=

> The function of *A* is *F* only if the performance of *F* by *A* confers a survival-enhancing propensity on the system *S* of which *A* is a part.
>
> (based on Bigelow and Pargetter, 1987)

CB (Christopher Boorse)=

> *A* is performing the function *F* in the *G*-ing of S at [time] t, *means*
>
> At [time] t, *A* is *F*-ing and the *F*-ing of *A* is making a causal contribution to the goal *G* of the goal-directed system *S*.
>
> (1976, p. 80, some variables changed for consistency)

FA (Frederick Adams)=

> A structure *a* has a function *F just in case*:
>
> (1) *a* does *F* in system *S*;
>
> (2) *F* causally contributes towards *S*'s outputting *G* (*through the causal feedback mechanism*);
>
> (3) *G* is (or itself contributes toward) a goal-state of *S*.
>
> (1979, p. 508, variables changed for consistency, his emphasis)

GS (Godfrey-Smith)=

The function of *a* is to *F* iff:

(i) *a* is a member of family *T*,
(ii) members of family *T* are components of biologically real systems of type *S*,
(iii) among the properties copied between members of *T* is property or property cluster *C*, which can do *F*.
(iv) one of the reasons members of *T* such as *a* exist now is the fact that past members of *T* were successful under selection in the recent past, through positively contributing to the fitness of systems of type *S*, and
(v) members of *T* were selected because they did *F*, through having *C*.

(1994, p. 359, some variables changed for consistency)

KS (Kenneth Schaffner)=

To claim that *a* has a function in system *s* is to claim

(1) System *S* has a goal *G* property,
(2) *a*, via *B(a)*, results in or significantly promotes *G*.

(1993, p. 404, some variables changed for consistency)

LW (Larry Wright)=

The function of *A* is *F* iff:

(i) *F* is a consequence (result) of *A*'s being there,
(ii) *A* is there because it does (results in) *F*,

(1976, p. 161, variables changed for consistency)

MB (Mark Bedau)=

The function of *A*'s *B*ing is *F* iff
*A B*s because [*A*'s *B*ing contributes to *F*ing and *F*ing is good]

(based on 1992a, p. 790, the square brackets are his and are intended to show the scope of the "because")

PG (Paul Griffiths)=

> Where *A* is a trait of systems of type *S*, a proper function of *A* in *S*s is *F* iff a proximal selective explanation of the current non-zero proportion of *S*s with *A* must cite *F* as a component in the fitness conferred by *A*.
>
> (1993, p. 418, syntax altered, some variables changed for consistency)

RC (Robert Cummins)=

> the function of *a* in system *s* is *F* relative to an analytical account *X* of s's capacity to *G* just in case *a* is capable of *F*-ing in *s* and *X* appropriately and adequately accounts for *s*'s capacity to *G* by, in part, appealing to the capacity of *a* to *F* in *s*.
>
> (based on 1975, p. 762, syntax altered, some variables changed for consistency)

RM (Ruth Millikan)=

Where *a* is a member of a reproductively established family *T* and *T* has the reproductively established or Normal character *C*, *a* has the function *F* as a direct proper function iff:

(1) Certain ancestors of *a* performed *F*.

(2) In part because there existed a direct causal connection between having the character *C* and performance of the function *F* in the case of these ancestors of *a*, *C* correlated positively with *F* over a certain set of items *I* which included these ancestors and other things not having *C*.

(3) One among the legitimate explanations that can be given of the fact that *a* exists makes reference to the fact that *C* correlated positively with *F* over *S*, either directly causing reproduction of *a* or explaining why *T* was proliferated and hence why *a* exists.

(1984, p. 28, some variables changed for consistency)

WW (William Wimsatt)=

> According to theory *Y*, a function of item *I*, in producing behaviour *B*, in system *S* in environment *E* relative to purpose *P* is to *F*
>
> (1972, p. 32, some variables changed for consistency)

Appendix B

Key to Abbreviations for Theories of Function and their Attendant Schemata, Sorted by Type of Account

A. Selectionist Accounts
A(i)—Etiological
A(ii)—Propensity

B. Systems-Theoretic Accounts
B(i)—Structural
B(ii)—Goal-Directed

A. Selectionist Accounts

A(i)—Etiological Accounts

GS (Godfrey-Smith)=

The function of *a* is to *F* iff:

(i) *a* is a member of family *T*,
(ii) members of family *T* are components of biologically real systems of type *S*,
(iii) among the properties copied between members of *T* is property or property cluster *C*, which can do *F*.
(iv) one of the reasons members of *T* such as *a* exist now is the fact that past members of *T* were successful under selection in the recent past, through positively contributing to the fitness of systems of type *S*, and
(v) members of *T* were selected because they did *F*, through having *C*.

(1994, p. 359, some variables changed for consistency)

LW (Larry Wright)=

The function of *A* is *F* iff:

(i) *F* is a consequence (result) of *A*'s being there,
(ii) *A* is there because it does (results in) *F*,

(1976, p. 161, variables changed for consistency)

PG (Paul Griffiths)=

> Where *A* is a trait of systems of type *S*, a proper function of *A* in *S*s is *F* iff a proximal selective explanation of the current non-zero proportion of *S*s with *A* must cite *F* as a component in the fitness conferred by *A*.
>
> (1993, p. 418, syntax altered, some variables changed for consistency)

RM (Ruth Millikan)=

Where *a* is a member of a reproductively established family *T* and *T* has the reproductively established or Normal character *C*, *a* has the function *F* as a direct proper function iff:

(1) Certain ancestors of *a* performed *F*.

(2) In part because there existed a direct causal connection between having the character *C* and performance of the function *F* in the case of these ancestors of *a*, *C* correlated positively with *F* over a certain set of items *I* which included these ancestors and other things not having *C*.

(3) One among the legitimate explanations that can be given of the fact that *a* exists makes reference to the fact that *C* correlated positively with *F* over *S*, either directly causing reproduction of *a* or explaining why *T* was proliferated and hence why *a* exists.

(1984, p. 28, some variables changed for consistency)

MB (Mark Bedau)=

> The function of *A*'s *B*ing is *F* iff
> *A B*s because [*A*'s *B*ing contributes to *F*ing and *F*ing is good]
>
> (based on 1992a, p. 790, the square brackets are his and are intended to show the scope of the "because")

A(ii)—Selectional Propensity Theory

BP (Bigelow and Pargetter)=

> The function of *A* is *F* only if the performance of *F* by *A* confers a survival-enhancing propensity on the system *S* of which *A* is a part.
>
> (based on Bigelow and Pargetter, 1987)

B. Systems-Theoretic Accounts

B(i)—Structural Accounts

BP (Bigelow and Pargetter)=

> The function of *A* is *F* only if (if and only if?) the performance of *F* by *A* confers a survival-enhancing propensity on the system *S* of which *A* is a part.
>
> (based on Bigelow and Pargetter, 1987)

RC (Robert Cummins)=

> the function of *a* in system *s* is *F* relative to an analytical account *X* of s's capacity to *G* just in case *a* is capable of *F*-ing in *s* and *X* appropriately and adequately accounts for *s*'s capacity to *G* by, in part, appealing to the capacity of *a* to *F* in *s*.
>
> (based on 1975, p. 762, syntax altered, some variables changed for consistency)

WW (William Wimsatt)=

> According to theory *Y*, a function of item *I*, in producing behaviour *B*, in system *S* in environment *E* relative to purpose *P* is to *F*
>
> (1972, p. 32, some variables changed for consistency)

B(ii)—Goal-Directed Accounts

CB (Christopher Boorse)=

A is performing the function *F* in the *G*-ing of S at [time] t, *means*

At [time] t, *A* is *F*-ing and the *F*-ing of *A* is making a causal contribution to the goal *G* of the goal-directed system *S*.

(1976, p. 80, some variables changed for consistency)

FA (Frederick Adams)=

A structure *a* has a function *F just in case*:

(1) *a* does *F* in system *S*;

(2) *F* causally contributes towards *S*'s outputting *G* (*through the causal feedback mechanism*);

(3) *G* is (or itself contributes toward) a goal-state of *S*.

(1979, p. 508, variables changed for consistency, his emphasis)

KS (Kenneth Schaffner)=

To claim that *a* has a function in system *s* is to claim

(1) System *S* has a goal *G* property,
(2) *a*, via *B(a)*, results in or significantly promotes *G*.

(1993, p. 404, some variables changed for consistency)