Dissertations                                                     Theses and Dissertations

2014

# Purpose and Design in Organisms and Artifacts: The Search for a Unified Philosophical Theory of Function

Mark L. Bourgeois
*Loyola University Chicago*

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Philosophy of Science Commons

## Recommended Citation

LOYOLA UNIVERSITY CHICAGO


PURPOSE AND DESIGN IN ORGANISMS AND ARTIFACTS:

THE SEARCH FOR A UNIFIED PHILOSOPHICAL THEORY OF FUNCTION



A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY



PROGRAM IN PHILOSOPHY



BY

MARK L. BOURGEOIS

CHICAGO, IL

MAY 2014

TABLE OF CONTENTS

# LIST OF FIGURES

ABSTRACT

This work examines the concept of function in both biological organisms and designed artifacts. Function is routinely attributed within both kinds of systems, yet it is unclear whether it can mean the same thing within each, and indeed whether function attribution within natural systems is legitimate at all. Function is typically understood to have teleological content; yet in true teleological causation, the end is somehow the cause of its own means. In artifacts the consciously envisioned design of an artifact is taken to be the cause of its parts and their roles. Yet in naturalistic science there is no prior design of organisms, and efficient cause works only in the opposite direction.

Three major approaches to function theory in contemporary analytic philosophy are examined in turn: etiological, systemic effects, and contributions. In addition a novel properties approach is proposed. It is seen that while the etiological approach promises to preserve an account of strongly teleological natural function, it is in fact unable to do so. The systemic effects approach, meanwhile, is found to not attribute function with adequate precision, nor to apply at all to contexts involving intention. The contributions account will be seen to require substantial modification to enable it to attribute function with sufficient discretion, including the necessity of distinguishing function performance from function possession, resulting in the novel Assigned Contribution theory. Finally, an entirely new approach to function informed by emergence theory is advanced. It is

posited that functional parts are those whose properties not only serve the whole, but whose properties have been constrained by the whole of which they are a part.

The four approaches to function will be seen to themselves turn on four different interpretations of teleology. Ultimately, the choice of function theory will be seen to rest on whether non-intentional teleology is a coherent metaphysical possibility. If it is, the properties approach offers important advantages. If it is not, the Assigned Contributions approach is superior. Yet even in this case, the properties approach shows enough promise as a full theory of function to merit further development.

CHAPTER ONE

INTRODUCTION TO THE PROBLEM OF FUNCTION

**Overview**

The subject of this work is the concept of function in both artifacts and organisms. The philosophical problem of function has a number of dimensions. Most obviously, function is typically regarded as somehow related to purpose – and in the context of natural organisms, positing teleology of any kind is seen as problematic. It is one thing to connect function with purpose in artifacts, where the intent of the designer can be appealed to (though even this is more problematic that it is usually taken to be). But without a divine designer to appeal to, attributing purpose to the parts of an organism is more challenging. A related task is to provide a coherent account of malfunction, and the nature and source of the apparent normativity implicit in this attribution. Underlying all these issues is the difficulty of simply articulating the principles which distinguish function from non-function, without which function attribution becomes meaningless.

There are three major strands of theory in the literature of the past half century of analytic philosophy: the etiological, the systemic effects and the contributions. I will analyze each in turn, relying on a paradigmatic thinker or two for each approach. I will also consider a novel approach founded on properties, and conclude by examining the picture of teleology implicit in each approach.

Etiological theory is the most widely accepted approach today, enjoying something approaching consensus (Buller 1999).  Nonetheless, this ambitious theory will be seen to be deeply flawed.  It seeks to ground a strongly teleological and normative attribution of function based on the origin, or etiology, of the functional part.  The basic principle is that an item has a function if and only if the functional effect explains the presence of the item – the paradigmatic examples being selected traits in organisms and design in artifacts.  Its problems derive from the overly tight connection stipulated between origin and function, which is seen to be unrealistic, as well as the strong teleology and normativity that is said to result, which are seen to be metaphysically unsupportable.  The primary thinkers for this approach are Larry Wright and Ruth Garret Millikan (Wright 1973, Millikan 1989b).

Systemic effects theory is essentially a reaction to the metaphysical problems of teleology in etiological theory.  It seeks a minimalist account of function, entirely without teleology or normativity.  Instead, it looks only to the effects of an item on the system in which it is found.  In order for it to exercise some discernment in attributing function, certain conditions for the system and the relevant effects are imposed – namely that the system must be hierarchically organized, and that the effect must contribute to a characteristic "systemic capacity".  However, ultimately these conditions are seen to be insufficient to prevent function from becoming attributable to nearly any causal effect in a minimally complex system.  The main thinkers for this approach are Robert Cummins and Paul Sheldon Davies (Cummins 1975, Davies 2001).

Contributions theory seeks to retain a central role for teleology, but avoid the complications of deriving this entirely from part origin. Instead, parts have function through making causal contributions to the goals of teleological systems. The primary thinker for this approach is Christopher Boorse (Boorse 1976). This approach is seen to require large amounts of expansion and correction, particularly in terms of what constitutes a goal-directed system and when a thing *has* a function rather than merely performing it – but eventually a plausible novel account, called Assigned Contribution theory, is derived.

Finally, following close study of these three conventional accounts, including the extensive rehabilitation of a contributions account, an entirely unique approach to function is developed. This "properties" approach is informed by emergence theory and appeals only to the formal properties of functional parts. It is posited that functional parts are those whose properties not only serve the whole, but whose properties have been constrained by the whole of which they are a part. This approach is a radical break with the existing precedents of function theory, and may point a new way forward not only for function theory but for its relevance to reductionist ontologies, including in philosophy of mind.

It will be seen that the choice of a theory of function turns on the unavoidable question of whether non-conscious teleology is metaphysically coherent. If it is taken to be, the properties approach seems to supply the best account of function. If it is not, the

contributions approach as elaborated in the form of Assigned Contribution theory seems to supply the best account of function.

## The Concept of Function

Function is one of the most apparent, familiar and intuitively compelling aspects of biological life. Everyone knows, for example, that it is the function of the mammalian heart to pump blood throughout the body. Every child learns this fact; but biologists and physiologists also endorse it and embed it in their more sophisticated explanations. But what precisely is the content of the concept of function, and what does its application to natural systems imply?

We share an intuitive sense that function is what most strikingly distinguishes a heart from a raincloud, or a watch from a hunk of ore. Intuitively, function calls to mind an ordered system composed of discrete components which themselves have function and order. Yet not all ordered systems are ordinarily thought of as having function; among them, we could name crystals, the atmosphere, and the solar system. Indeed, all the nonliving aspects of nature which human beings have not transformed or somehow harnessed seem to lack function. This is true even for ordered physical systems which exhibit regular behaviors, and even when they are of use to humans. Some astronomers, for example, have postulated that the planet Jupiter has often shielded Earth from meteor impacts by sweeping them up in its massive gravitational field. None have suggested that this is *the function* of the planet Jupiter, save perhaps metaphorically. What exactly is the distinction between function and non-function being made here?

There are many contexts in which the concept of function finds application. We speak of the function of the parts of an animal and the parts of a sentence; of language itself; of social institutions; of individuals' professional roles; of the parts of a building; of machines and devices and their parts; of aspects of works of art; of policies and laws; of human and animal behavior, and more. What does this vast range of contexts have in common for each to merit functional attribution?

What is most notable about this range of application (which is representative if not exhaustive) is its bifurcation into objects which are natural and those which are human-made. The range of human-made objects is diverse, spanning artifacts to social institutions. Yet those which are natural all relate directly to biological organisms – their parts or their behaviors. Why is it that the concept of function is applied to natural systems exclusively in the form of organisms when every other application is to human-made systems? Most of the human cases involve conscious intentions in rather obvious ways. The natural cases do not; no one believes, anymore, that the human heart was intentionally designed the way that a water pump is. But given this, what then do they have in common which allows them both to be characterized as having a function? Biologists frequently say things such as, "the function of the heart is to pump blood"; "the function of the kidneys to purify the blood of toxins"; or "the function of the lungs is to exchange carbon dioxide for oxygen in the blood". Are such functional characterizations really appropriate? Or are they just convenient but metaphysically unlicensed metaphors, as if the astronomer were to say that the function of Jupiter is to

protect the Earth?  Is the attribution of function to *any* natural system legitimate given

that no intentions are involved?  Again, what is the precise nature of the concept of

function being invoked here?  And what implications does it convey?

**The Value of Function Attribution**

Let us first consider what is prima facie at stake in function attribution

generically.  At the most general level, the value of a functional description is that it

supplements a strictly 'mechanical' *depiction* of physical causality – the brute push of

one thing on another – with a more meaningful *interpretation* of the same physical

event.[1]  It is the difference between "two metal objects of a certain shape sliding and

pushing against one another" and "a camshaft lobe pushing down a rocker arm to open a

valve".  One description is freighted with meaning in the form of context and

explanation; the other is not.  The event they describe is the same.

The functional description is illuminating in several important respects.  All by

itself, it helps us to understand where these entities are located (within an internal

combustion engine).  In doing so, it also helps us to understand what role the objects are

playing within this system, if we are familiar with such systems (they are opening a

passage to a cylinder in order to allow gasses in or out).  At the same time, the description

makes it obvious where these objects came from (they were designed and produced by

human engineers).  A functional description thus takes what would otherwise be a simple

physical depiction and implicitly illuminates things such as the systemic context in which

---

[1] A functional description is a description of an event involving the attribution of function to various parts, implicitly or explicitly.  It is of course impossible to say more precisely what would constitute such attribution without prejudicing the various theories of function to be introduced.

the objects are found, the place of the objects within the system, the contribution they make there and how the objects arose. All this is implied in the simple functional characterization of valve, camshaft and rocker arm.

Such a functional description also invites more finely detailed characterizations of the action. Does the valve open and close smoothly? Is the lash of the rocker arm correctly adjusted? Does the valve seat properly, preventing pressure loss in the cylinder when closed? These further descriptions pertain to the *evaluative* performance of the function, and as such they are applicable only to functional descriptions. In the strictly physical description, it would make little sense to ask whether the rounded object strikes the flat metal piece well, whereas the questions above are actually posed by mechanics and engineers every day.

To be sure, the mechanistic and functional descriptions are not at odds. The purely mechanistic depiction remains intact. But such a narrow description often omits – or at least obscures – what is most significant or relevant in functional systems. Without wanting to stretch the analogy too far, it might be said that a mechanistic picture provides the raw text, but the concept of function helps us to read and interpret that text in a way that is more meaningful to us.

For example, an entirely accurate and complete mechanistic description of the cardiovascular system would be immensely useful. But if that was strictly all we were limited to, we would not know what to do with it or what to make of it. We would have no idea of the reasons behind the various interactions beyond those provided by the laws

of physics. We would understand, for example, how oxygen bonds to hemoglobin on the molecular level; but from this alone we would have no idea why this interaction exists, what, if anything, it accomplishes for the larger system, or how it relates to why this molecular structure is present and not some other. Such a description would be, to borrow Toynbee's colorful caricature of history, "just one damn thing after another". A functional description is thus a critical supplement to a basic physical description, allowing it to be contextualized within a larger, more meaningful picture of how and why it may be happening.[2]

Yet it should also be clear that, in comparison to the artifact example, the attempt to apply such a rich level of interpretation to a biological system – where intention can no longer be appealed to – raises at least as many questions as it hopes to answer. These include: Does attributing functions to biological systems imply that they have a purpose? If so, how can this be specified and justified naturalistically? Must attributing functions to biological systems help to explain how such systems arose? If biological systems have functions, what is the basis, if any, for evaluating their performance? Is the concept of function as applied to biology similar to or different than that applied to artifacts? These questions and others present substantial challenges to the deep intuition that biological organisms can be characterized functionally. Moreover, while these questions are interesting in their own right, there are also important reasons to care about their answers.

---

[2] Of course some kinds of higher-level explanations are available apart from functional descriptions. For example, when meteorologists study the atmosphere, they treat large air masses as discrete units, and derive general laws which govern their interaction. This explanatory model does not involve any attribution of function. However, this is still essentially a physical level description – just a more abstract one.

**The Stakes of Natural Function Attribution**

The question of the status of function in natural systems is about much more than just the propriety of a metaphor or the appropriateness of a terminology. We cannot linguistically sidestep these issues simply by couching function language in caveats partly because the concept of function in biology has important ramifications in a number of areas. How it is to be interpreted and where it is to be applied is a significant philosophical question with material ramifications. The question of natural function is crucial not only to philosophy of biology, and to some extent philosophy of science more generally, but also lies at the nexus of several critical issues in philosophy mind, philosophy of medicine, and via the latter, even bioethics and policy.

A major philosophical contact point with biological function is found in philosophy of mind. Most famously, the theory of functionalism explicitly invokes the concept of function in order to provide a foothold for mental states within a behaviorist account in terms of the functional role of their content. In essence, functionalism interprets mental states as functional roles within an essentially deterministic input-output system in the brain. As Ned Block describes it, "According to functionalism, the nature of a mental state is just like the nature of an automaton state: constituted by its relations to other states and to inputs and outputs" (Borchert 1996). Because these same abstract states can be concretely instantiated in different ways in different systems, functionalism provides a means to identify mental states independently of their physical correlates. Thus, function is utilized to provide a characterization of mental states in non-mental

terms, as required by behaviorism's black-box approach to the mind. Function in the brain is conceived of as a relatively straightforward set of dispositional input-output maps. However, this is essentially merely a formal definition of mental states. The only characterization of function is as multiply-realizable schematic relationships of input and output. The richer set of functional ramifications detailed below, including purpose and teleology, are not relevant to functionalism.

These deeper aspects of function are relevant to other areas of philosophy of mind, including a defense of higher-level mental properties against physical reductionism. Psychology, for example, can be thought of as a holistic functional interpretation of the physical neural hardware, one which might survive reduction schemes applied at that physical level. A direct appeal to molecular interactions as the ultimate explanatory level in the brain has tremendous potential for dehumanization. Such low-level mechanistic explanations effectively displace higher-level explanations, such as the psychological. The idea of function provides a conceptual model that resolves this dichotomy by integrating mechanistic explanations with equally valid higher-level explanations without either undermining the other. While a functional device follows mechanistic laws, it does so in the performance of an effect which has meaning and value for the system in which it is found. In a like sense, neurochemical and neurological explanations need not obviate higher-level and more semantic psychological explanations. This more intelligible psychological meaning can thus be considered a crucial supplement to a mechanistic picture of neural action, rather than

being invalidated by it or at best made superfluous. In other words, function is explanatorily anti-reductionist. Despite this, a precise articulation of the concept of function is rarely attempted within philosophy of mind, and it is clear that not all of the available theories of functions would serve this purpose.

The normative, rather than the explanatory, dimension of function is also appealed to in addressing other aspects of the problem of mental states in philosophy of mind. For example, one of the best known and most complete theories of function, Ruth Millikan's etiological theory, was developed with the express goal of providing a naturalistic basis for a theory of mental contents. In her theory of teleosemantics, the accuracy and reliability of mental contents are ultimately informed and underwritten by the normativity of biological function, such that properly functioning neural apparatus can be relied upon to provide accurate mental representations of the world.

The issue of normativity in biological function also holds important implications beyond philosophy of mind. Functional norms in biology have obvious relevance for defining health in the organism. Therefore, the precise nature of a theory of function holds profound ramifications for theories of human health, and in turn for various issues in philosophy of medicine and bioethics. Health, clearly, is a goal of medicine. But is this the same as proper biological functioning? Is it something more or different than proper biological functioning? Is there even such a thing as "proper biological functioning" at all, implying as it does a single standard against which all biological behavior can be judged? As various theories of function will answer this last question in

very different ways, they may lead to various incompatible theories of health, with material ramifications for medicine. Few theories of function are developed expressly to address these questions, though Christopher Boorse's is the notable exception. Nonetheless, this is one of the most consequential impacts for any theory of function.

Finally, the most obvious significance of biological function is its status as a problem in the philosophy of biology, and more generally in philosophy of science. The primary issues are whether there can be any natural basis established for teleology and normativity, if these are taken as necessary for functional ascription. As we will see, these are the primary obstacles to theories of natural function, but these questions obviously resonate more broadly in the philosophy of biology. The disposition of the concept of natural function is thus of vital importance to multiple philosophical and even ethical issues surrounding human and biological nature.

**Function in Artifacts**

While natural function is the more problematic area, function in artifacts cannot be taken for granted. While artifact function is paradigmatic, and occasions fewer controversies than natural function, there are still many puzzles, as we will see. For example, to what extent does function in artifacts depend on intention? If it does depend on intention, does only the intention of the designers count? And if so, what qualifies as design, and distinguishes it from mere improvisation? Can a thing only have function as its designers intended, or is any use an example of function – in which case a tree limb or a rock might have function when used as a tool, despite having no designer.

Another challenge is to construct a theory of function that is as symmetrical as possible between natural and artificial cases, such that there is a single cohesive theory of function at work in both, and not two disconnected ones.

**Detailing the Concept of Function**

As we have begun to see, function itself is a difficult concept, even when it is not being applied to natural systems, arguably its most problematic context. It is a good starting point to say that "function is what a thing does". But it is not what all things do, as we have already seen. And paradoxically, even if a thing never does anything, like a knife forever in a drawer, it may still have a function. When a thing has a function, it is not even all the things which it does that are its function (e.g., a heart makes sounds, but we do not count this among its functions).

Let us now sketch a more detailed overview of the issues in the concept of function generally. What follows is an overview of the three primary aspects at stake in the concept of function. Ultimately, each of these aspects finds its roots in our intuition of what function means and how the term is actually used. Collectively, they capture most of what is at stake in the concept of function, and in the various philosophical disagreements over it. As we will see, different theories render different verdicts on the nature and importance of each of these aspects of function. Surveying each of them briefly will allow us to begin to understand the issues that separate diverse theories of function.

**Function and Explanatory Value**

As discussed above, perhaps the most basic hallmark of the concept of function is that the attribution of function offers meaningful explanations which go beyond mere physical description of an object. It does this in part by implicitly contextualizing the basic physical action of a function within a systemic context. For example, the action of the heart can be described at the physical layer as the contraction of ventricles and the ejection of fluid and the motion of valves. But this tells us relatively less than the functional description, which is that the heart is pumping blood, and the valves are regulating flow. Where it is applicable, a functional description helps to give meaning and relevance to physical systems, distinguishing them from blind mechanical interaction (as we saw above in the camshaft example).

In doing so, function also defines higher-level entities on the basis of their function. For example, the heart, whose function is to pump blood, contains valves, whose function within the heart is to regulate the direction of the flow. Without the concept of function to appeal to, it would not be as easy to distinguish these complex contiguous entities as precisely. Function enables the definition and reference to a mass of complex physical structures by a single term, unified under the rubric of the function they collaborate to provide. Of course, function is not the only way to make holistic references. Some non-functional entities or systems have parts so distinct that their holistic identification is simple: the individual planets of the solar system, for example.

But function provides a particularly powerful and convenient way of 'picking out' complex entities, perhaps with many disparate elements, as wholes.

It would seem that some kind of explanatory value is implied in any concept of function. As Larry Wright, one of the originators of the function literature put it, "functional ascriptions are-intrinsically, if you will, explanatory. Merely saying of something, X, that it has a certain function, is to offer an important kind of explanation of X." (Wright 1973, 154). Without some kind of explanatory value, function attribution would seem to have nothing to offer. We would therefore expect all philosophical theories of function to retain an explanatory dimension. It is perhaps the only universal hallmark. However, that is not to say that all theories of function explain the same things, or explain things in the same fashion. Wright, for example, had a very different notion of what a functional ascription should explain than did Robert Cummins. It is only to say that some explanatory value must be part of anything aspiring to be called a concept of function.

**Function and Purposefulness**

Intuitively, function seems to imply purposefulness. Indeed, in ordinary usage it is nearly synonymous with it. Saying that the function of the heart is to pump blood is typically to say that this is its purpose – what it is *for*. The same holds for artifacts, whether a car battery or something as simple as a screwdriver: their function seems just another term for their purpose. But if this is true, then there must some account of the teleology behind such purposefulness.

In the case of artifacts, the general source of this teleology seems apparent: the intentions of the people who make and use the artifacts. However, note there is no guarantee that these two sources of intentions will in fact align perfectly. It may be the case that an artifact designed for one purpose is in fact put to another purpose in the hands of a user. This may be as simple as a screwdriver being used to pry open a paint can, or as sophisticated and subtle as modulated digital data being sent over an old-fashioned analog telephone switch, rather than the voice call which it was designed to carry. This distinction between design intention and use intention in artifacts is not as incidental as it may seem, and will be explored in what follows.

In the natural case, things are, as will typically be the case, more problematic. It remains to be explained how the notion of purpose can be fleshed out in the context of natural organisms who, we hold, lack a conscious creator. Function in biological systems thus faces one of its biggest quandaries: a source of natural teleology that can ground the intuitive function-as-purpose concept in organisms; or, alternatively, a theory of function that can do without the affiliation with purpose. We will see both approaches attempted.

**Function and Normativity**

As we know too well from everyday life, a functional entity does not always flawlessly fulfill its function. Hearts becomes diseased, and cars can fail to start. What is notable in these familiar situations is that the functional entity in question does not seem to lose its functional attribution. We would never think to say that an enlarged heart is no

longer supposed to pump blood simply because it no longer is in fact able to do so. We would typically say that it *should*, and if it does not then it is *bad*.

In short, we make a firm distinction between a thing *having* a function and that thing *fulfilling* its function. And we tend to think that it continues to have the function even when it fails to perform it. This is where the normative dimension of function is revealed: we say that the heart should pump, and that the car should start. When they do not, we do not merely say that it is bad that they do not; we go further, typically, and say that the items *themselves* are bad. A car that won't start may have a 'bad battery', and a heart that won't effectively pump blood is may be referred to as a 'bad heart'. Of course, there are a great many things that also will not pump blood or start an engine. A glass of water substituting for an electrochemical battery will lead to a non-starting car. But we would never think to say that a glass of water is a bad battery (though we might say it *makes for* a poor battery). In other words, a battery is a battery even when it no longer functions, just as a heart is a heart even when it fails. But if a thing is identified as what it is via the purpose that it serves, it is at least curious that it does not lose that identity even when it fails to serve that purpose.

Here, too, we begin to see the particular conceptual challenges with attributing function to natural entities such as hearts. When a battery fails, there is obvious recourse to a standard *by which* it failed: the intentions of its human makers. A car battery was always *intended* to start engines, and when one fails or ceases to do so, all things being equal, we may say that it has failed or is bad *by the lights of* those original intentions.

But there are no comparable intentions to refer to in the case of nature.  If the heart belongs to a human, or is even just one that a human cares about, we of course have human preferences and desires to appeal to.  We can say that it is bad, by the standard of our desires, if the heart declines in the quality of its functional performance.  But this is a far cry from the establishment of standards of performance based on pre-defined intentions, or some other pre-existing standard of performance, which the heart has now failed to live up to.  If function is to be attributed to the parts of organisms, this issue will have to be confronted one way or another.  Just as with purposefulness, what is required is a non-intentional standard for normativity, or else the normative dimension of function must be dismissed as somehow not integral to the concept of function.

## A Unified Account of Function

The hope is to arrive at a single detailed concept of function which fits both the natural and the more familiar artifact cases.  Most philosophers aim to produce a single theory of function, applicable to both natural and artifact cases (as well, often, as other contexts).  Unification of the concept is typically an explicit goal.

Two points should be made clear at this juncture.  All writers agree that the concept of function, in some form, can and should be applied to organisms.  None of those we will encounter advocate abandoning the application of function to biological life.  This is probably so for a number of reasons: the fact that biologists, physicians and lay people alike would continue to do so anyway, and the fact, which may explain why they do, that the intuition of function in biological life is so compelling.  Moreover,

eliminating function references in biology would leave us with a substantial cognitive deficit. As we explored above, the explanatory and interpretive value of function is significant. If it were abandoned in biology, it is not clear that we could make do without it, or what if anything could replace it. Abandoning function would also leave us with a basic inarticulateness; we could not even refer to a gene as encoding for a trait, since such terms are inherently functional. So, abandoning function talk in biology is not an option. The goal, then, is to find a concept of function that is still explanatorily fruitful but which can be applied in good naturalistic faith, without importing intentions, evaluative norms or teleological purposes that cannot be justified naturalistically.

Secondly, most writers agree that the concept of function to be applied to natural systems should be as much like the one we apply to artifacts as possible; they agree that there should not be artifact-function and organism-function in isolated conceptual universes. They seek a *unified account* of function.

### The Plan for the Work

The primary task is arriving at a theory of function which covers both artifacts and natural organisms in a common framework while accommodating as many of our intuitions about function as possible. Each chapter in what follows will be dedicated to explaining and critiquing a single closely-related class of theories of function, which I term an "approach". An approach is not a particular theory, but a small family of theories united by their assumptions of what function attribution is about. Within each approach, I will detail one or two prominent theories as paradigms of the approach. After fully

examining at least one of these, I will subject it to critique focusing primarily on its logical and metaphysical consistency, as well as its consonance with ordinary intuitive function attributions.

The two main poles of the 'function debate' as it is typically framed in the philosophical literature today are the etiological approach and the systemic effects approach. These will be examined in order in chapters 2 and 3, respectively. We will then turn to the less popular option of contributions approaches, and develop a novel version of such an approach, to be known as Assigned Contribution theory, in chapter 4. Following this, in chapter 5 we will consider the prospects for a more radical and entirely novel alternative approach, which applies many of the lessons of the contributions account but takes its inspiration from emergence theory and looks to the formal properties of the item to attribute its function. Finally, the theories, and the notions of teleology within each, will be compared in chapter 6. An abstract for each of the remaining chapters follows:

 **Chapter Two**

This chapter will address what is currently the most widely accepted class of function theory, the etiological theories as exemplified by Larry Wright and Ruth Garret Millikan. It will explore the problems with tying etiology so closely to function, including the difficulties with knowing the etiology of parts, whether function can be attributed without knowledge of etiology or with an inappropriate etiology, and the tenuous relationship of function etiology to natural selection. This chapter will also

include a substantial diversion into how a theory of function should be evaluated generally. Overall, the etiological account will be found to introduce many intractable difficulties and to conflict with some basic intuitions regarding function attribution, as well as with important aspects of natural selection. In addition, the natural teleology that it seeks to establish will be seen to be metaphysically problematic in its own right.

**Chapter Three**

This chapter will treat the opposite pole in the function debate, known as the systemic effects account. It will look to Robert Cummins and Paul Sheldon Davies as the primary proponents of this class. Systemic effects theories entirely reject teleology in natural function as naturalistically untenable. With this shift, problematic etiology becomes irrelevant. Systemic effects theory is an attempt to keep things simple, and to stay well within the confines of naturalism. This account of function looks only at the effects of parts in their respective systems. But is this too minimal? Can function be non-teleological and remain recognizable? Is an account of function that denies malfunction tenable? The additional problem is that function as mere effect threatens to apply to anything, including things we would never otherwise be inclined to attribute function to, such as rainclouds. Thus, the approach requires some account of just which effects are functions, and which are not. Davies suggests a limitation to hierarchical systems with 'capabilities'; but in formal terms this turns out to look a lot like purposes. If we can't avoid invoking purpose (even implicitly) to give meaning to function

attribution, how do we supply the requisite teleology naturalistically and avoid the both

the methodological and metaphysical issues encountered with the etiological approach?

 **Chapter Four**

While etiological theory is the most popular account, and systemic effects its most

notable challenger, other varieties of function theory do exist.  One of these is the

contribution account, which looks to the contributions that parts make to the teleological

systems in which they occur in order to attribute function to those parts.  Taking cues

from both etiological and systemic effects, I will develop an elaborated, novel version of

this kind of account.  The virtue of this approach is that it lets the teleology of the system

inform the teleology of the functional parts, avoiding the problems associated with

tracing the etiology of each individual part.  This approach will also allow us to

distinguish function performance from function 'assignment' in a manner not possible in

either previous account.  This distinction is key to solving a number of conundrums.  But

this approach obviously still relies on some account of the teleology of the overall

system, which for natural systems remains problematic.  This means that natural function

can only exist if organisms as a whole can be shown to be teleological.  A further

difficulty is that a strict contributions account is not easily able to distinguish on principle

between the contributions (and thus the function) of proprietary differentiated parts such

as the heart and of necessary but nonproprietary and undifferentiated components, such

as water.  Combining the insights of both these issues will drive us to the final alternative.

**Chapter Five**

  This chapter will consider a final, radical alternative: a completely formal analysis of function, one which looks only to the properties of the functional item and the way in which those are shaped by the system to which it belongs. This is an attempt to secure teleology entirely without etiology. It does so by finding the feedback pattern of teleology evidenced in form alone, rather than in causal relationships. This will still allow us to distinguishing function performance from function possession based on degree of fit in properties between part and system. This approach also unifies natural and artifact function to a greater degree than the alternatives, as it automatically applies to both artifacts and natural systems equally due to its complete independence from etiology. Without any reliance on etiology, there is no longer any principled gap between natural and artifact systems in the first place. This theory of function provides a new and perhaps promising paradigm for further work in function theory.

**Chapter Six**

  The concluding chapter will consider what teleology is most centrally about, and evaluate each approach anew in terms of the concept of teleology implicit in each. To do this, the concept of teleology itself will be examined in what seems to be its purest form: conscious intention. It will be seen that such teleology is essentially a pattern of mutual interaction between means and ends, such that ends are the cause of their own means. Ultimately, the choice of function theory will be seen to turn on whether non-intentional teleology is a coherent possibility. If objective natural teleology is considered

metaphysically plausible, the properties approach offers important advantages. If it is not, the Assigned Contributions approach is superior. But even in this case, further development of the properties approach into a full theory of function is recommended.

**A Note on Strategy**

This work is a search for the best philosophical theory of function, in terms of internal consistency, coverage, explanatory power, and the correspondence of the resulting function attributions with ordinary usage and intuitions of function. The job of each chapter is to present the best possible case for each approach to function: etiological, systemic effects, contributions, and properties. Each chapter examines an entirely distinct approach on its own merits, repairing it when possible to make it as strong a competitor as possible before subjecting it to critique and seeing where weaknesses remain. An approach may be found inferior or inconsistent, and then exploring it was generally good only for seeing where it went wrong. However, the theories themselves are not generally subject to revision once the chapter concludes. By then, we will have already said all that we can about them – what will remain is to see what the alternatives are. Sometimes, the strength of one approach may highlight the relative weakness of another. In many cases, the weaknesses of one theory will propel us towards seeking a particular sort of alternative approach.

CHAPTER TWO

THE ETIOLOGICAL APPROACH TO FUNCTION

**Introduction**

**Overview**

This chapter will address the major class of function theories known as the etiological, which are currently the most popular class of function theories in analytic philosophy (Buller 1999), although critiques of the approach seem to be gathering momentum and that dominance may not last very much longer (Davies 2001, Lewens 2005). In this class of theories, function is tied explicitly to origin, such that for a thing to have a function is for that thing to have arisen in order to perform it. Below, I will examine the nature, strengths and weaknesses of this type of function theory at length. I will focus on the two most prominent proponents of the approach: Larry Wright, who is credited with inventing the genre; and Ruth Garret Millikan, who is generally recognized as having developed the most sophisticated version of the account.

I will consider Wright's theory of etiological function first, as it is the prototype upon which most later theories of the type are based. Wright's theory also remains the most schematic and essential account, and is therefore an excellent introduction to the logic and intuitions motivating this class. However, because Wright's theory is closer to a conceptual foundation than a fully worked-out theory of function in its own right, it is not an ideal subject for extensive theoretical critique.

As it is generally regarded as the most sophisticated and well-developed account in the literature, I will treat Millikan's theory, known as the theory of "proper function", as the paradigm of etiological function theories. I will explicate this theory in substantial detail, for its subtleties (perhaps due to the density of Millikan's presentation) are often overlooked, yet they can make appreciable differences in the way that the theory should be approached.

One such important "detail" is the fact that Millikan objects to her theory of proper function being taken as a general theory of function at all. Nevertheless, it is almost universally taken to be one in the function literature. However, taking care to closely examine what it is that Millikan objects to and why will prove useful, for it raises difficult but important issues about just what a theory of function should aspire to in the first place. This discussion will therefore concern issues applicable to every theory of function, and enlighten the overall project. Other important refinements of Millikan's which are often overlooked in discussion of her theory will also receive attention below, such as her distinction between different iterations of function orders, the rationale for making all proper functions dependent upon type membership, and the way in which these two ideas are connected.

Following this close exegesis, I will subject the theory of proper function to critique, often turning to arguments advanced by two of its leading critics, Robert Cummins and Paul Sheldon Davies, both proponents of the competing systemic account of function to be examined in chapter 3.

As a general theory of function, proper function is susceptible to a number of problems resulting from the narrowness of its interpretation of function, and the constraints, both real and epistemic, which it places on attributing function.  One notable counterexample will figure prominently: how William Harvey was able to deduce the function of the human heart correctly centuries before Darwin devised the theory of evolution by natural selection, thereby accounting for its etiology.  For a theory that predicates natural function attribution on selection etiology, this is problematic in a number of ways.  Converse examples, where no history of natural selection is applicable while biological function is nonetheless typically attributed, will also be considered.  Separate from such issues, there are problems relating to the characterization of natural selection upon which the theory of proper function relies so heavily.  If natural selection in fact does not work quite as the theory of proper function demands, this poses serious problems for the integrity of the theory.

Ultimately, however, the most profound problems will be seen in proper function's most profound (and proudest) claim: to be able to account for both natural purposes and natural malfunctions.  Both the metaphysical/ontological dimensions and the logical consistency of these claims will be closely examined.  We begin now with the basic intuitions that motivate the etiological approach to function.

**The Link Between Function and Etiology**

What makes something – anything – have a function?  This is perhaps the simplest question that can be asked about function, and, as simple questions often are, among the hardest to answer.

Since natural functions, lacking any direct appeal to conscious intentions, are clearly more problematic, we might do well to begin by thinking about artifacts.  What makes an artifact functional?  This begins with what makes something an artifact in the first place.  An artifact can be nearly anything – from a sophisticated microchip to a simple stone tool. Let's again start with the simplest case.  What makes a stone tool an artifact rather than simply a stone?  We might distinguish it first on the basis of its overall form, as an archaeologist might.  Does the stone have a sharp edge?  A place for holding that fits a human hand?  These are good starting points.  But they are perhaps not decisive.  A stone might have very similar features via natural processes like erosion or being fractured in a landslide.  To defend against this possibility, an archaeologist might look more closely for clues.[1]  Does the rock show a pattern of chips, as if it had been struck with another object to arrive at the form we observe?  Or does it look as if it simply fractured naturally?  Notice what is now being sought here: although we still look to the stone for evidence, it is no longer precisely about the stone itself at this point.  Its properties and how closely they match what it can be used for are secondary.  Instead, in looking for evidence of handiwork, what the archaeologists are looking for is evidence of

---

[1] An archaeologist would of course also look beyond the object itself into the environment in which it is found – but let us keep our focus on the object itself.

how that form *came to be* – they are looking for clues not to its properties but to the *origin* of those properties.

And so we come to what seems to be *the* central factor in what makes a thing an artifact: whether or not it was intentionally fashioned by human beings. A stone is just a stone until human handiwork transforms it – however modestly – into an artifact. This is all that separates a stone knife from a sharp stone. Indeed, the definition of an artifact is just something made by human hands, as evidenced by the word's etymology.

Given that it is origin which defines an artifact, and that artifacts are the clearest paradigm of function, many thinkers believe that origin is the best place to ground a theory of function.[2] In fact, the contemporary philosophical debate about function is generally traced back to this intuition, as expressed by Larry Wright in 1973.

## Wright's Etiological Function

**Design Intention and Artifact Function**

Wright saw the first task of any theory of function as distinguishing functions from accidents. In the example above, a rock that had been in a landslide and been broken into a shape like a stone knife does not have any function. On the other hand, a stone that had been chipped away at by a human with the conscious intention of producing a stone knife does have a function. The former is an accident – the latter is a

---

[2] This is not to prejudge that all artifacts have function. For example, there are byproducts produced by humans which meet the technical definition of artifacts, such as a discarded scrap from a sawn board, which seem to have no function. But the immediate issue is the nature of function, not whether all artifacts have it. If there are indeed artifacts without function, then the theory of function would inherently have to account for that, just as it would explain why not all natural entities have function. We still have to arrive at a theory of function first.

functional artifact.  For Wright, origin is overriding.  Even if the accidental stone is

sharper, easier to grasp, and cuts better, it has no function.  And even if the stone knife is

dull and fragile, it still has the function of cutting.  Usefulness alone is not a criterion

because it "does not make the right function/accident distinction" (Wright 1973, 149).

What determines the function of an object is entirely the intention, if any, behind its

origin.  "If X was designed to do Y, then Y is X's function regardless of what

contributions X does in fact make or fail to make." (Wright 1973, 153).

Of course, that does not prevent the sharp stone from being *used* as a knife, and

this involves a kind of intention.  But Wright draws a sharp line between intention in use

and intention in creation.  "When something does something useful by accident rather

than design … we signal the difference by a standard sort of "let's pretend" talk. Instead

of using the verb "to be" or the verb "to have," and saying the thing in question has such

and such a function, or saying that is its function, we use the expression "functioning as""

(Wright 1973, 147).  As is clear from the reference to it as mere pretense, "functioning

as" is not considered true function, but only a loose allusion to function.  True function

only obtains via the intention behind the creation of an artifact.  One uses a stone knife,

but one only uses a sharp stone "as a knife".  There is no function without a

corresponding design intention.  Yet, with such intention, function performance is not

even required.

For example, suppose the function of a switch on the dashboard of a car is to

control the windshield washers.  For Wright, this is true just so long as this was the actual

intention behind its design and manufacture. Suppose that, despite this intention, a

worker on the assembly line forgets to connect a wire on a given car. As a result, unless

repaired, that switch is permanently disabled for the life of the car – the driver flips it,

and nothing ever happens. What is the function of the switch? According to Wright, its

function is the very same as if it worked flawlessly: to operate the washers. The fact that

it does not and cannot is incidental. Its function was affixed when it was designed and

made (however poorly), and that function does not depend on what we might call its

'functionality' or 'ability to function'. Thus, "intention is so central here that it allows us

to say that the function of I is to do C, even when I cannot even do C. If the windshield

washer switch comes from the factory defective, and is never repaired, we would still say

that its function is to activate the washer system; which is to say: that is what it was

designed to do." (Wright 1973, 146). Note also that while ordinarily the labeling of the

switch would be a clue to its intended function, this too is formally independent. To

extend the example, suppose there is a switch labeled "washers" in James Bond's Aston

Martin, but what Q actually designed it to do is release the oil slick. This is then its

function, even if it does not work, and even if the driver does not know that this is its true

function. For Wright, the intentions behind the origin of the artifact <u>alone</u> determine its

function.

**Functional Etiology**

Whatever its overall merits, this account is at least clear enough for artifact

function. But the goal of a theory of function is typically to account for function not only

in artifacts, but in natural organisms as well, a goal Wright explicitly embraces: "I am looking for an analysis which will unify conscious and natural functions" (Wright 1973, 148). As a result, he is "committed to finding a way of stating what it is to be a function – even in the conscious cases – that does not rely on an appeal to consciousness" (Wright 1973, 143). How then can basing function in artifacts on conscious intention in creation also be applicable to natural organisms unless one posits a parallel conscious Creator? As Wright puts it, "it might appear that this [account] commits us to the view that natural and consciously contrived functions cannot possibly be the same sort of function. If conscious intent is what determines the function an artifact has got, there is no parallel in natural functions." (Wright 1973, 146).

However, Wright claims that this is a mistaken perception. He has not, despite appearances, in fact based artifact function on intentions. Artifacts by definition only come to be through intentional design – but this is simply a specific truth about artifacts. The key for function attribution is not intention *per se* but the general schema: functions are the result of an origin process which brought them about. But to have function, not just any origin process will do. The key for function is that the origin of the object must be *traceable* to its function. Functional objects (i.e., objects that have function) must arise *due to* their functional properties. For example, the function of a screwdriver is to drive screws; this is so not simply because it can, or even because it is in fact used to do so. The function of the screwdriver is to drive screws because the screwdriver *was only created in the first place so that it might do so*. The tool owes its very existence to its

function.  It would not have been created at all but for the sake of this functional capacity. In other words, its function depends on its origin just because its origin depended on its function.  There is a bi-directional or co-dependent relationship between origin and function.  "So when we say that Z is the function of X, we are not only saying that X is there because it does Z, we are also saying that Z is (or happens as) a result or consequence of X's being there." (Wright 1973, 160).

One important benefit of this approach is that it makes functional ascription helpfully explanatory: attributing function implicitly provides an explanation of origin. We know something about origin by knowing something about function, and vice-versa. This gives the theory of function some positive content, illuminating some distinctive aspect over and above simply pointing to what the item does or can do.

But for Wright "making sense of the function/accident distinction [is] perhaps the *primary* aim", and he sees this account as achieving this aim better than any alternative account (Wright 1973, 142, 158).  Simply citing what a thing can do or is used for does not assure us that the object in question and its useful properties did not come about by accident, like the sharp stone.  But if having a function implicitly means that that function played a role in the origin of the object, then by definition the function of the object can never arise by accident.  Function has been set apart from accident on principle, and so the concept of function is preserved as distinct.

With the core logic of function attribution clarified, Wright concludes that in general, "functional ascription-explanations are in some sense etiological, [they] concern

the causal background of the phenomenon under consideration. … functional explanations, although plainly not causal in the usual, restricted sense, do concern how the thing with the function got there." (Wright 1973, 156).

But what kind of "causal background" can this be? Surely, the claim cannot be that the properties of a present object were directly causally effective in bringing about its own origin. This would be to reverse the temporal order of causality itself. How could the properties of an object that does not yet exist 'cause' its own creation? In the case of designed artifacts, the appropriate mechanism is clear enough: the conscious, intentional imagination of the designer mentally envisions the properties of the object he has yet to create, and this mental vision then guides its actual creation. In this sense, the (envisioned) functional properties of the object are responsible for (causally contribute to) the creation of the object that arises as a result. The actual properties of the object cannot predate and cause its own creation – but the envisioned properties in the imagination of the designer certainly can, and typically do. In this sense, conscious intention indeed plays an essential role in accounting for the function of artifacts, but this is but a special case of a more general logic. What in nature could substitute for conscious intention and still provide the requisite "causal background" for functional objects? Wright claims that evolution via natural selection can play this role.[3]

---

[3] Interestingly, Wright reserves selection etiology exclusively for natural systems, while design is the exclusive etiology for artifacts for Wright. Yet human-driven selection (e.g., improvisation of parts) may be an integral part of artifact creation as well. In addition, it could be argued that wholesale improvisation, like the sharp stone selected, retained and regularly used as a knife, should also be considered tools with function, if not artifacts. These issues will resurface in chapter 4.

**Functional Etiology and Natural Selection**

"We can say that the natural function of something – say, an organ in an organism – is the reason the organ is there by invoking natural selection. If an organ has been naturally differentially selected-for by virtue of something it does, we can say that the reason the organ is there is that it does that something.  Hence we can say animals have kidneys because they eliminate metabolic wastes from the bloodstream; porcupines have quills because they protect them from predatory enemies; plants have chlorophyll because chlorophyll enables plants to accomplish photosynthesis; the heart beats because its beating pumps blood." (Wright 1973, 159).  Just as imagined functional effects accounts for the presence of functional properties in artifacts, natural selection accounts for the presence of functional traits in natural organisms by way of the selective benefits they conferred.  Thus, it is because the heart's pumping action helps keep an animal alive that it was selected for.  If we want to know why animals have hearts at all, we need only look to the selective benefits they confer – and those selective benefits are their functions.

This makes natural and artifact function tractable by the same general schema: "Not only is chlorophyll in plants because it allows them to perform photosynthesis, photosynthesis is a consequence of the chlorophyll's being there. Not only is the valve-adjusting screw [in an engine] there because it allows the clearance to be easily adjusted, the possibility of easy adjustment is a consequence of the screw's being there." (Wright 1973, 160).  Artifacts and natural organisms have very different origins.  But in both cases, function is accounted for by etiology: a thing has a function if (and only if) its

origin depends on it having that function.  As Wright puts it, "Both natural and conscious functions are functions by virtue of their being the reason the thing with the function "is there"" (Wright 1973, 164).  Anything else is mere accident: "The etiological condition is what distinguishes consequences that are functions from those that are not" (Wright 1976, 78).  Since distinguishing function from accident is often taken to be a core task of a theory of function, this general account has great intuitive appeal and has enjoyed widespread acceptance and further development in analytic philosophy over the past several decades.  David Buller goes so far as to call its core ideas "as great a consensus as has been achieved in philosophy" (Buller 1999, 19-20).

Indeed, while Wright is generally credited with originating the so-called etiological theory of function, others have developed versions of the theory with far greater detail.[4]  The most rigorous and extensive successor is Ruth Garrett Millikan's theory of "proper function".  Rather oddly, Millikan claims that, "Various published remarks to the contrary, there is no overlap at all between Wright's analysis of function and mine." (Millikan 1989b, 297, note 5).  The basis for this disclaimer is her contention that Wright's etiology is not sufficiently 'etiological', in that it does not go back far enough in history, and is not precise enough in describing that causal history.  Yet given that Wright makes explicit appeal to natural selection to explain function attribution in organisms, and did so 11 years before Millikan, it seems reasonable to read Millikan as a successor to Wright, "overlap" or not.

---

[4] It should be noted that Robert Cummins (Cummins 1975) actually traces the original idea of function accounting for the presence of the functional item to (Nagel 1961) and (Hempel 1994a).

**Millikan's Proper Function**

Wright proposed the schematic of etiological function as a self-contained general theory of function – that is, it was developed as an attempt to systematically answer questions such as: "How does one distinguish function from accident?" and "How does one account for function attribution in both artifacts and natural organisms in a common fashion?" His ideas seem to answer these questions, at least to his own satisfaction. Yet he left some of the important details and implications of these ideas unexamined in (Wright 1973) and even in his book expanding on that landmark paper (Wright 1976).

**What is Proper Function?**

Millikan's proper function was conceived with a somewhat different goal in mind, as we will shortly see. Yet in its core concept, Millikan's notion of 'direct proper function' is quite like Wright's: "A function F is a direct proper function of x if x exists having a Character C because by having C it can perform F". As Millikan immediately adds, "Notice how close this is to the idea that x exists in order to perform F." (Millikan 1984, 26). We are to understand this 'because' in terms of historical origin: things like x with C exist "because there were things that performed F in the past due to having C". As in Wright, the presence of the thing is explained via the ability of the thing to perform the function attributed to it: the existence of the enabling characteristic C is due to the fact that with C x is able to perform F.

Millikan offers a more complete formal definition of proper function; however it will take some time to unpack all of its implications and terminology. This complete definition is as follows:

> (1) Certain ancestors of *m* 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 *m*, C correlated positively with F over a certain set of items S which included these ancestors and other things not having C.
> (3) One among the legitimate explanations that can be given of the fact that *m* exists makes reference to the fact that C correlated positively with F over S, either directly causing reproduction of *m* or explaining why R was proliferated and hence why *m* exists. (Millikan 1984, 28)

Roughly, the elements in the definition can be understood as follows:

*m* is the item or trait to be attributed function (a token)

F is the function to be attributed to *m*

R is the reproductively established family to which *m* belongs (a type)

C is the 'characteristic' or the relevant properties which allow *m* to do F

S is the set of actual items wherein the correlation between C and F obtains

To begin interpreting this further, note first that the connection between the item *m* (with the characteristics C that allow it to perform its function) and the actual performance of the function need not be airtight. The relevant criterion is correlation – not universal consequence. Does the presence of *m* with C tend to be associated with the performance of function F more often than the alternative scenarios which are missing either m entirely, or have m but without its relevant properties C? If so, then we can say that there is a correlation between them, which satisfies this part of the condition for proper

function.  This is a sufficient statistical correlation, not necessarily a reliable causation: "such a function should be thought of not as an invariant function or as an average function but as a function performed in a critical mass of cases of actual use" (Millikan 1984, 32).

Note also that the definition allows that there may be cases with *m* and C but without F occurring.  *M* and C need not be either necessary or sufficient for the performance of F – they must just make the performance of F more likely.  How much more likely is a judgment that is up to us.  This correlation, however, must be the result of a direct causal connection between *m* with C and F.  For example, chlorophyll makes the occurrence of photosynthesis far more likely because chlorophyll makes a causal contribution to photosynthesis.  But the plant may be in the dark, so none actually occurs even with chlorophyll.

Hypothetically, there could also be other compounds which also enable photosynthesis, such that there could be cases of it without chlorophyll present. However, another entity that is *better* correlated with the function is held to be unlikely: "If certain other things had correlated better with these functions, the chances are these other things would have been reproduced or would have proliferated instead." (Millikan 1984, 27).  Nonetheless, the theory does not seem to prohibit the existence of other (particularly lesser) correlations in principle.

Moreover, as the inclusion of S makes clear, this correlation is not theoretical or abstract – it is to be empirically observed to obtain or not obtain in a given set, S, of actual historical cases for which we have data (or we at least imagine that we do).

Finally, then, given that *m* with C is positively correlated with the performance of F, direct proper function requires that there exists a "legitimate explanation" that explains the presence of *m* with C by virtue of this correlation. This "presence" may either take the form of direct reproductions of *m*, or of the proliferation of its overall type R. Here again is the core content of all etiological function theories: "products of evolution have in common with various other kinds of products the fact that they are reproduced or continue to be proliferated because they, *rather than certain other things,* have been associated with certain functions" (Millikan 1984, 27).

It now remains to explain two further components of the above account, the reference to the family R, and the reference to the function as 'direct'. Each of these represents a substantial and rigorous expansion of Wright's much more casually sketched idea, and because of their complexity each merits its own section here.

**Millikan's First Major Refinement: Reproductive Families**

One of Millikan's major refinements is to systematize the notion of etiology by making it a matter of reproduction based on class membership. Whereas Wright spoke in very general terms about origin via 'design' or 'selection', Millikan takes care to define etiology more precisely in terms of membership in a functional class. Thus, functional

origin is not usually origin *sui generis*, but origin as a member of a family – a family

which has, as it were, survived by making a 'family business' of one or another function.

For example, "Artifacts that have been serving certain functions known to those

who reproduce them and that are reproduced on this account (e.g., household

screwdrivers) have these functions as direct proper functions" (Millikan 1984, 28). A

token's membership in this family, whose lineage owes to its function, has this function

automatically – even if the token cannot itself perform this function, or is never called

upon to do so. The fact that it is a member of a class which has been perpetuated (*as a*

*class*) due to this function means that every member has this function as part of its

inheritance: "if any member of a reproductively established family has a direct proper

function, all members of which this member is an ancestor have this proper function too."

(Millikan 1984, 28).

By making items part of a functional class established by lineage, Millikan seems

to provide a stronger basis for establishing what constitutes appropriate function. In

artifacts, for example, Wright was forced to appeal to the intentions responsible for the

creation of the artifact. Although he did so under the general principle that the presence

of the functional properties must be due to of those functions, the only content we have to

appeal to is design intent. Thus he says, rather loosely, that a screwdriver has the

function of driving screws because this was the intention behind its creation. Millikan

seeks to systematize this, sidestepping the explicit appeal to intentions, and strengthening

the parallel between function in natural and artificial objects. She, by contrast, says that a

screwdriver has the function of driving screws because it is a token (copy) of a type which has historically been perpetuated due to this service. If a class of objects, whether screwdrivers or hearts, has survived (i.e., been reproduced) by virtue of that function, then all members of that class have that function as part of their inheritance, *even if given tokens cannot actually perform it*.

The basis for establishing purpose, and thereby normativity, has thus been subtly enhanced. We no longer need to show that a given token was *itself* made to serve a particular purpose on a per-case basis; we need only show that this token is a token of a type which has as a type been perpetuated only so that it might serve this purpose. By appealing to class membership, we can more readily establish function, and enhance the parallels between manufacturing of artifacts and natural selection in organisms. Thus, "A direct proper function is a function that an item has *as* a member of a reproductively established family" (Millikan 1984, 27).

As we will see, although Millikan does not make this explicit, the most critical benefit of this appeal to classes is to align the theory better with natural selection, which operates on classes (most notably species) but cannot be said to account for the properties of any given token in the same fashion.

**Ordered Reproductive Families**

Appeal to reproductive classes also allows Millikan the opportunity to more precisely classify the nested types of propagation that may be involved: "There are two kinds of reproductively established families, first-order and higher-order" (Millikan 1984,

23). First-order reproductively established families are similar items that are all reproductions from the same original model. For example, in language, tokens of all established words are in a sense direct 'copies' of some original neologism. More significantly for our purposes, tools such as screwdrivers are (we may assume) mass-produced copies of an original prototype model. This, too, is first-order.

Most notably, genes are all direct biological copies of each other in a way that the traits and organs they code for are not. Livers, for example, are not direct copies of an original liver. Even if we imagine the liver as an organ appearing all at once (which is nearly impossible), any descendent livers would still only be produced via the reproduction of the entire containing organism, and in that as a result of the genes which coded for the liver. Genes are thus members of a first-order family, while organs are (more indirect) *higher-order*.

Higher-order reproductively established families are defined as those which are themselves the functional products of a lower-order family. An additional criterion applies, which is that this production occurs "in accordance with Normal explanations", or according to the 'usual' manner which is that proper function (Millikan 1984, 24). Because genes code for traits and organs as their *own* proper function, genes are first-order while their products are higher-order.[5] Higher-order families can also be constituted by a process whose proper function is to repeat a successful item, such as a learned behavior; the learned behaviors themselves would then constitute a higher-order reproductively established family.

---

[5] Millikan does not enumerate orders beyond the first.

There is one critical corollary to this family ordering schema, intended by Millikan to allow for the possibility of malfunction in higher-order classes.  The corollary is essentially that if a thing is produced according to the mechanisms of a higher-order family, then that thing is a de facto member of the family, *even if it is not entirely similar to the other members*.  "If anything … has been produced in accordance with an explanation that approximates in some (undefined) degree to a Normal explanation for production of [such members], then [it] is a member" (Millikan 1984, 25).

This stipulation is designed to allow for some latitude in what counts as an unsuccessful product versus something so different that it cannot be counted as a member of the family at all.  The vagueness of the stipulation is thus intentional, reflecting the inherent judgments call involved in such determinations (e.g., when a malformed organ is *so* malformed as to no longer count as a member of the family).  It is noteworthy that the definition she offers appeals exclusively to the mechanisms of production, and not (directly) to the nature of the actual product.  The presumption appears to be that only variation in process will result in variation in the product, and vice-versa that variations in product must always point to some variation in process.

Like the appeal to classes in general, this appeal to mechanisms of production as primarily determinate of class membership seems designed to conform the theory to the principles of natural selection, and to attempt to insulate the theory from charges of circularity in determining class membership.  We will later examine how successful this line of defense actually is.

**Millikan's Second Major Refinement: A Taxonomy of Function Iterations**

The second important innovation, one which complements the establishment of reproductive families, is Millikan's taxonomy of levels of function inheritance and instantiation. In the foregoing, including the formal definition, function has been referred to as a 'direct' proper function. This is the simplest form, as well as the core to which the rest ultimately appeal. The general schema beyond that is one of increasing specificity and contingency of instances of function performance, rooted in a more general direct function characterization.

For artifacts, this is often very simple, and reduces to merely a matter of more specific instances of the same kind of function – for example, driving a particular screw as opposed to screws in general. For physical artifacts or basic natural function, in particular the functions of bodily organs, the core notion of direct proper function will often suffice, since the instantiations do not vary widely. As she notes, "Body organs have for the most part only direct proper functions" (Millikan 1984, 49).

Yet Millikan requires this sophisticated taxonomy to handle cases far more complex than screwdrivers – for example, words, sentences and propositions. The same sentence or word can have very different intended meaning in different contexts, and the same object or intent can be referred to by several different words, whether in the same or in different languages. The taxonomy of such levels thus plays a central role for Millikan, and consumes a large fraction of her writing on function. Moreover, some of these principles are relevant to critiques of proper function as a general theory of

function. For example, the notion of items whose function it is to produce other items that have their own functions (as do genes) requires a schema to make such distinctions. Therefore, a basic exposition of this taxonomy is necessary.

First, a reminder that 'proper function' simply means a function whose presence is in some fashion due to the fact that it performs the function that it does. Somewhat confusingly, although providing evaluative normativity is the motivation for the idea, 'proper' in this usage does *not* pertain to any evaluation of function performance. Rather, 'proper' refers to the 'proprietary-ness' of the function to the item that has it – the sense in which the item 'owns' this function (Millikan 2002, 116).

*Direct* proper function is the most general and most basic: "A direct proper function is a function that an item has as a result *as* a member of a reproductively established family." (Millikan 1984, 27). Simply by virtue of being a heart or a screwdriver, each object has the direct proper function of pumping blood or driving screws, respectively. This level of function is the most general possible description of the overall *purpose* of an item. As such, it tells us what the item is 'doing there' – why its type exists at all in the manner in which it is typically found. Since the function applies to the type (or 'family'), it is not actually necessary that any given token (family member) be able to perform this function. This, of course, is what allows for the possibility of malfunction.

A '*relational'* proper function describes functions that involve obtaining fits between a system and its context. For example, a function that promotes a fit between an

organism and its environment would be termed 'relational'. "A device has relational

proper function if it is its function to do or to produce something that bears a specific

relation to something else" (Millikan 1984, 39).[6] For example, animals are frequently

engaged in aligning their status to the demands of their environment – for example,

hiding from predators, or avoiding temperature extremes. A chameleon has as one of its

functions matching its skin color to the background environment. This can be done either

by moving to a background that matches its present color, or by changing its skin color to

match its present environment, so long as the relationship between the two is aligned. An

animal learning a simple pattern that is not a built-in instinctual behavior, and that results

in the fulfillment of a basic need (whether in a lab or in the wild), is also an example of a

relational proper function. Functions which are not relational include those only internal

to the system – for example, the regulation of blood sugar or circulation.

Whether a function is relational or non-relational is independent of whether and

how it is a direct function. In addition, it is apparent that some functions have both

relational and non-relational aspects (or at least close contact with other functions that are

both relational and non-relational). For example, maintaining correct blood sugar is

partly a matter of internal regulation of hormones and digestion, and partly a matter of

obtaining nourishment from the environment.

Furthermore, when a relational proper function is operating in a specific given

context, we say that it performs an *adapted* proper function – adapted, that is, to the

specific present environment. With the chameleon, for example, the 'relational' function

---

[6] By 'device' Millikan simply means a functional item; this is not to imply that the device is an artifact.

is color-changing based on context; an 'adapted' function is, for example, brown with green skin in a brown and green environment.

**Derived Proper Functions**

Unlike relational or adapted functions, *derived* proper functions are to be contrasted with *direct* proper functions.[7]  Derived proper functions are second-order functions in that they are "functions derived from the functions of the devices that produce them" (Millikan 1989b, 288-289).  That higher-order producing function may itself be a direct function or yet another derived function, but ultimately the chain must terminate in the bedrock of a direct proper function which is not also derived.[8]  Derived proper functions are ultimately *derived* from more general direct proper functions.  It is worth noting that conscious intentions, depending on their specificity, may be either direct or derived.

Because derived function is higher-order, it often applies to functions which are themselves complex enough (in practice) to be satisfied in multiple ways, depending on the environment.  Thus, exclusively derived functions (that is, functions that are not also direct in their own right) are applicable to behavior and language, and less so to homeostatic temperature regulation, or circulation, where the repertoire of functions is relatively limited, and not subject to ad hoc expansion as the environment presents

---

[7] Note, however, that while they are contrasted in concept, they may overlap in practice, as described below.

[8] It is, however, an open question as to what, if any, functions are truly direct without also being derived. Millikan refers to systems that are hard-wired into the animal at birth; however, such systems are also derived from genetics, making them, too, derived.  Genes themselves may be exclusively direct, but even this is arguable.  The direct/derived schema, despite its utility, seems liable to unending regress.

opportunities for new ways in which more basic needs might ultimately be satisfied (for example, new hunting strategies, or new forms of shelter). When we rise to the level of intentional language and behavior, there are almost no limits as to how intentions might be expressed. Even so, the chain of function must ultimately terminate at "a system that was built into the animal at birth" (Millikan 1984, 47). At this level, we are dealing with direct proper functions.

For example, consider a conditioned behavior, learned by an animal as a means to satisfy a basic need. The behavior that the animal learned to do as a reliable means to, for example, obtain food is not itself a 'built-in' function of the animal. Rather, the behavior was conditioned by the stimulus the animal received, and is contingent on the environment in which the animal found itself. For that reason, "all conditioned behavior, whether relational or nonrelational, is behavior that has derived proper functions. Such behavior will have derived proper function only if the specific mechanism or program responsible for it has a proper function to produce it and, by doing so, to perform some function(s) beyond. Whence came the proper function of this *program*? Obviously it has no *direct* proper function. The program is not a reproduction, and it was not hard-wired into the organism at birth. So I must show how such a program is an adapted device, to create it having been an adapted proper function of the mechanisms that made *it*…" (Millikan 1984, 46). Due to the variety of expressions possible, complex behaviors are generally derived (that is, produced by some kind of program on an *ad hoc* basis) rather than direct (that is, built-in).

Yet derived functions are not limited to complex behaviors, and are not always exclusive of coinciding direct proper functions. It is possible, indeed common, for a device to have *both* direct and derived proper functions simultaneously – in fact, both may apply to the same function of the same device at the same time. As Millikan notes, some "proper functions are direct, some are derived, and some are both direct and derived" (Millikan 1984, 48). This is apparent when we consider tools: "Tools that have been reproduced (as have traditional carpenter hand tools) because of their success in serving certain functions have these functions as direct proper functions. But all tools have as derived proper functions the functions that their designers intended for them. When a tool has both these sources of proper function, they usually coincide. But the proper function that derives from the intention in design is always there in the case of tools. Tools simply *as such* have only *derived* proper functions." (Millikan 1984, 49). This can be made somewhat clearer by applying it to an example. Screwdrivers are reproduced because of their utility for driving screws. Driving screws is their *direct* proper function. Yet the only reason that driving screws is a utility at all is that human beings *desire* to drive them to attain some further end. Screwdrivers were intentionally produced by humans (as Wright pointed out) only as a result of this need. Insofar as they are the result of this higher-order desire, screwdrivers are *also* derived products – derived from intentions independent of the tool itself. Thus, calling a function derived or direct does not tell us much about the nature of the function, much less what the function

actually is. Rather, it has to do with the path of inheritance by virtue of which it has its proper function – whether its function is 'owned' independently, or by proxy.

Such examples can also be drawn from biology. Genes have as their direct proper function the production and regulation of entire organisms; the organs and systems which the genes code for have their *own* direct proper functions – for example, the blood-pumping function of the heart. However, the heart also expresses the *derived* proper function of the genes that coded for it. In other words, genes have as their function the production of entities which have their own direct functions. Thus, the schema is generally applicable, and integral to the theory of proper function.

**The Connection Between Reproductive Families and the Taxonomy of Function**

The two major innovations described above – reproductive families, and the taxonomy of function, are closely related. Millikan requires this elaborate taxonomy due to her reliance on reproductively-established families in attributing function. There is little obvious (though some hidden, as we will see) difficulty in saying that the heart as a type was selected to pump blood – and that, as a result, any heart token (defined as that which arises due to processes similar to the other tokens in the family) has blood-pumping as its direct proper function. Yet instances of more plastic behaviors, and certainly human intentions and language, may be unprecedented or innovative in ways that make them unique, or nearly so. Any given hunting behavior intuitively has the function of obtaining food for the animal, satisfying a basic and universal need. But the

contingency of such behavior, its particular sequence of actions, makes it more challenging to say that it is part of a reproductively established family of the same act.

Thus, the delineation of the instantiation and inheritance of proper functions is necessary to establish the *lineage* of the functional token, and thereby provide it with proper function attribution through reproductively established families. If the particular, perhaps unique, token is the expression of a more general functional system, one which *can* be readily characterized as but one member of an established family, then the function attribution scheme Millikan has devised can be applied. The taxonomy exists in order to allow the relevant reproductively established families to be identified. Although she is not explicit about this, Millikan's taxonomy of function is thus designed to allow the connections to be made that ultimately root the 'family tree' in direct proper function.

**Purpose and Normativity in Proper Functions**

As noted above, the entire purpose of Millikan's theory of Proper Function is to provide a basis for both teleological purpose and normativity in natural function (and function generally). That is, quite literally, the proper function of Millikan's theory of proper function – it was developed in order to provide a basis for attributing normatively evaluable and teleological function to natural organisms, their physiology, and their behavior, including most significantly the linguistic and intentional behaviors of human beings. By extension, the normative teleology also applies to the artifacts which human beings produce.

Having evaluative normativity means that there are functional criteria by which an item is performing its function well, poorly, or not at all. This standard is not a merely descriptive, statistical one based on prevailing conditions. Normative function means that an item can have a function that it is not in fact performing, but that it '*should* be': "For malfunctioning devices still have proper functions" (Millikan 1984, 41). This means that, unlike a descriptive norm, even entire classes of items can begin failing to perform their functions without in any way ceasing to have that function.

The only statistical or descriptive dimension to the account is the fact that proper function means that there was an historical tendency for the item to perform the function, and that this tendency is the reason why the item's family persisted or was reproduced, as described earlier.

A heart which is not pumping blood strongly or efficiently enough for the organism to thrive is simply not doing the job which its evolutionary history has 'assigned' to it by virtue of its membership in a type that has been perpetuated due to this function. Thus, "Having a proper function depends upon the history of the device, not upon its form of dispositions." (Millikan 1984, 29)

Integral with this normativity is a strong account of teleology in proper function. Devices with proper function are *for* something in the strongest teleological terms: they have a true purpose. The teleology of the device is inseparable from this evaluative normativity and vice-versa. It is because the purpose of the heart as a type is to pump blood that the purpose of the token is to do the same. In turn, it is only by having such a

purpose that there can be any standards by which the token succeeds or fails to do its job. Its norms are derived from its purpose.

I will examine the plausibility of these claims in what follows. However, before we proceed to critique the theory of proper function, there is a substantial side-issue that must be addressed.

## Proper Function as an Etiological Theory of Function

Before we begin critiquing the theory of proper function in earnest, a fundamental difficulty must be confronted: Millikan expressly claims that proper function is *not* intended as a general theory of function (or, in her terms, a *conceptual analysis* of function). Indeed, she considers such analyses illegitimate in principle: "Argument over THE correct analysis of THE concept of function either in ordinary life or in biology is, I believe, quite pointless" (Millikan 2002, 114). Instead, "'Proper function' is intended as a technical term. It is of interest because it can be used to unravel certain problems, not because it does or doesn't accord with common notions such as 'purpose' or the ordinary notion of function'." (Millikan 1984, 18). Yet if it is immaterial whether or not it or its referents comport with our intuitive ideas of function, or with other philosophical theories of function, this makes critique far more difficult, if not impossible.

Unlike Wright, Millikan did not develop her theory of Proper Function as an overarching theory of function. Rather, she did so because she "needed a term that will do a certain job" (Millikan 1984, 18). The job that she needed done was providing an account of normative function – and in particular a naturalistic account of normative

biological function. Millikan requires normative function in order to provide a grounding for the intentional content of the thought and language of humans as biological creatures. In this ultimate picture, intentionality itself comes to be seen as functional: "explicit or conscious purposes and intentions turn out to have proper functions that coincide with their explicit or conscious contents." (Millikan 1989b, 289). Since Millikan's ultimate goal is to anchor intentional content in naturalistic biological function, she requires an account of natural function that provides for both purpose and normativity. Otherwise, intentionality could not have reliable content.

Therefore, while her work on function develops and defends important implications and details of etiological function theory that Wright left relatively unexamined, she does so with a very specific agenda that Wright did not share. And one which, she claims, should insulate the functional attributions of the theory of proper function from comparisons to the many different attributions of function we may otherwise be inclined to make.

Where does this leave us? How do we evaluate an account of function which, we are told, is not to be taken as an account of function? For that matter, addressing this question forces us to confront the even more fundamental question of what it means to look for a theory of function *at all*. There are many possibilities and little obvious consensus for what a having philosophical "theory of function" really means. Perhaps a theory of function is just an account of how the term is actually applied by people in general, or even just by biologists or engineers. In that case, the best theory could be

selected based on a simple survey. But perhaps a theory of function is not just about how a word is used but about some core truth behind a shared, preexisting yet vague intuition. Then, we must dig a bit deeper; but the criteria remain descriptive of human usage. On the other hand, a theory of function might be primarily about empirical facts of the world itself. In this case, the question is essentially one of natural science. Yet it seems that it cannot be only this, for function attribution, and thus any theory of function, also seems to implicitly involve human application of a concept to the world. While to some extent both these dimensions are always implicit in scientific theories, they seem to be at stake in a deeper, more intertwined sense in theories of function – for function seems to be at least in part about how we *interpret* phenomena in the world. This confluence of the conceptual and the material is just what makes function worthy of serious philosophical attention.

What is ultimately at issue in Millikan's claim, therefore, is not just whether or not proper function can be judged in part by how it compares to other instances of function attribution, but whether or not – or in what sense – there is *any* single or best account of function to be had at all, and *how* we should judge between candidates. Thus, a diversion to explore this fundamental – and rarely addressed – issue seems necessary.

**Proper Function vs. William Harvey**

The paradigmatic counterexample to proper function is William Harvey's attribution of function to the heart. In 1628, Harvey was the first to discover what the heart was really for: pumping blood. Yet he obviously did so without knowing (or

hypothesizing) anything about its origin via natural selection, a theory over two centuries hence. If function attribution can be made accurately without any knowledge of etiology, where does this leave the theory of proper function? Was Harvey just lucky, making the right attribution for incomplete reasons? Does function attribution have to be made for the right reasons to be correct? Or was Harvey, despite appearances, doing something other than attributing function? [9]

Millikan recognizes that critics "argue against an account of biological function that presupposes evolution by natural selection on the grounds that Harvey didn't know about natural selection when he proclaimed the discovery of the heart's function, or that evolutionary theory would have to be conceptually true to play any such role in the definition of function." Yet, as we have seen, she goes on to claim that "Such criticisms are valid only if the project is analysis of the concept of function. … My purpose, my program, is an entirely different one from that of conceptual analysis." (Millikan 1989b, 290). What the theory of proper function is, she insists, is a *theoretical definition* of function. What is meant by a theoretical definition? The only *independent* answer Millikan offers is to note that, "A theoretical definition is the sort the scientist gives you

---

[9] Exemplifying the lack of careful attention paid to this aspect of Millikan's thought in the literature, while Davies does address the general issue, he notes that "Unless Harvey had in mind the theory of selected functions – and presumably he did not – then, on Millikan's view, Harvey's notion of functions was in error" (Davies 2001, 113). This is not accurate, since, as we have seen, Millikan does not insist that anyone else's concept of function is mistaken, *per se*. But she does insist that "there is no need for me to compete with" alternative accounts of function, and that the search for any one true account of function in general is misguided (Millikan 1989b, 290). As we will see, she explicitly leaves room for additional accounts.

in saying that water is HOH, that gold is the element with atomic number 79" (Millikan 1989b, 291).[10]

**Critiquing the Claims of Proper Function as a Theoretical Definition**

Paul Sheldon Davies, a proponent of the competing systemic effects account of function to be examined in the next chapter, offers what is likely the most sustained and detailed critique of Millikan's proper function specifically and the etiological account in general in (Davies 2001). Davies takes issue with Millikan's characterization of proper function as a theoretical definition, comparing it unfavorably with the respectable examples of theoretical definitions she offers. He notes that "When Lavoisier hypothesized that water is H2O, there were reasonably clear methods with which to test his hypothesis. And the methods available today are clearer still. We need only collect samples of the stuff we typically find in our rivers and lakes and examine its microstructure in the laboratory. Similar considerations apply to the hypothesis that the atomic number of gold is 79" (Davies 2001, 113). Davies notes that no such test is available to Millikan: "in what lab are we supposed to test the theory of selected functions?" Yet such tests are required because, "A theoretical definition, in this sense, is an account not of a concept but of stuff in the world and its underlying structure, and the

---

[10] Millikan does offer a more complete definition of a theoretical definition – but unfortunately it is one that presupposes the integrity and validity of the theory of proper function itself: "Now I do have a theory about what theoretical definitions are, a theory about how the theoretical definition of "theoretical definition" should go. Unfortunately this theory rests upon a theory of meaning that rests in turn on the notion "proper function", the very notion under scrutiny." (Millikan 1989b, 291). This is obviously unhelpful.

definition is true or it is false quite apart from the structure of our concept." (Davies 2001, 112). There is much to analyze here, beginning with the last claim.

Clearly, a theoretical definition is *not* independent of the overall theory to which it belongs. Thus, the plausibility of the proposed theoretical definition cannot be established without implicitly depending on some content of the theory. Davies overstates things in this regard. The examples make this clear. One cannot test water for oxygen and hydrogen composition in the lab without at least implicitly appealing to the definitions of oxygen and hydrogen embedded in the theory. Perhaps this test could be construed as theory-independent if the theory is taken to consist strictly of water being a conjunction of oxygen and hydrogen, each of which is themselves independently defined *outside* the theory. But such an approach, even if viable for H2O, is more problematic when it comes to gold being atomic number 79. This necessarily implicates a vast amount of atomic theory in its definition, in a fashion virtually inseparable from the theoretical term. Thus it would be impossible to validate the theoretical definition without invoking the theory itself.

What *is* independent in both cases, however, is the referent itself: 'gold' or 'water'. These *substances* exist independently of any theoretical definition of them – and more importantly, they were commonly referred to *before* their respective theoretical definitions were composed. We can therefore determine whether the referent of AU-79 or H2O matches that of the ordinary terms, 'gold' or 'water', respectively.

What then does it mean for them to 'match'?  Presumably, it is for each theoretical definition (of gold and of water) to be able to consistently and correctly pick out these entities from the world.  For any given candidate entity, the theory must be able to determine empirically (at least in principle) whether or not the entity fulfills the definition: presented with ethanol and water, or pyrite and gold, the theory should be able to sort them *consistently* and *correctly*.

In terms of consistency, we can imagine a scenario where the term, accompanied by a theory that proposes certain tests for the application of the term, makes consistent designations.  That is, for the very same sample, the test reliably makes the same determinations, run after run.  The same sample will not test as gold in one instance, and pyrite in another.  We may also imagine that the test is consistent under various circumstances.  The same sample will not test as gold at one temperature, for example, and pyrite at another.  Moreover, we might imagine that similar samples (perhaps defined as coming from the same source) will also test alike.  Yet the basic idea of consistency says absolutely nothing about accuracy or 'correctness'.  What tests reliably as gold may yet be pyrite.  The test could be consistently *wrong*.

What would make it *right*?  Abstractly, a correct test for gold means that gold is identified as gold – and *only* gold is identified as gold.  But what is gold?  If we define gold as just that which conforms to the test (implementation issues aside), the test can never fail.  It would be correct by definition – gold just *is* whatever passes the test.  This is logically consistent, but it cannot be all there is to the definition.

When the theory and its tests were first proposed, there had to be some external standard by which the theoretical definition *itself* passed or failed. It cannot immediately have been the case that whatever the test approved was automatically henceforth to be considered gold. Gold was known and used for centuries before atomic theory was developed. The term 'acid test' refers to a much earlier technology for distinguishing gold from other metals: nitric acid dissolves almost any material that is not gold. Such tests are still in common use today; atomic theory is essentially useless to a jeweler who wants to know whether the material he is working with is in fact gold.

Imagine that an early draft of atomic theory proposed a definition of gold as atomic number 80 rather than 79, and that some form of atomic resonance tests were developed on this basis. The confirmation process would be to obtain a sample that had already been verified as gold by the acid test, and subject that sample to the atomic resonance test. If that test rejected the sample as gold under the proposed theoretical definition, it is the definition – *not the sample* – that would need to be rejected. In other words, the theory and its test have to *explain* what gold is – not arbitrarily stipulate it, even in contradiction of existing standards.

Similar logic should apply to a theoretical definition of function. Given that function is a preexisting term with at least some uncontroversial referents, it will not do to say that the theoretical definition of it is independent of these.[11] To be sure, the lines

---

[11] To be fair, Millikan does hasten to add that her definition, while theoretical, is "not merely stipulative", and that '""has a function" does as a matter of fact correspond, in a surprising diversity of cases, to having a proper function."(Millikan 1989, 289, 291). But it is clear that this is a mere convenience, not a criterion with which to evaluate the theory.

around the phenomenon are far less clear that they are around water or gold. But the basic principle must be the same. This is also not to imply that function is (necessarily) a natural kind as are water or gold. It is simply to say that function is a concept that is in common (if occasionally fuzzy) usage outside of any particular theoretical edifice, and that this 'preexistence' must be respected by any theory which purports to give it definition, even if – perhaps *especially* if – that definition is 'theoretical'.[12]

Of course, existing standards are not inviolable, and cannot have the final word. Indeed, the point of a theoretical definition is that it will, if successful, effectively *supplant* other standards and become authoritative. Imagine if H2O was seen to refer to a large subset – but only a subset – of the samples which had all previously been referred to as 'water'. What would we do with each term – the generic word 'water' and the theoretical term 'H2O'? At the least, we would have to fundamentally reexamine our use of the word, as we could not look at water again the same way. Perhaps we would be forced to recognize that what we took to be a singular referent is really a loose class of transparent liquids of varying composition. In the case of gold, those entities that may have previously passed under the label 'gold' would likely be considered only a convincing fake. When and if a theoretical definition *supplants* the generic term, the generic term becomes merely a pointer to the theoretical definition.

---

[12] Note that this analysis does assume that the target of the terms remains roughly constant over time. This is not always the case, as the term "atom" demonstrates. What once meant something indivisible now refers to a very divisible constellation of particles. Yet such dramatic shifts in reference would seem to be apparent when they occur.

Such is certainly not the case with proper function. At least, it is not presently and has not been historically. Yet the same was true of H2O and AU-79 at certain points in time. Thus, it could still be argued that this is simply because the theory has not yet found acceptance to the degree required for such replacement.

We might then ask why this is so, and what would change it – is it the case, potentially, that proper function will someday 'succeed' as a theory to the degree that the generic term 'function' becomes merely a pointer to it? To do so, it would have to offer compelling advantages over the generic term and its looser collection of referents. In the case of water and gold, these advantages are apparent: greater precision in referring to the class of entities that seem to have been *already* intended by the term. And this seems the key to a theoretical definition: Does it capture what is already intended, but with greater precision and explanatory power?

Yet this puts us back precisely where we began: proper function cannot be a private internal theoretical account, but, as a theory of function, must attempt to account for many standard, public instances of function attribution, and do so in competition with other accounts. This is just the problem being addressed in the present work overall and it is the general object of the philosophical literature on function. Indeed, broadly speaking, this just *is* the problem of function in philosophy. For better or worse, there is no cheap way to sidestep the challenge as Millikan attempts to do.

**Novel Theoretical Definitions**

To be thorough, there is one last possible scenario for a theoretical definition: that the theoretical term is entirely novel. In that case, external verification of the referent would be difficult if not impossible without relying entirely on the corresponding theoretical content. Such may be the case, for example, with the quark in subatomic physics. This term itself was chosen more or less arbitrarily to name a wholly new hypothetical entity, one which is only ever indirectly observed. In this case, the theory and its test seem inseparable. Quarks were invoked to fill explanatory roles within the theory, not to offer an explanation of an *already* familiar phenomenon. Because the term and its referent are both novel, no extra-theoretical tests seem possible, at least until other methods of observation are developed outside the original theory.

Yet, as precedent, common usage and competing theoretical accounts all attest – such is emphatically *not* the case with function. Therefore, it seems that Millikan has no right to hide behind proper function's status as a purported theoretical definition in order to defend it from problematic comparisons. Proper function *is*, whatever else it may be, a philosophical account of function. If it were not, choosing the term "function" at all would be badly misleading. As such, it is inherently in competition with other such accounts.

**Proper Function as a Theory Of Function**

Despite Millikan's disclaimers (and, it should be noted, typically without a close analysis of this disclaimer), as the most well-known and very likely the most fully

developed example, proper function has become the paradigm of etiological function

theory in the literature (Buller 1999, 1). Relying on the foregoing argument, I will

embrace this tradition and treat proper function as the most fully-developed example of

etiological function theory. It therefore remains to analyze the viability of the proper

function as a general theory of function.

## Critique of Proper Function

### Etiology and Proper Function

Millikan's claim that her theory of proper function is simply a theoretical

definition, and as such not be compared with ordinary function attribution, has now been

considered and dismissed. The question now becomes how proper function, as the main

representative of etiological function, fares as a candidate for the best philosophical

theory of function – a theory whose job is to account for what it is to attribute function,

whether in artifacts or natural organisms. As a prime example of the conflict between

proper function and more ordinary, intuitive function attribution, we therefore return to

the case of Harvey and the function of the heart. What does it mean for proper function

as a general theory of function that Harvey was able to attribute the correct function to

the heart without knowing it origins?

### 1. Function Attribution with No Etiological Knowledge

According to the theory of proper function, the heart has the function of pumping

blood *only* because the heart as a type was retained in the species which have them due to

the fact that hearts pumps blood. Let it be granted for now that it is unproblematic to say

that the heart as a type has been selected simply for the function of pumping blood (in practice things are not quite this simple, as we will see). The question is what it means for the prospects of the theory of proper function that Harvey was able to correctly attribute that function without knowing this history.

One thing apparent in the example is that, at least in this case, a determination of function was easier to make than a determination of origin. This is clear if we consider the way in which Harvey arrived at his breakthrough insight: by studying the circulatory system empirically. At the time, the received wisdom from Galen was that the heart served to heat the blood, and its motion was simply to expel waste. Because the venous and arterial systems were believed to be independent, the heart's role as the agent of circulation was not perceived. But Harvey was able to grasp this role by experimenting on living animals, including humans. By paying close attention to how the behavior of veins and arteries correlated with the expansion and contraction of the heart itself, he was able to arrive at an accurate understanding of the heart's function, and in the process correctly hypothesize the capillary system joining veins and arteries. In short, Harvey was able to attribute function based only on immediate empirical observations of the properties and relationships of isolated tokens. These careful observations told Harvey what the heart tokens were really doing – but it seems unlikely that they could have told him where the heart as a type came from.

Given the era, it seems safe to assume that Harvey believed animals and humans to have been originally created by God, according to a divine plan for their inner

workings.  In other words, he likely held that God designed animals in much the same way that humans design artifacts, and thus that function in both could be credited to intentional design.[13]  Publication of Darwin's *On the Origin of Species* was, after all, not for another 231 years following Harvey's *De Motu Cordis*.  But what seems clear is that no empirical experiments of the kind which allowed him to ascertain the function of the heart would allow him to determine its origin as a type – whatever he held that origin to be.  Etiology, being fundamentally about what happened in the past, is not as easily enlightened by empirical experiment in the present.

Since the attribution of 'proper function' relies on having a correct account of type-origin, it too is not easily subject to empirical investigation.  This makes reliable function attribution according to the theory of proper function inherently difficult; moreover, this makes any purported proper function attribution difficult to either confirm or falsify, since sufficiently detailed evidence of origin is very often difficult to come by (as we will see in further detail below).  As Davies points out, the evidential bar for selected function is rather high: "the mere existence of complexity does not justify the attribution of selected functions to every efficacious trait within the system.  What is required is evidence concerning the actual selective history of the system; otherwise we have no grounds for claiming of any specific trait that it was selected for." (Davies 2001, 110).

---

[13] It is, of course, the loss of this assumption which creates the problem of natural function as distinct from artifact function.

At the very least, this example shows that establishing etiology, even when possible, can be difficult. This in turn makes etiological function attribution difficult. Yet this alone does not mean that it is not a viable basis for a theory of function in principle. It may be that function, when defined rigorously, is simply more difficult to attribute correctly than it has generally been taken to be. Yet the difficulty of obtaining reliable origin info should prompt us to ask: is such knowledge really necessary for function attribution? Only if it is should we consent to live with the implications.

## 2. Etiological Origin with No Function Attribution

If the theory of proper function bases function attribution in etiology, what is function apart from etiology? It seems that it must be *something* because there is, after all, a *reason* why an item-type gets reproduced, according to the theory of proper function. A heart, as a type, was selected by natural selection only because it plays an essential role in the differential rates of survival and reproduction in the animals that have them. This basic logic is at the core of etiological function attribution. But it also betrays a crucial role for something *other* than etiology in attributing function: the value or role of the function for the system. An example where this is missing will make this clear. The human appendix is reproduced as a type just as much as is the heart – every "normal" human has one. Yet our strong instinct is to *not* attribute function to the appendix just because it is a routinely reproduced trait (in Millikan's terminology, a

member of a reproductively established family).  The basis for this instinct is obvious:

unlike the heart, in present organisms, the appendix *doesn't seem to do anything useful*.[14]

Millikan is well aware of this counterexample.  Her response is that, "the

historical correlation pattern responsible for the appendix tokens in humans is quite

remote from the current appendix tokens.  Thus a direct proper function of a device may

be called 'historically proximate' or 'historically remote'; the direct proper function of

our appendix is historically remote." (Millikan 1984, 32).  Yet this merely highlights the

fact that origin alone is not sufficient to attribute function.  The appendix is still with us,

still reproduced just as much as the heart is – even if the appendix in humans today does

nothing more than become infected.  To say that its proper function is 'remote' and

therefore irrelevant to contemporary function attribution is to admit that the criteria for

current function goes beyond selection history.

It therefore seems that a theory of function based on origin may have things

backwards.  The only reason we have to go inquiring after the origin process of the heart

in the first place is that we *already* recognize – as did Harvey – that hearts indeed *have* a

---

[14] As it happens, very recent medical science (c. 2009) has begun to uncover an important role for the appendix as a reservoir of helpful bacteria which can replenish the gut in the event of an imbalance. However, because this discovery is so new, even recent philosophers like Davies, writing in 2001, continued to take the appendix to be the epitome of a useless vestigial trait.  For the sake of continuity, I will maintain this example.  Thus 'appendix' should be read as a stand-in for 'vestigial trait without current function'.  For a more current example, consider non-coding 'junk' DNA.  There is ongoing debate about how much of this DNA, comprising some 98% of the human genome, has functions.  Some likely does. However, studies in organisms with far simpler genomes have concluded that in those cases, *all* of their non-coding DNA is in fact non-functional 'junk' – strongly suggesting that at least a large fraction of 'junk' DNA in humans is genuinely useless.   Thus, whether or not the human appendix is in fact an example of a non-functional trait is not the issue here; if it is not, there are almost certainly other traits that are.  It seems unlikely that there is no single example, in any species, of a previously selected trait that has been retained but lacks current function, as Millikan implicitly acknowledges.  Even if that were somehow known to be the case, the hypothetical would still be plausible and worth exploring, though it may lose force as an argument against etiological function.

function. This seems to indicate that function attribution is *not*, in the end, fundamentally about etiology. Davies makes a similar argument, observing that, "We cannot discover the selected function of any trait without first knowing its systemic function. If we do not know the systemic function of a trait, we have no guide with which to seek historical evidence for the claims that this trait was selected for the specified functional task." (Davies 2001, 55). If this is so, then proper functions must be more generic "functions" first (for Davies, 'systemic functions'), otherwise their etiology is irrelevant.

## 3. Function Attribution with No Etiological Origin

This contention is strengthened by considering converse examples. The appendix is a case where no function is evident even though a selective origin seems to apply. There are, however, also well-known cases where function is evident and *no* selective history seems to apply – or at least, not one relevant to the currently attributed function. Such is the case, for example, with the partitions in the skulls of animals. In mammals, these partitions aid the birth of the live young by allowing the skull to elastically deform as it passes through the narrow birth canal.[15] Yet, as Darwin himself realized, these "sutures" are present even in reptiles and birds, which derive no such benefit from them because their young are hatched in eggs. Given that reptiles evolutionarily pre-date mammals, it is impossible that skull sutures – however essential they are for live births – were evolutionarily selected for that reason. Their aid to the process of live birth is

---

[15] Although I am not aware of any medical evidence to the contrary, like the appendix example, the case of skull sutures should be read broadly, as a case of a beneficial trait that could not, historically, have been selected for the benefit it now confers. As with the appendix, other examples of the general principle are no doubt possible.

merely a happy coincidence. Yet while their presence is an accident in terms of this effect, function is nonetheless attributed to them – originally, as it happens, by Paley, who took them as further evidence of divine design.

This points to what seems to be another latent paradox in proper function. Recall that proper function applies only to types. An item must be a member of a "reproductively established family" before it is eligible for proper function attribution based on the family to which it belongs having been perpetuated through performing its function. Yet when we consider how such families are established by natural selection, it is clear that the function performance, at least of a prototypical token, must predate the selection process. Natural selection cannot select for anything that does not already exist. Thus, whatever function natural selection is selecting for must *already* be performed by the item-type being selected. Hearts, for example, must already pump blood in order to be selected for doing so. And yet, according to proper function, function attribution can only be made *after* selection occurs. This introduces two problems.

First, why is the initial item to achieve the relevant effect not eligible for function attribution, given that it is the prototype which all descendants model? Here too, Millikan does recognize the difficulty. She explains, "*F* was not of course the function of the very first *T* tokens selected, even though they did *F* and that is why they were selected. They were not selected "because they had functions" but because they did *F*." (Millikan 1989a, 174). In other words, her response to the problem is simply to differentiate between the *effect* for which a thing is selected for, and the *function* that

effect has once it has actually been selected. Yet given that the "effect" and the "function" are the very same thing, this seems to be merely a semantic distinction, one which simply insists that functions are selected. It remains unclear whether this is really what function attribution is centrally about.

Moreover, this response does not address the second problem, which is: just how many generations must be reproduced before its "selection" becomes official and the trait becomes eligible for proper function attribution for doing what it has, as a type, always done? Is one round of reproduction enough? Two? Or must there be a more established lineage? No principled answer to this question seems possible from within the theory of proper function – and yet without one we cannot make proper function attributions.

The case of apparent function with no corresponding selection etiology assumes its most extreme form in the thought experiment of 'instant organisms' – hypothetical beings which, like Davidson's original swamp-man, simply pop into existence (for example, through "quantum fluctuations"). Obviously, instant entities would not have selective histories with which to attribute proper functions to their parts, no matter how similar those parts are to those of identical organisms that arose through natural selection. For her part, Millikan frankly admits that this is the case and embraces the consequence, however unintuitive. Denoting "proper function" by the equivalent (for Millikan) "purpose", she notes that "even though many people would be prone to say it did have a purpose, that apparent heart in your double's body really would not have a purpose. It would merely display enough marks of purposiveness to fool even very sophisticated

people." (Millikan 1989b, 292-3). Yet this is a deeply troubling conclusion. Consider that, by hypothesis, an instant organism is indistinguishable from one of natural origins (Davidson 1987). This means that if we knew such a thing existed but did not know which individual it was, we could not identify it – there is nothing physically distinctive about it. And yet this one token alone, out of all the seemingly comparable tokens, lacks proper function or purposes internally – and we could never know.

Reflecting the lessons of the last several sections, even Millikan admits to the intuition that, "Surely there is a paradox in the claim that to be entirely certain about the function of any trait one would have to be entirely certain about its evolutionary history." (Millikan 1989a, 175). How then does she address this apparent paradox? Surprisingly, it is by consigning the things that proper function leaves out to an entirely different account of function.

**Proper function and systemic function.** "What has been left out, I believe, is acknowledgement of another kind of function, or another sense of "function,"" (Millikan 1989a, 175). In essence, Millikan is admitting that proper function covers only a *subset* of routine function attributions. Her solution is essentially to divide function attribution into two kinds: proper function and non-proper function. To account for the remainder – the many times when we feel inclined to attribute a function that the theory of proper function is unable to account for, we must appeal to a *different* theory of function. The additional sense she has in mind is the systemic account of function, which is the focus of the next chapter. For now, it will suffice to understand that this alternate species of

function is a fundamentally different one, lacking any teleological or normative connotations. Systemic function, in a nutshell, is simply a reference to the effect of the item in the system. But the type of function being invoked is less important for the moment than the fact that there is another kind being invoked at all.

Once again, the theoretical dimension of proper function becomes a factor here, because Millikan's strategy relies on the potential for the cooperative coexistence of two very different theories of function, rather than their competition for the most adequate general theory of function. It is as if, to return to her examples of theoretical definitions, what was previously commonly referred to only as gold has been redefined into two different classes of metals, each with their own distinct theoretical definition.[16] We now need *two* theories of function to account for what seemed, on the surface, to be one kind of function attribution. The question in such cases is: does this bifurcation reflect fundamental facts about reality, or a shortcoming in the candidate theory? Are there really two distinct kinds of function? Or is proper function too restrictive as a theory, such that it fails to account for many legitimate examples of function, necessitating appeal to an entirely separate account as a backstop?

To answer this, we must further investigate both the remaining issues with proper function (i.e., those apart from its failure to reflect many typical attributions of function),

---

[16] It should be emphasized that this analogy is to the different *concepts* of function, and the function they attribute in any given case. Depending on their etiology, some items will not have proper functions, but will have systemic functions (e.g., skull sutures). Yet it does not follow that each theory of function would necessarily attribute different functions to the same item. For example, both proper function and systemic function attribute the same function of pumping blood to the heart. What is different for each is the grounds for function attribution and in turn the nature and meaning of that function – not necessarily *what* function gets attributed.

and also the nature of the alternative theories of function. I will do the former in the remainder of this chapter, while the close examination of competing theories will be the focus of the subsequent chapters.

**Proper function as an account of malfunction and purpose.** In comparison to the alternative systemic theory of function Millikan refers to, it must first be appreciated what makes the theory of proper function unique and, purportedly, valuable. Between the two, only proper function provides an account of the teleology and normativity of the function. Proper function is *purposeful*, and if something with proper function fails in the performance of its function, we have a basis on which to say that it is *malfunctioning*. If a heart, for example, is malformed or diseased, and unable to pump blood effectively, we can say, under an account of proper function, that it is malfunctioning – that it is not doing the job that a heart is 'supposed' to do.

By contrast, those beneficial biological traits that are not *known* to have been competitively selected for do not enjoy the right kind of etiology to qualify as proper function. On Millikan's account, they then default to mere systemic functions. If there are useful traits "that were not in fact ever selected for, then these will have [systemic] functions but not selected functions." (Millikan 1989a, 175). Because they lack proper function, no account of normativity or teleology is available for these traits. Because the trait was not selected for, there is no basis to say what the trait is 'there for'. Thus, if a fetus for some reason did not develop the skull sutures useful for vaginal birth, there would be no basis to say that there is any defect, or that the skull is not as it 'should be'.

This is so even though we may continue to refer to the skull sutures as having a systemic function when they *are* present. Proper function, then, is a very special variety of function – the only one, purportedly, that includes an account of teleology and normativity. Systemic function, it seems clear, is only function nominally for Millikan. It exists as a label that gives us leave to continue to refer to commonly attributed functions left out of proper function *as* functions. Yet the primary notion of function for Millikan remains proper function.

An account of purpose and norms, particularly one that applies to natural systems as well as artifacts, has great appeal as it corresponds well with our basic intuitions of function. This is likely the reason why etiological theories are regarded as the consensus account in contemporary analytic philosophy (Buller 1999). But this is a good approach only under three conditions:

1. That normativity and teleology are in fact essential for an adequate theory of function, including natural function.
2. That no other, more amenable theory accounts as well for teleology and normativity.
3. And that proper function really is able to provide an account of teleology and normativity consistently and coherently.

The first criterion will be addressed in the next chapter; the second in the chapter after that. The third divides into two main considerations which I will address presently, in order: First, does proper function comport with the theory of natural selection on which it relies, or does it demand things that natural selection simply cannot provide?

Secondly, are the teleology and normativity which proper function promises to provide even for natural functions philosophically and naturalistically coherent? Or are the natural teleology and normativity it promises illusory, inherently unsupportable naturalistically?

## 4. Proper Function and Evolutionary Theory

Proper function explicitly appeals to the theory of evolution by natural selection: "I do need to assume the truth of evolutionary theory in order to show that quite mundane functional items such as screwdrivers and kidneys are indeed items with proper functions." (Millikan 1989b, 290). This means that proper function relies on natural selection to supply it with certain elements. Is natural selection, according to orthodox evolutionary theory and modern biological understanding, able to supply proper function with these elements? If the theory has problems connecting to the theory of natural selection without conflict, this debilitates it both as a theoretical definition, and as a theory of function.

**The problems of reliable knowledge of selection.** Recall that in order to attribute proper function to a trait, we must know the trait to have been selected for that function. We cannot conclude from the mere fact that a trait is present and useful that it has been selected for, as the case of the skull sutures reminds us. Indeed, to have been 'selected for' carries specific historical implications: the trait must have had competitors, and the variety of the trait to which we are attributing proper function must have allowed the organism with this trait to outcompete those with the inferior variations in terms of

differential survival and reproduction rates.  How do we know that this is the case for many of the routine biological function attributions we are inclined to make?  How do we know, with any degree of certainty, that a trait even had competitors?  As Millikan says, "Suppose that the genes responsible for some highly useful trait, like having eyes, freakishly just never happened to have any competitors in the gene pool. Could it be that no matter how useful eyes were, useful over no matter how many generations, eyes would not then have a biological function?" (Millikan 1989a, 175).  We have already seen what happens to traits which cannot be shown to have been selected for: they are relegated to merely nominal systemic functions.

The question is thus pressing: how can we ever *know* for certain that any given trait was in fact selected for?  Because it is about particular past events as opposed to current states or universal laws, this is a somewhat more intractable version of the problem that applies to all empirical science.  As we saw, it was easier for Harvey to demonstrate empirically what the heart actually did than it would have been for him to demonstrate where the heart came from.  The confidence with which we can attribute proper function cannot exceed the confidence we can have in saying that the specific trait in question was competitively selected for.  This limitation becomes worrisome when we turn to consider in detail what it would really mean for a functional trait, like a beating heart or good running speed, to have been 'selected for'.

**The problem of attributing 'basic' function via fine-grained natural selection.**

Consider again our stock example as interpreted by the theory of proper function: the

proper function of the heart is to pump blood because this is why it was selected as a trait.

But what does the claim that the heart was 'selected for pumping blood' really mean

when we unpack it?  First, it implies that the heart was a randomly generated variation,

spontaneously produced through genetic mutations.  Second, it implies that there were

competing organisms that did *not* have hearts, because selection of any trait requires the

presence of variations lacking the selected trait.  Third, it implies that because hearts

served a useful role in the organism, they allowed the organisms that had them to survive

and reproduce at rates differentially superior to those organisms that lacked hearts.  All

these points follow directly from the definition of selection as embedded in the theory of

evolution by natural selection, and the claim that the heart was selected because it pumps

blood.

Yet, as Larry Cummins points out, when spelled out so explicitly, it is clear that

this entire scenario is biologically implausible, to say the least.  Hearts as we know them,

or anything recognizably similar, did not simply pop into existence fully-formed through

a single genetic mutation, like the instant organisms imagined earlier.  Like any complex

trait, they are the result of eons of incremental refinement through natural selection,

which operates through gradual alteration, not wholesale invention.[17]  Thus, functional

---

[17] Note that this is not necessarily the case with all functional traits – i.e., some may in fact be generated by a single mutation (or small coincident set).  Wing color in moths is such an example; once the wing structure is in place, variations in their coloration may arise spontaneously in a single mutation event, whereupon they may be selected for.  Thus, the wings' function as camouflage is not subject to this

traits and organs are the result of thousands of fine-grained selections for *performance*. This enhanced performance is adaptive, to be sure – but such adaptions do not constitute basic function: "Adaptiveness is a matter of degree; having a function is not" (Cummins 2002, 167). In other words, evolutionarily recent selection may account for the fact that the heart pumps blood *as well* as it does, or *the way* that it does – but not the fact that it pumps blood *at all*. But etiological theories like proper function do not merely attribute minor functional refinements or adaptions to selected traits – they assign full-fledged "basic" functions, such as blood pumping.

Why is this so problematic? It would seem that the period of selection simply needs to be adjusted to include initial selection for basic function. There are in fact two profound problems with this response. The first is that the trait or organ first selected for basic function was likely nothing like the modern examples. Because the traits we attribute function to were generally not the products of single selection events but the gradually accumulated result of thousands or millions of such events over millennia of evolution, traits which first had the functions now attributed were very likely nothing like those we see today. The first hearts, for example, "were not even [in] vertebrates, and the structures in question were nothing like our hearts" (Cummins 2002, 164-165). Because the refinements would have been so gradual and so piecemeal, it is quite difficult to

---

critique. However, the wings' function as enabling flight is very likely not as sudden a development but rather a series of refinements. Thus, some traits may gain their function in a single 'leap', and then it can fairly be said that their function was 'selected for'. However, many more functional traits are complex enough that they have arisen only gradually, and their selection histories are not so direct. This is not to imply that we can necessarily tell the difference – only that many traits cannot have had as simple a selection history as the theory of proper function implies and requires.

imagine what would constitute the very *first* heart.  We can only speculate, or at best appeal to the fossil record.  Even then, it will likely be a judgment call as to when the sophistication of the trait is sufficient to qualify it as the first 'heart'.  Thus, even if that organ could be said to have been selected for its function, if it is profoundly morphologically different, on what grounds can it be said to be of the same type (for Millikan, 'reproductively established family') as a modern heart?  If it is not of the same type, it cannot share function by virtue of membership in the type.  One can see this, for example, by considering that the norms of performance of a heart from a different species many millions of years ago would not apply to the heart of modern humans – this is just what it means to have undergone further selection for performance.  Therefore, we know that they cannot be members of the same family for Millikan, since family membership is what determines the performance norms that apply.

Once an organ that pumped blood and resembled modern hearts enough to merit the name evolved, subsequent selection events would have been for refinements of efficacy or efficiency.  This brings us to the second, even more serious problem: the selection events that occurred following selection for basic function were not only just for adaptive refinements – they were actually *against* competing variations which *also* performed the same basic function.  At this point it becomes not merely imprecise but inaccurate to say that function is being selected.  The modern types "were selected because they were better at performing some function that was also performed by the competition" (Cummins 2002, 165).  But these selection events were by then amongst

organs that had the same essential function, and varied only in performance. "Selection requires variation, and there was no variation in function in the structures in question, only variation in how *well* their functions were performed." (Cummins 2002, 165, emphasis added). This pattern is of course not particular to hearts, but is nearly universal. "*Having* a function is not what drives selection, but rather functioning *better* than the competition. What the function of a wing is should be distinguished from how well it performs it." (Cummins 2002, 167, emphasis added).

An artifact example involving deliberate, human selection rather than natural selection may serve to make this point clearer. Suppose that I am choosing between two different model cars at a dealership. One is a sports car with a 300 HP engine, and the other is an economy car with a 150 HP engine. I choose the sports car. When asked to justify that choice, I respond that I selected the sports car because it drives (i.e., it is self-propelled). The basic function of a car is indeed to drive. Yet this explanation is nonsense. *Both* cars drive. Therefore I cannot justify the choice *between these two cars* on that basis. I chose the sports car because it drives *faster*. If, on the other hand, I were choosing between a sports car and a horse-and-buggy, then I could legitimately say that I chose the sports car because it drives. But this scenario is implausible, because historically sports cars did not compete against horse-and-buggies. Sports cars are evolutionarily much later. Horses competed against Model A Fords, for example. So, under normal circumstances, to say that the buyers of a 2013 Mustang choose it because it drives is inaccurate. Its buyers choose it because it drives (let us assume) faster than its

competition; but its competition also drive.  Even if we assume that the Mustang is directly descended from the Model A, and even if the Model A was in fact chosen because *it* drove, this does not mean that the Mustang's later selection was simply because it drove.  By then, driving was a given; Kias drive too.  They just don't drive as fast.  In the same way, it is problematic to claim that modern hearts were selected for because they pump blood; they were selected for because they pump blood more effectively or efficiently than their competition.

One might still try to rescue the notion of selection-for-function by insisting that the relevant selection events include not just the final or most recent rounds, but earlier rounds as well – indeed *all* selection events that led to the present form.  Yet the problem with this response is obvious: the regress does not end until we reach the beginnings of life.  As we have seen, if we try to limit this regress by stopping it at the first version of the trait that was selected for the performance of the modern function, the trait in question will almost certainly be very different than the modern examples.  Moreover, we still cannot say that the modern examples were selected for their distinctive function, because the variations they were selected *against* had the same basic function.

One last response to this problem is possible for etiological proponents: rather than trying to fix the problem on the selection end, adjustments could be made on the function end.  In other words, function attribution itself could expand to include fine-grained performance characterization.  Thus, the function of the heart would no longer be simply to 'pump blood', but to 'pump blood at a certain volume, pressure, and rate, etc.'.

This way, the distinctive selection history of the actual modern type is reflected in its function. One issue with this is obvious: it makes function attribution overly precise and thus terribly unwieldy. It also departs substantially from the ways in which function is actually attributed in ordinary use. But these issues are not the biggest problem, which is that this approach merely shifts all the epistemological issues with selection history onto function itself. Precisely *what* functional attributes was the heart actually selected for (note the "etc." above)? It will do no good to characterize normal performance on a statistical basis for current examples, because we have no assurance that every trait was actually selected for. Which ones were? As Davies notes, "we often cannot determine which traits are being affected directly by selection" (Davies 2001, 56). This is the recurring problem with the etiological approach, and shifting the burden onto function characterization in order to dodge the problems with fine-grained selection histories does nothing to alleviate it.

Even if the problems with selection epistemology were strictly practical, with no theoretical difficulties, they would still present real obstacles to actually attributing function via the theory of proper function. But while that might be daunting enough, the problems in fact run deeper than this. In the final analysis, "the targets of functional characterizations and the targets of selection just do not match." (Cummins 2002, 169). Thus, according to evolutionary theory, modern hearts *cannot* have been selected for the simple function of pumping blood, even though they descended from some organ that perhaps was.

**'Basic' function and contributing traits.**  The more the biological details of selected function attributions are examined, the worse the picture for selected functions becomes.  In addition to the issue of natural selection for function, there is the problem of what constitutes a selected trait in the first place.  The problem here is how we can ever know, or show, "for any type of entity that can evolve via selection, which among its various traits are causally responsible for its selective trajectory" (Davies 2001, 123).

Many high-level traits to which we might want to attribute function are the result of the interaction of a multitude of lower-level traits.  For instance, Davies offers the example that running speed "is the orchestrated effect of several lower-level phenotypes, including lung capacity, muscle strength, blood chemistry, etc."  All of these factor into the top-level trait 'running speed'.  If that speed aids survival and reproduction, which enabling trait was thereby selected for?  And thus to which can the theory of proper function attribute function?  Not only would it be very hard to tell, epistemically, but the question may have no clear answer in principle, as it is their confluence that resulted in success.  It is at least possible that on occasion "some phenotypes screen off others" such that, "these other phenotypes, in consequence, are barred from possession of selected function." (Davies 2001, 126).

The corollary is that, in general, if organisms are the level at which selection occurs, there is little assurance that an organism's traits have *all* been selected for.  The organism may prosper because of some traits, and in spite of others.  It would be circular

to claim that *all* beneficial traits are selected for. How then can we ever hope to untangle the web of interactions which led one species with one set of collective traits to prosper and survive, especially given that whatever occurred likely happened millions of years ago? Yet etiological theories of function such as proper function demand that we do so.

## 5. The Questionable Metaphysical Naturalism of Teleology and Normativity in Proper Function

As noted earlier, the main draw of the theory of proper function – indeed, the reason why Millikan developed it in the first place – is not simply that it is a way to attribute functions systematically, but that it is a way to imbue the functions attributed with both purpose and norms of performance. Teleology and normativity are thus literally the proper functions of the theory of proper function; they are its *raison d'etre*.

Let us ignore for now the problems explored above with natural selection, epistemology or origin, and correspondence with ordinary function attributions, and imagine that proper function is entirely viable on all these counts. Can it really then provide us with a naturalistic account of purpose and norms, even for natural functions untouched by conscious intent? If it can, there may still be incentive to try to salvage the idea. If it cannot, it would seem time to move on in our search for an optimal theory of function.

To begin with, we must ask some profound questions: What would it really mean for an organ or trait to have a *purpose* without any intention to give that content? What

would it mean to have norms apply to its performance?  Is selection for function an adequate base on which to establish these properties?  What kind of properties are these anyway?  And can they be natural?

We know that for Millikan, in Davies' words, "Nothing about the current intrinsic physical properties of the token is relevant; the right sort of historical relation is everything." (Davies 2001, 137).  Given that historical relation, which is essentially membership in a reproductively established family, Millikan's claim is that the token has proper function.  By that virtue, it thus has a purpose, and norms that apply in relation to this purpose, such that it is malfunctioning when it fails to function.  I will examine the logical structure of malfunction claims in the next section.  But the ontological and metaphysical aspects of these claims are puzzling when examined closely.  In both areas, I will rely heavily on condensed forms of arguments devised by Davies, with additional considerations offered.

Davies' primary question is, "What kind of a property are we attributing to a token when we attribute a selected function?" (Davies 2001, 137).  Obviously, actual function performance is not required of selected function tokens, or there could be no such thing as selected malfunctions – the existence of which is the very point of the theory.  This means that there are functional obligations imposed upon a token solely by virtue of its heritage as a selected type.  Millikan must "hold that the emergence of selected functions involves the emergence of a functional office or role, including a norm of performance that applies to tokens of the functional type, a norm that remains even

when the requisite capacity is lost." (Davies 2001, 141).  But what could the naturalistic basis for such an immaterial 'office' be?

Davies contends that when we look carefully, the connection between selection and teleology seems more like an assertion, without much argument to back it up. "We require an account of the natural mechanisms or the natural causes that give rise to the functional office and roles. This part of the theory is asserted but never defended." (Davies 2001, 139).  There is no apparent natural basis on which to claim that a thing is 'supposed to' do something that it *physically cannot do*, as in the case of natural malfunctions.  No causal properties seem adequate to produce such norms.  "It is hard to see how differential reproduction resulting from variation among traits makes it the case that descendent tokens are "for" the performance of some task." (Davies 2001, 139). "What natural features of the causal-mechanical processes that constitute a selective history have the power to determine that descendent tokens are for the performance of some task?" (Davies 2001, 139).

On a strict naturalist ontology, then, it seems that there can be no such things as natural malfunctions.  Or, if there are such teleological obligations, the ontological commitments involved seem to violate or transcend naturalism.  On this reading, we could hypothesize that there are functional offices or roles somehow created by selection history.  But this involves abstract and non-causal properties without apparent basis or necessity.

If functional purpose and norms were required for explanations in biological science, we might be motivated to try to salvage them somehow. But Davies claims that biological theory in no way requires such properties; causal explanations suffice for biological science, just as they do in the rest of science. Moreover, the existence of such offices seems impervious to confirmation or falsification. Thus they are both "*explanatorily extraneous*" and "*beyond empirical investigation*", making them scientifically superfluous (Davies 2001, 146, 147). In addition, the precedents for such properties, such as purpose via divine design, are largely theological, and as such are not encouraging. "I conclude that the theory of selected functions, taken as a whole, is insufficiently naturalistic" (Davies 2001, 148).

More generally, Davies is led to assert that "We should reject the claim – a claim central to the theory of selected functions – that an explanation of why a trait exists is equivalent to an account of what a trait is for. There are good accounts of why a trait has persisted, but there are no good accounts of what natural traits are for." (Davies 2001, 149). Of course, this is the founding logic of etiological function theory in general: it is presumed, largely on intuition, that to explain what something is *for* is to explain *why it is there*, or how it got there in an etiological sense.[18] Wright was explicit about this, as we saw earlier.

But this logic does presume that it is coherent to postulate that there could be something the thing is *for*. In the case of natural functions, this presumption is far from

---

[18] There are of course other senses of 'why it is there' besides the etiological, some of which we will consider later in the work.

obviously valid.  In that case, showing why a thing is there (i.e., how it got there as a type) would not demonstrate its purpose.  In an artifact, it is logical to assume that if the token item did not have functional purposes, then it would not have been produced.  We assume that a device on an engine is there for a reason – that it is *for* something, placed there *in order to* achieve some desired effect.  However, this 'for' ultimately refers to and depends on the conscious intention behind the production.  Wright wants to abstract from this case and apply the same logic to natural functions.  But it is not clear that the analogy holds.  In abstracting from intention, Wright may be dispensing with the only real source of meaning in the concept of *purpose*.

If intention is the ultimate source of the appeal to origin in artifacts, then it is essential, and the logic does not clearly transfer to the natural realm.  It may be true that if hearts as a modern type did not serve to pump the blood then they would not have been selected for as a type.  But does this mean that this is what hearts are metaphysically 'for'?  With no intention to appeal to, does this mean that if a heart token does not pump blood that it is not doing as it 'should', or what it is 'for' in some genuine yet natural sense (in other words, in a sense not reliant on any human contribution)?  How can it?  What would that mean?  I would suggest that it may be more likely that it is we humans who make that inference, inaccurately, based on our experience of design.  As we will see, Cummins makes a similar argument (Cummins 1975).  For his part, Davies offers an alternate explanation based on epistemically-informed human expectations of function for selected-type tokens.  He notes that "selective success may prompt us to *expect*

descendent tokens to carry on in the manner of their successful ancestors, but that is a fact (if it is a fact) about the effects of selection on our psychology." (Davies 2001, 140).

[19] In any case, natural teleology – one independent of any human (or other conscious) contribution – seems a very curious concept. Thus, despite its appeal to the scientific theory of natural selection, Davies concludes that "the naturalistic standing of selected functions is illusory" (Davies 2001, 156).

This discussion is still preliminary, and these considerations will receive a more thorough hearing in the chapters to follow. For now, we note the exceedingly strange metaphysical implications attached to claims of purpose and norms as applied to natural objects such as hearts – namely, the immaterial, untestable, binding 'office' which the object is obligated to perform, even when it cannot. Moreover, in addition to these crucial metaphysical considerations, it is not clear whether proper function is able to offer a *logically* consistent account of natural malfunction. We turn now to this issue.

## 6. The Logical Consistency of Malfunction Claims in Proper Function

The promise to account for natural malfunctions is one of the biggest attractions of the theory of proper function. One argument against this claim is offered above: it seems to require very strange ontological categories (namely, a functional office that applies even when that function performance is impossible). But there are additional problems worth exploring with the basic logical consistency of malfunction claims within the theory, even apart from these metaphysical problems.

---

[19] I will return to the relationship between expectation and malfunction claims in a later chapter.

Logically, malfunction claims rely on the definition of generic types. Any theory of malfunction incurs the obligation of defining the generic type of the things which have function.[20] Within that generic type, there are at least two subsets: those that function and those that fail to function. For example, hearts must be defined as a generic type that includes both functional and non-functional examples. They cannot simply be defined as that organ which successfully pumps blood, for otherwise tokens which fail to do so could no longer be considered hearts at all. In other words, theories of function which seek to explain malfunction must supply a type standard *independent* of successful function. A heart must be a heart by virtue of something other than what it actually succeeds in doing in order for there to be the logical possibility of a malfunctioning heart. Etiological theories of function struggle with this requirement.

The difficulty stems from the fact that selected functions are defined by those tokens that functioned successfully. Those that failed to function (i.e., failed to produce the relevant effect), were by definition *not* selected for. So how can it be that they are a member of the selected type at all? If successful tokens define the selected type, then no tokens which are unsuccessful are members of the selected type. Therefore, Davies argues, there can be no such thing as selected malfunction, as selected functions lose membership in the relevant class as soon as they fail to function: "membership in a selected functional category … requires possession of the property selected for." (Davies 2001, 200).

---

[20] This is true in principle even if there exists only a single token item, for accusations of malfunction would still implicitly appeal to an idealized functional version of the item that exemplifies its 'type'.

In other words, "generic types and selected functional types are not coextensive. … Selected functional types, in a word, are success types. Generic types, by contrast, are not." (Davies 2001, 199). Therefore, "the category 'heart', which contains the efficacious and the impotent, cannot be a selected functional type." (Davies 2001, 203). Given that the theory of proper function bases function on selection, its claim to account for malfunction on this same basis is questionable.

Millikan has a potential defense against this argument, as the argument above does oversimplify Millikan's account in one important respect. As noted earlier, she defines family membership not directly according to *properties*, as Davies alleges she must, but according to *processes*. As noted above in the discussion of ordered reproductive families, on Millikan's account an item is the same sort of thing as another item if it came about by the same general sorts of processes as the first. This is why a malfunctioning heart is still a heart, even when it cannot function as one: it has an origin history very similar to the selected type. It is worth quoting the relevant clause in full here that was only excerpted earlier:

> "If anything *x* (a) has been produced by a device a direct proper function of which is to produce a member or members of a higher-order reproductively established family R, and (b) is in some respects like Normal members because (c) it has been produced in accordance with an explanation that approximates in some (undefined) degree to a Normal explanation for production of members of R, then *x* is a member of R." (Millikan 1984, 25).

Her next sentence leave no doubt that the purpose of this arrangement is to address the problem of functional classification: "The vagueness of the question, in some cases,

whether a bit of matter should be called 'a malformed eye' or merely 'a glob of

misplaced matter on the forehead' is reflected in the vagueness of 'in some respects like'

and 'approximates in some degree to a Normal explanation.'" (Millikan 1984, 25). By

relying on common or analogous productive causes, Millikan attempts to establish a

generic type logically independent of functional success, even though functional success

is what constituted its selection.

Let us consider the problem more carefully in light of this logic. Rather than

appealing directly to functional properties to define the type, Millikan looks to the

processes that bring about the type. The implicit claim is that as a result, this type can be

considered a *generic* type rather than a *functional* type. In the biological case (which is

our main concern since artifacts can always resort to design intentions to cash out

malfunction), genes are the primary mechanism of inheritance – the means by which

successful traits are propagated. Indeed, this is built into the theory of natural selection,

as evidenced by the fact that Darwin had to hypothesize a mechanism of inheritance even

before Mendel demonstrated it. In that case, in Millikan's terms, the genes themselves

have first-order direct proper functions, while the hearts they code for are higher-order

products.[21] When Millikan refers to trait $x$ as a product of something which has as its

direct proper function the production of something like $x$, we can imagine a set of genes

coding for, for example, a heart. Her argument then is that as long as a thing is produced

---

[21] Indeed, Millikan is explicit about this: "Hearts and livers are not members of first-order reproductively
established families because one heart or liver is not a direct reproduction of another. But although my
heart is not a copy of my parents hearts, it was produced under Normal conditions in accordance with the
proper functions of certain of my genes which were directly copied" (Millikan 1984, 24-25).

by these mechanisms, that thing is a member of the relevant generic type. Or, in our example, as long as the organ is the product of the relevant set of cardiac genes, that organ is a heart. This is convenient, because whether or not that organ is intact and fully functional, it may still be considered a heart, whose 'job' it is to pump blood, even when it fails to do so. Malfunction is possible, because it is lineage that determines its functional office, not its ability to fulfill that office.

But then we must ask: what are the relevant genes, and what would it mean for a non-functional item to be produced by them? We may presume that errors in these genes would produce hearts with anatomical anomalies leading to malfunction (as, for example, in trisomy 18). But what constitutes 'errors' in the genes? What would make them the 'same' genes as were selected for (or, for that matter, as those they were immediately copied from) if there are in fact discrepancies in them? Here we face precisely the same problem of generic types and functional types as we did before – only now the problem has been pushed back several levels to the biological mechanisms responsible for the production of the functional organs. The problem has not been solved – only buried. In terms of basic logical structure, it remains unresolved. If we look to the productive processes for genes to escape the problem once again, we risk complete regress.

It is also important to consider the issue in light of the precise terms which Millikan lays out for function selection, as described in detail above. Of note is the fact that function selection is characterized as a prevailing historical *correlation* – not a strict unfailing causation: "[selected] function should be thought of not as an invariant function

or as an average function but as a function performed in a critical mass of cases of actual

use" (Millikan 1984, 32). To be selected for function, the presence of the functional type

must only *tend* to be associated with the performance of the function more often than the

alternatives. Could this provide enough flexibility to escape the charges of conflating

generic types for success types by capturing both under the cover of correlation?

Yet this will not do. Millikan included the correlation characterization primarily

to account for cases of function failure due to lack of opportunity, such as plants with

chlorophyll in which no photosynthesis takes place due to a lack of light, or sperm which

fail to impregnate due to the lack of an ovum. It cannot address the problem of only

functional types being selected for. Consider that the selected functional type either

refers only to success types, or it refers to generic types, successful and unsuccessful

alike. If it refers to functional types, the problem of selection for success types only

remains. If it refers to generic types, then these types are being defined independently,

outside of the selection process. Yet this would be improper, for as the issue is what

types are actually selected for, selection *must* do the work of defining the type. To do

otherwise is to smuggle in a type definition, and then attribute the type to the selection.

That the problem remains can be seen in another way. Natural selection,

according to orthodox evolutionary theory, requires variation against which to select. We

can imagine early hearts, for example, competing with other modes of circulation. What

we must notice here is that functional heart tokens were competing for selection not

merely against radically different functional types (for example, a form of peristalsis

rather than centralized pumping) – but also against non-functional or sub-functional tokens of their *own* type.  In other words, far from belonging to the same selected type, functional hearts were selected for *against* non-functional hearts.

Millikan remains locked into a conundrum.  If the generic type 'heart' includes functional as well as non-functional hearts, then it cannot be a selected type.  On the other hand, if we appeal to the selected type, then non-functional hearts do not belong to the type.  If they do not belong to the functional type, then they do not have this function as their selected 'office', and thus they cannot malfunction.  As Davies puts it, "incapacitated traits, insofar as they do not qualify as members of the selected functional type, cannot possess the corresponding selected malfunction" (Davies 2001, 205).

## Conclusion

The examination of the etiological account of function in the form of Ruth Millikan's theory of proper function is now completed.  Millikan's proper function has been shown to be problematic in its methods and questionable in its very aims.  In general, proper function should be viewed as an attempt to provide teleological and normative content for natural function by appealing to etiology in the same fashion we do for artifacts.  But if intentions are in fact ultimately underpinning the purpose and norms of artifacts, the etiological parallel between the two may not produce the same results for natural function.  Meanwhile, tying all function to etiology leads to all manner of complications in attributing function, including conflicts with natural selection and the persistent problems surrounding securing reliable etiological knowledge.

Thus, it is not clear that proper function can accomplish its explicit goal of attributing purpose and normativity to natural functions simply on the basis of etiological heritage. For that matter, it is not entirely clear what having real natural purposes or norms might truly mean, even in principle.

A great many problems thus stem from this heroic attempt to provide function, especially natural function, with built-in teleological and normative dimensions. As discussed above, this attempt was motivated by the presumption that these dimensions are important – indeed essential – for any account of function. But are they? If the attempt to provide them produces such difficulties – methodological, metaphysical and even logical – perhaps the wiser choice for a theory of function is decline to make the attempt – indeed to bypass these aspects entirely. Why does function need to be teleological? Why should an account of function incur the need to account for malfunction? The group of theories known as systemic effects adopts this strategy and is the subject of the following chapter.

I have also considered what it means to look for a theory of function at all, and how we should choose between them. In short, a theory of function seems evaluable by many of the same criteria we use when evaluating any theory of science. This may include striking a balance between breadth of coverage of phenomena (i.e., the range of function attributions which the theory can make) and explanatory depth (i.e., the depth of what it means to attribute a function). Etiological theories such as proper function

prioritize the latter over the former, as we have seen.  Other theories, such as systemic

function, which I will examine next, do the reverse.

Therefore, I now turn to consider the systemic effects account of function, which

is in many ways the polar opposite of the etiological variety.

CHAPTER THREE

THE SYSTEMIC EFFECTS APPROACH TO FUNCTION

## Introduction

**Overview**

Is function an inherently teleological concept? Or can a workable, coherent

theory of function dispense with teleology, and thereby avoid its metaphysical

complications when applied to natural systems? We have already seen how problematic

it can be to insist on the definitional link between function and origin, as one attempt to

cash out the intuitive link between function and purpose. I now consider the inverse

alternative: function based only on observed systemic effects, without teleology, purpose

or normativity of any kind. If teleology is held to be incompatible with naturalism, this

variety of function theory, known as systemic effects theory, will be free of such

problems because it is designed to omit all teleological aspects. This strain of function

theory was developed by Robert Cummins in 1975 and later expanded, defended and

promoted by Paul Sheldon Davies in 2001.

Because systemic effects theory was developed as a reaction to the problems with

etiological function, it is important to begin with an examination of these roots. Although

critiques of etiological function were thoroughly explored in the previous chapter,

Cummins' diagnosis of the work on the philosophical problem of function as presuming a

link between origin, contribution and purpose will be instructive both in highlighting

where systemic function departs and in glimpsing how the three theories to be covered (etiological, systemic effects and contributions) might relate.

Following this, I turn to an exegesis of Cummins' and Davies' theory of systemic effects function. In a nutshell, the theory of systemic effects function holds that the function of a part is whatever effect that part has within a system that helps to produce the capacity of the system at stake in the analysis. Because systemic effects function is inherently relative to a given systemic capacity, this capacity must be identified before function can be attributed. Identifying this capacity is thus necessary for the application of the theory. This arrangement also means that systemic effects function is essentially contingent function; it is only ever the weaker "function *as*", never the more definitive "function *of*", as with proper function. Thus, there is also no such thing as malfunction in systemic effects theory.

I then proceed to a critique of the theory of systemic effects function, which, in many respects, mirrors the prospects for any non-teleological theory of function. The first issue is just what it is the theory of systemic function takes itself to explain, which is quite different than the explanatory focus of etiological function. Another issue is the awkwardness of making function attribution without the possibility of malfunction attribution, and the discontinuity this creates for the application of the theory to fields such as engineering and medicine, which require normative judgments to accompany function attributions. Looming over these issues, the same central difficulty appears in different contexts: given that systemic function attribution is relative to a systemic

capacity, how do we specify that systemic capacity? If any system or capacity will do, function attribution threatens to become meaningless. The most sophisticated response to this problem comes from Davies, in his one major refinement of Cummins. Davies insists, with good reason, that the theory only applies to hierarchically organized systems.

As all its details are considered, Davies' improvement will be seen to imply more than it first seems, and in one sense it brings systemic effects and contributions approaches closer together. Yet ultimately this provision is insufficient to prevent the proliferation of functions, and this betrays the underlying arbitrariness inherent to function attribution under systemic effects theory – and likely to any theory of function which excludes any concept of teleology.

In the end, this attempt to have function without teleology is found to produce as many problems as it solves, and a third way that retains some teleological underpinnings for function while avoiding the metaphysical complications encountered with etiological function will be sought.

## Cummins' Analysis of Etiology and Contribution in Function

### Cummins' Analysis of Etiological Function's Premises

It was a core conviction of Wright and Millikan that to attribute function is also to account for origin – that the effect of an item that is its function is just that which is implicated in explaining the item's presence. The heart produces a number of effects: among them, it pumps blood and it makes sounds. But only one of these is its function: the one which accounts for its selection as a type and thus any token's presence in an

organism. The heart was not selected for its sound effects; it was selected for its pumping effects. Etiology, it is claimed, is thus the key to properly attributing function, because etiology necessarily picks out the salient effects which are the type's function. In turn, proponents claim, this distinguishes function from accident in principle, lends function attribution explanatory power and provides a basis for teleological and normative aspects of function.

Robert Cummins initiated a broad strain of opposition to such etiological theories of function just two years after Wright's landmark paper, with the publication of his paper "Functional Analysis" (Cummins 1975). Notably, Cummins begins by reducing the logic of etiological function to two independent premises: (A) that the goal of a functional explanation is to account for the *presence* of the functional item; and (B) that for something to function is for it to *contribute* to a system. A "fusion of (A) and (B) constitutes the core of almost every recent attempt to give an account of functional analysis and explanation" (Cummins 1975, 741). Yet, Cummins claims, neither premise is ever actually defended – and both are in error. Moreover, he claims that the two are inappropriately conjoined, as etiological explanation (A) cannot be derived from systemic contribution (B).

Although the previous chapter has already considered and largely dismissed etiological function theories involving both premises, the status of each premise independently is a different matter. Indeed, on its own each remains pertinent to the overall project. (A), linking origin with purpose, is a core intuition of teleology in

general. (B), linking purpose with contribution, will inform an alternative approach to function to be considered in the next chapter. It is therefore important to examine Cummins' critique of each.

First Cummins presents an argument against deducing (A) from (B): simply that a thing is making a contribution to a system does not in itself explain the origin of that thing. Moreover, even if its contribution can be construed as necessary for the existence of a system of the given type, that alone cannot tell us why a thing of any *particular* type is there to make it. For example, even if we assume that a blood circulation organ is necessary for the proper functioning of vertebrates as we know them, the mere existence of healthy vertebrate organisms does not allow us to deduce the presence of hearts as we know them. In principle, some other circulatory arrangement might have worked as well. Even granting evolutionary history as we know it, the fact that a particular organism is currently alive cannot prove that it has a natural heart, since today there also exist artificial hearts. In sum, Cummins claims, any "attempt to explain the presence of something by appeal to what it does – its function – is bound to leave unexplained why something else that does the same thing – a functional equivalent – is not there instead." (Cummins 1975, 745).

However, the deductive relationship Cummins critiques is *not* in fact a feature of most etiological theories of functions – and in particular does not factor in the thought of either Wright or Millikan.[1] There is nothing in either which stipulates a logically

---

[1] To be fair, Millikan wrote several years after Cummins' paper. However, Wright preceded him and is explicitly included in Cummins' analysis.

*necessary* relationship between origin and function – a contingent *actual* one is sufficient.

Cummins is addressing the original sketches of etiological function found in (Hempel

1994b) and (Nagel 1961), each of whom presumed that "explanation is a species of

deductive inference" (Cummins 1975, 745). But later etiological theorists did not adopt

this premise and do not require that hearts as we know them be deducible from the status

of an organism. The posited *de facto* relationship that hearts were *in fact* selected for due

to their pumping of blood suffices for the attribution of this as their function, according to

both Wright and Millikan. Thus, it is clear that the relevant sort of explanation for

Wright and Millikan is not of a deductive or necessary type as it was for Hempel and

Nagel. These later etiological theories have serious difficulties of their own, as was

demonstrated in the previous chapter; but invalid logical deduction of presence from

contribution is not among them. However, it is not Cummins' critique of the relationship

between the etiology (A) and contribution (B) premises that is most significant, but rather

his critique of their independent propriety as features of a theory of function.

**Cummins on Etiology**

We now proceed to Cummins' treatment of (A), the etiological premise, which he

summarizes as, "the point of functional characterization is to explain the presence of the

item" that has the function (Cummins 1975, 741). This, he insists, constitutes "an

undefended philosophical hypothesis about how to construe functional explanation"

(Cummins 1975, 748). This is not to say that an etiological approach to function is

*always* unwarranted. An appeal to intention to account for origin is clearly appropriate in

the stereotypical case of artifacts: a need is perceived, an artifact is envisioned in the imagination of a designer to fulfill that need, and the artifact is created according to that vision, with luck successfully fulfilling that need. In such cases origin is unambiguously and uncontroversially tied to function. Although he is careful to highlight its fallibility, Cummins takes no issue with a link in principle between function and origin in artifacts. However, he views any attempt to generalize this approach to natural functions as a serious error: "For it seems to me that the question, "why is x there?" can be answered by specifying x's function only if x is or is part of an artifact. … because it rationalizes the action of the agent who put it there by supplying his reason for putting it there." (Cummins 1975, 746). In short, Cummins claims that expecting an item's function to *always* implicate its origin is to improperly apply the valid pattern of conscious intention to situations where (it is agreed) none obtains. Without an agent to appeal to, the origin/function premise is false. If the assumption of deliberate intent is in error, then "specifying the thing's function will not answer the question" of its origin (Cummins 1975, 747). Wright's attempt to abstract away from the role of the agent is therefore misguided: "Once we leave artifacts and go to natural systems, however, this approach is doomed to failure" (Cummins 1975, 748).

What should then be made of Wright's contention that natural selection supplies the requisite origin story in natural systems just as well as does intention in artifacts? Why is *natural selection* for function not an adequate substitute for *intention* for function, when in principle both seem able to link the origin of a type with its function? Cummins

counters that the form of any organism is controlled by its genetic plan alone. Given a particular genetic plan, the organism that results is entirely determined by it. If the plan specifies traits that turn out to be beneficial, the organism may prosper. Yet if the plan specifies traits which are not beneficial (or at least not in the given environment), the organism will nonetheless express those traits just as reliably. In short, the genetic plan is utterly insensitive to the prosperity of the organism. It simply specifies a blueprint that is dutifully implemented by the biological machinery "totally insensitive to what the structure does" (Cummins 1975, 750). Any "alterations in the plan are due to mutation … regardless of the function or survival value of [the resultant traits] in those organisms" (Cummins 1975, 750). This much is uncontroversial.[2] Yet it is also true that genetic plans that do not adequately equip the organism to survive will be eliminated, and conversely that those which do will become more popular. In Cummins' interpretation, however, "this is not to explain why [functional traits] occur, it is to explain why the sort of [organism] incorporating these [traits] occurs" (Cummins 1975, 751). In other words, because natural selection seems to operate at the level of the organism, it is at best the presence of the overall organism which natural selection can explain – not any of its particular traits.

Davies, it may be recalled, advanced a similar argument as a challenge to the epistemology of trait selection; but for Cummins this objection is universal and overriding. Traits for Cummins are apparently *never* selected for; only organisms are.

---

[2] Or at least, it is accepted for the purposes of our argument. In fact, epigenetics complicates this picture substantially, though it was unknown at the time Cummins wrote.

Yet this is surely too extreme, for at some level the argument devolves into mere semantics. In the simplest case, if two populations of organisms differ only in one advantageous trait (think of the famous moth wing-color example), and the one with the trait prospers while the one lacking it is eliminated, we would surely say that the trait was selected for. Ultimately, this is an argument which defenders of etiological function must take up but which we can avoid settling here. In any case, Cummins, for his part, uses this logic to dismiss the appeal to natural selection in attributing natural function on an etiological basis.

**Cummins on Contribution**

Cummins next considers part (B), the contributions aspect of etiological theory, which he summarizes as: "for something to perform its function is for it have certain effects on a containing system, which effects contribute to the performance of some activity of, or the maintenance of some condition of, that containing system" (Cummins 1975, 741). This too he considers an undefended – and indefensible – premise. Cummins' objection is twofold. First, function-as-contribution requires the principled identification not only of the function but of the special higher-level effect to which it contributes. Second, attempts to address this problem seem to initiate a regress (or, perhaps more accurately, an upward 'progress') without any apparent means to halt it.

The problem begins with rigorously identifying the effects to which a function contributes. For function to be contribution to a privileged, significant higher-level effect, we must know which systemic effects are eligible for contribution towards in

order to know what counts as a function. In the example of the heart and its dual effects, sound and pumping, which of its effects are its functions and which are incidental depends on what "special" systemic-level effect to which each contributes (or fails to contribute). To identify functions, then, we must first identify "some *special* condition of, or the performance of some *special* activity of, the containing system" (Cummins 1975, 752, emphasis added). Otherwise, any effect, including heart sounds, could be said to contribute to *some* higher-level effect, even if that effect is merely an overall soundscape. We therefore need "some principled way of selecting the relevant activities or conditions" in order to distinguish them from mere ancillary effects; we require "a general formula that identifies the appropriate effects" (Cummins 1975, 152).

An obvious candidate for such a formula would be effects that are somehow system-wide, or those which seem to exhibit a kind of homeostasis: effects, in other words, which are essential to maintaining the life of the organism. Yet herein lies the second act of the problem for Cummins. For "if we identify the function of something *x* with those effects of *x* which contribute to the performance of some activity *a* or to the maintenance of some condition *c* of a containing system *s*, then we must be prepared to say as well that a function of *s* is to perform *a* or to maintain *c*." (Cummins 1975, 753). Here, Cummins seems to posit a sort of reflexive property for function: if *x* as a part of *s* contributes to *a*, then *a* must be a function of *s*.

For example, if the function of the carburetor as part of an engine is to mix fuel with air to contribute to the mechanical power production of the engine, then it seems

implicit that the function of the engine is to make power. And indeed, we would normally agree that it is. This contention is strengthened when we consider that the engine too is part of a higher system to which it in turn contributes: it is part of a car, the function of which is to drive. Notably, at each of these steps we have the content of deliberate human intention to appeal to in justifying the attribution of functional contribution to each part. We might wish to terminate the analysis at the level of the car, yet if we wished to continue and view the car as a component of a still-larger transportation system, we could continue to do so on the same intentional basis.

But by definition, intention is not a resource which natural systems have to appeal to – and this creates problems once a single instance of such contribution is posited. If the function of the heart, as an organ of the circulatory system, is to circulate blood, then, according to Cummins, the function of the circulatory system must be to circulate blood. This conclusion is eminently plausible. But, continuing on, if the function of the circulatory system as a subsystem of the organism is to circulate blood, then what is this a contribution to? It would seem to be to maintain the organism's life. But must we then say that a function of the organism is to maintain its own life? According to Cummins, we would. But on what basis could we say that organisms have *any* functions? Moreover, if function is the contribution to a higher level activity, does it make sense to say that the organism contributes to its own survival? Or does the organism have a function in the broader ecosystem to which it contributes?

Without intention, Cummins claims that we have no obvious basis on which to justify any of these attributions of function as special contributions. This may be a genuine problem – and it is one that we will address again. But it cannot be problematic for the precise reasons that Cummins adduces. For if function is defined as the contribution to a special activity, why must we presume, as Cummins explicitly does, that that activity *is itself also a function*? By equivocating between 'activity' and 'function' in his description of what I have called the reflexive property above, Cummins has engaged, perhaps unintentionally, in some terminological sleight of hand. Logically, function can be defined as contribution to a 'special activity' *without* implying that this activity is *itself*, as well, a function. This is clear when we consider that by hypothesis function is causal contribution to a systemic effect. Thus, the functional effect must be causally 'upstream' from the systemic effect to which it contributes. If the systemic effect is *itself* also a function, it can be one only by virtue of still-higher-level super-systemic effects further downstream to which *it* contributes. To insist that the systemic level effect to which function contributes is itself a function of the system goes beyond what was stated; indeed it is a mere presumption. It is this presumption which sets up the regress (or progress) on the basis of which Cummins (erroneously) dismisses the contribution approach.

None of this is to say that there are not genuine challenges for a contributions account directly related to what Cummins has touched on here. Indeed, there are; including: How exactly do we identify the relevant 'special' activities to which functions

contribute? How do we distinguish activities which are themselves functions from those which are not? And, if the chain of functional contributions must eventually terminate in activities which are *not* themselves functions – that is, which are ends-in-themselves of some form – is this coherent in teleological terms? If it is, is this applicable to non-intentional, natural contexts? We will be led to confront these issues in the next chapter when we consider an elaborated version of the contributions account. For now, it is enough to note them, and to note that they are part of why Cummins rejects contributions accounts and continues his search for a still-more conservative account of function.

Having rejected both halves of what he characterizes as the basis for etiological approaches to function – that function must relate to origin, and that function must relate to contribution to special higher-level effects – the obvious question is: What's left? If function is about neither origin nor contribution, much less the two together, what can function be for Cummins? Without etiology or teleology, function becomes merely the causal effect an item has within a system.

### Cummins' and Davies' Systemic Effects Function

Cummins arrives at a fundamentally different theory of function by fundamentally reconceiving the task of function. Freed of the expectation to account for origin or special contributions, the explanatory role of function is reduced to mere analyses of efficient cause. Systemic effects function is thus not about *how* the heart arose or *why* the heart pumps, much less how the two might be related – but simply *what* the heart does that produces the observed larger-scale effects of its containing system. Davies, a fellow

systemic effects proponent, summarizes the approach this way: "On this view, functions are attributed relative to the larger system within which the trait operates and relative to certain capacities of the larger system. The functions of any trait are those effects that, within the context of the system, contribute to the exercise of some higher-level capacity." (Davies 2001, 3).

Oddly, and perhaps tellingly, for a system that was developed specifically to address natural function lacking teleology, the initial examples that Cummins offers are all of plainly teleological artifacts. "Assembly-line production provides a transparent example of what I mean. Production is broken down into a number of distinct tasks. Each point on the line is responsible for a certain task, and it is the function of the workers/machines at that point to complete that task." (Cummins 1975, 760). Of course, that individual or machine is capable of many other acts or effects, perhaps even simultaneously. But these other effects, for the purposes of the capacity of the line at issue, are ancillary and not its function. Thus, "we may pick out a certain capacity of an individual exercise of which is his function on the line. Of the many things he does and can do, his function on the line is doing whatever it is that we appeal to in explaining the capacity of the line as a whole." (Cummins 1975, 760). Cummins goes on to explain that "Functional analysis in biology is essentially similar. The biologically significant capacities of an entire organism are explained by analyzing the organism into a number of "systems" – the circulatory system, the digestive system, the nervous system, etc. – each of which has its characteristic capacities." (Cummins 1975, 760-761).

The claim that function is based on the production of higher-level "capacities" sounds superficially similar to the contributions approach which Cummins roundly rejected in the previous section. But in fact this is a radical shift, for there is a crucial difference. The contributions approach relied on the existence of a particular inherent "special" effect of the system to which a function contributed – and by virtue of which functional effects were distinguished from ancillary effects. Despite the examples above which strongly encourage the impression, systemic effects theory does *not* in principle rely on the presumption of any special systemic capacities (e.g., a systemic function). If it did, it would indeed come uncomfortably close to a contributions theory. Instead, systemic effects theory simply *relativizes* function to a *given* systemic-level capacity. An example of the same function as attributed by each theory may help to clarify this difference.

Contributions approach: Because the function of the car-system is to drive, the function of the engine is to make mechanical power. This is because that is the relevant causal contribution the engine makes to enabling the function of the system it is part of.

Systemic effects theory: Relative to the observed capacity of the car-system to drive, the engine's empirical causal contribution is making mechanical power. Driving is simply the systemic capacity we have *chosen* to analyze here, likely because it is the most characteristic or interesting; and in reference to it, the engine has this function.

The difference comes down to two very different senses of the word "contribution": one teleological, the other strictly causal. In endorsing particular higher-level effects as 'special', or functions themselves, a contributions approach is implicitly seeking *purposeful* contributions, contributions which, while causal in nature, will *attain* these special ends. Systemic effects theory, on the other hand, is content with tracing causal contributions behind, in principle, *any* given higher-level effect.[3] Inevitably, there will be certain higher-level effects that we are most likely to be concerned to account for, such as the productive output of the assembly line, or the blood circulation of the circulatory system. But the choice of systemic capacity – the choice which sets the standard for function – remains external to the theory, and thus from a theoretical standpoint is effectively arbitrary.[4]

The perspectives of systemic effects theory and a contribution approach are thus reversed. A contributions approach effectively considers the systemic-level effect an end, and treats the lower-level functional effects as required to get there. Systemic effects theory takes the systemic-level effect as a given, and works backwards to decompose the causal structure which produced it. Thus a contributions approach looks at function causally 'forwards' to reach a pre-designated point, while systemic effects looks causally 'backwards', starting with an observed capacity and decomposing the underlying causal structure.

---

[3] Restrictions on the nature of this global effect will be explored later.
[4] In practice, as we will shortly see, this freedom creates perils in attributing functions too liberally, so additional conditions will be invoked by Davies to limit this arbitrariness. But the point remains that it is essentially explanatory interest alone which drives function attribution in systemic function theory.

**Systemic Capacity Identification**

In practice, the functions attributed by the systemic effects theory and a contributions approach will frequently overlap, since the systemic capacities we will be most interested to explain will often be just those we think of as special or intrinsic to the system. The typical systemic effects version of cardiac function is thus as follows: "Hearts have the function of pumping blood, relative to the capacity of the circulatory system to distribute nutrients throughout the body, because pumping blood contributes to the exercise of this capacity." (Davies 2001, 3). But it is crucial to note that it is only the choice of the capacity of the circulatory system to "distribute nutrients throughout the body" that establishes the heart's function relative to it, and thus attributes the same function which proper function or a contributions approach would attribute to the heart.

Therefore, a systemic effects attribution of function should not be confused with attributing *the* singular function of the entity, even within a given system. Rather, it is always relative to a certain systemic capacity. It may seem unlikely that we would choose, for example, the capacity of the circulatory system to make sound as a relevant capacity for function attribution. But in principle, it seems that we could choose other capacities to analyze, whereupon the function of the heart will change – perhaps even to the degree that its sound effects do become functional. Indeed, as Cummins notes, "It is sometimes suggested that heartsounds do have a psychological function. In the context of an analysis of a psychological disposition appealing to the heart's noise-making capacity,

"The heart functions as a noise-maker" (e.g., as a producer of regular thumps), would not even sound odd." (Cummins 1975, 762, note 21).

For Cummins, therefore, function is only ever *function as*; there is no such thing as function simpliciter, or with the definite article, for *all function is contingent* and relative to an externally designated systemic capacity. Anything that enjoys any systemic function does so only by virtue of two things: a given systemic context, and a given observed capacity of that system which we are at pains to explain. Proper function (and likely all plausible characterizations of function) is at least implicitly contextual as well. This is why Millikan was forced to couch functional attributions in terms of "Normal" background conditions. But systemic function is explicitly contextual, not only to a specific systemic characterization, but also to a particular capacity of that system. It is ultimately only this stipulation which allows us to pick out effects which are functions from those which are not.

Whereas Wright took the first task of function to be to make a principled distinction between function and accident, for Cummins the distinction itself holds no meaning. This is deliberate. In shifting the onus of designating the high-level effects to be explained from the concept of function itself to our own independent, external choice, systemic effects theory seeks to free itself of all teleological connections and their attendant issues. Whether it is successful in doing so, and whether what remains is useful, will be discussed below.

In the end, for Cummins and Davies, function in the form of systemic effects theory is simply a way to trace efficient-cause relationships backwards. It is no more than causal analysis. If we envision a network of all actual causal relationships, to attribute a systemic function is merely to choose a set of systemic-level effects (capacities) and work backwards to identify those entities and their effects which are implicated in the capacity. Relative to that chosen system-level effect (capacity), functions are the causal effects of the lower-level components which produce it. Indeed, Davies entitles the chapter in which he lays out the basic of systemic effects theory "Tracing Links of Causation" (Davies 2001).

**Systemic Effects and Malfunction**

Finally, note that for there to be systemic function, the relevant capacities *must* exist. If they do not – for example, if the heart in a circulatory system does not pump – the heart simply does not function as a pump for the purposes of the analysis of the circulatory system's capacity to distribute nutrients. Indeed, in this case the would-be analyzed capacity itself is not present, so in this sense there is nothing to analyze. Systemic function, having deliberately avoided all teleological implications, has no resources with which, nor any inclination, to attribute malfunction. Attributions of malfunction can thus only be made through appeal to the intentions of an agent in the case of artifacts; yet even then the malfunction attribution would seem unattached to the systemic effects theory of function, and remain simply a reflection of the desire for function on the part of the agent. The situation is similar for natural function: while we

might wish to see the heart function as a pump in the circulatory system, there is no

theoretical basis on which to say that it malfunctions should it fail to have this effect.

Systemic effects theory is strictly descriptive; it is only about what actually happens –

never about what should or does not.

<p align="center">**Critique of Systemic Effects Theory of Function**</p>

Systemic effects theory was conceived to exclude all teleological aspects from

function in order to make it metaphysically narrow enough for application to natural

systems. Yet does the concept of function absent teleology retain its value?  Is it even

coherent?  The issues with the systemic effects theory of function fall into at least three

related areas: Does it supply sufficient explanatory value and retain enough coherence to

serve as a general theory of function?  Does it retain applicability in contexts where

intentional teleology is involved?  And, finally, does it attribute function with sufficient

discretion?

**Functional Explanation without Teleology**

All function theorists agree that function attribution should be explanatory in

some fashion.  Wright took the primary task of function theory to be to distinguish

function from accident.  Millikan expanded on this and held that function must always

explain type origin.  Others, as we have begun to explore above and will see more fully in

the next chapter, claim that function should explain how the part aids the ends of a

system.  Without an explanatory dimension, function attribution seems no different than

citing ordinary causal effects.  But given that causal effects are what the systemic effects

theory of function is all about, what more, exactly, does its brand of function attribution

explain?

In short, Cummins claims, it explains how things work.  This is not trivial.

Indeed, as Cummins is quick to point out, figuring out how systems work comprises the

majority of the research mission in fields like physiology. Following the assembly line

example, Cummins notes that "Schematic diagrams in electronics provide another

obvious illustration.  Since each symbol represents any physical object whatever having a

certain capacity, a schematic diagram of a complex device constitutes an analysis of the

electronic capacities of the device as a whole into the capacities of its components. *Such*

*an analysis allows us to explain how the device as a whole exercises the analyzed*

*capacity*, for it allows us to see exercises of the analyzed capacity as programmed

exercise of the analyzing capacities." (Cummins 1975, 760, emphasis added).  Functional

explanation in natural systems – that is, in organisms – is no different: "The biologically

significant capacities of an entire organism are explained by analyzing the organism into

a number of 'systems' – the circulatory system, the digestive system, the nervous system,

etc. – each of which has its characteristic capacities.  These capacities are in turn

analyzed into capacities of component organs and structures."  Indeed, the analytical

strategy of systemic effects moves so seamlessly between organisms and artifacts that

they nearly overlap: "We can easily imagine biologists expressing their analyses in a

form analogous to the schematic diagrams of electrical engineering, with special symbols

for pumps, pipes, filters, and so on."  Both machines and organisms are tractable by this

strategy because in essence this is simply a program of causal reductionism, and both artifacts and organisms operate according to causal laws. "Ideally", Cummins remarks, "this strategy is pressed until pure physiology takes over." Again, systemic effects function is a form of efficient-cause analysis.

The explanatory value of such analysis is inarguable. Yet the question persists: however important this work, is it what function attribution is distinctively about? Or is systemic effects theory guilty, like proper function before it, of taking a concept in general if imprecise use, and distorting it beyond recognition – or, alternatively, of taking the term and applying it to a rather different concept? Consider the contrast with the other two theories we have encountered. Function attribution in proper function tells us why a thing is there, and by extension, what it is for. Similarly, a contributions account tells us what assistance to the end of a system is provided by the part at issue. Both get at what parts are 'for', not just what they happen to do. It is just this teleological aspect which has, deliberately, been omitted from systemic effects theory. But is function still function without it?

Here, the question is complicated by the many senses of the word "function". We have already encountered one central semantic distinction: the difference between "having a function" and "serving a function" (or "functioning as"). This distinction is directly relevant here. Proper function theory attributes functions which are possessed by or proprietary to the part that has them (hence the name of the theory). Systemic effects theory denies the legitimacy of such possession; for it, all function is contingent.

Attributing function to a part thus becomes an exercise in pointing out what a part is doing from the standpoint of the systemic capacity at issue. Therefore, as noted earlier, all function in systemic effects theory is effectively "function as". The heart does not *have the function* of pumping – which would imply a general obligation, one which may even apply when it cannot pump – it simply *functions as* a pump; and that only from the standpoint of the circulatory system's current capacity to circulate blood. This is the essence of systemic effects theory.

But a different semantic distinction in "function" is relevant here as well: the difference between "having a function" and "how a thing functions". Asking *how* a heart functions is rather different than asking *what* the function of a heart *is*. Indeed, the questions are aimed in opposite directions. To answer the latter we would intuitively look to the context of the system the heart is in (the systemic "set" to which the heart belongs) and to something about what the heart is doing there. To answer the former, we would look within the heart itself (the "set" which comprises the heart itself) and to the action of its parts. Thus, when asked, "What is the function of the heart?" it simply will not do to reply that, "The heart works by means of the contraction of muscular chambers and the action of valves". Nor will it do to reply that, "The heart pumps blood" when asked, "How does the heart function?".[5]

---

[5] I suspect it is the surface plausibility of these exchanges which systemic effects theory trades on in presenting itself as a directly competing theory of function. But deeper reflection makes it clear that the questions involved are quite different. It is perhaps because the function of the heart is so familiar that the equivocations go unnoticed.

And yet, it is clear that these two types of questions are deeply connected. A thing must *have* a function before the question of *how* it functions is intelligible; and to ask *how* a thing functions presumes that it *has* a function. Crucially, therefore, *explaining how a thing functions depends on knowing what its function <u>is</u>*. In terms of systemic effects theory, this means that in order to attribute any function, we must know what the salient systemic capacity is. For example, it may happen with a mysterious organ like the thymus that we do not initially know what the significant systemic capacity we should wish to explain *is*. In such cases, we will inevitably be drawn to ask questions around what the organ is "for". To answer this, we might investigate how the organ arose, why it was retained, or what benefits it seems to currently confer to the overall system. In this, we are again looking for functional attributions of the etiological and teleological kind. Thus, such accounts, rather than being obviated by systemic effects analysis, could ironically be interpreted as prerequisites for it to meaningfully apply.

Systemic effects theory is only able to get away with focusing on the "how" to the exclusion of the "what" if the systemic capacity is identified before the functional analysis begins. But at this point it is not obvious how, if at all, systemic effects theory aids that identification. The capacity identification issue noted earlier has thus returned with new urgency. We can now see that the reason that this issue has resurfaced is that such identification is integral to the basic concept of function attribution. To have a function, or even to serve a function, seems to require an *end* that that function is oriented *towards*. This is what seems to give function meaning. Without a teleological dimension

to aid in identifying the capacity, the task seems to fall to some interest or agent outside

the theory.  But this task is so integral to the concept of function that it cannot remain

outside – its absence will be continually implicated in the weaknesses of the theory.

Thus, to remain viable, systemic effects theory will need to show how it can aid this

identification from its own resources.  Otherwise, the capacity will remain essentially

arbitrary, and so will the resulting functions attributed.  We will return to this important

issue when we consider the related "promiscuity" objection commonly leveled at

systemic effects theory.

**Function without Malfunction**

As noted above, the lack of teleology in systemic effects theory means that it has

no place for functional norms – and thus no concept of malfunction.  But what are the

implications of function attribution absent even the *possibility* of malfunction?  Is it

coherent (or, if so, is it helpful) to say that a thing has a function (even if only relative to

a particular systemic context and systemic capacity) and yet if that part breaks down, no

malfunction has occurred?

It is easier to address the coherence question first, for here systemic effects theory

has a sound defense.  It may be logically incoherent to say that a thing *has* a function yet

does not malfunction when it fails to perform it – for what else is function "possession"

but the "obligation" to perform the function?  And if the obligation is not met, then an

attribution of malfunction seems apt.  But, as we have been at pains to note already,

systemic effects theory does not attribute function possession in the same way as proper

function.  For systemic effects, function is always "function as", and merely *serving* a

function implies an ad-hoc relationship which need not, intuitively or logically, imply

malfunction when it ceases to be performed.  This is for the same reason that we would

not accuse a rock of malfunctioning when we attempt to use it unsuccessfully as a

hammer – we have no prima facie right to expect function from it, for the rock holds no

apparent functional office.  Likewise, expectations are not part of the role in systemic

effects function, for the point of moving to systemic effects function is escape the

metaphysical quandaries associated with such roles in natural systems where there is no

conscious intention to appeal to.  The question now becomes: Is mere "function as", with

its "no fault" functional implications, sufficient to fulfill our intuitions and typical use of

function attribution?

At the least, such a shift interferes with the application of function to a number of

areas where natural function attributions – and more particularly natural malfunction

attributions – are routine.  Medicine is the notable example here.  Taking the basic

mission of medicine as repairing the body, function attribution and medicine part ways if

in systemic effects theory there can be no such thing as a malfunctioning organ; for

example, a bad heart.[6]  Of course, we may still evaluate the state of affairs normatively

on other bases.  We can say, for example, that the heart is bad because the effect it is

having is not as we prefer.  But if we want to make any malfunction attributions we are

---

[6] The loss of malfunction attribution in organisms would also have seriously disruptive effects on other
philosophical areas that refer to or rely on normative function.  For example, in philosophy of mind no
normative references to the function of the senses or the brain in representing the world would seem
possible.

forced to go outside of the theory of function to make them or we must repackage them in terms that make no reference to function. Both avenues are possible, but are seriously disruptive to our intuitions of function and its relationship to matters which look to function for normative evaluations.

There is also a somewhat infelicitous phenomenon relating to the loss of function due to the loss of capacities, rather than the reverse. The heart helps to explain the capacities of the circulatory system, and it has its typical function in virtue of this. But what if the capacities of the circulatory system are impaired by something *other* than a loss of function in the heart? Suppose, for instance, that there is a blood clot, a loss of blood, or a problem with the lungs. In these cases, the relevant capacities of the circulatory system *as a whole* are lost. But with this loss of systemic capacities, which once anchored the function of the heart relative to it, logically, the heart ceases to have any function. Once there ceases to be a systemic capacity to explain, all systemic function relative to it goes missing as well. This seems counterintuitive, but it is the self-consistent result of systemic effects theory given the contingent nature of function in it.

We should readily admit that the health of the heart becomes less relevant, at the extremes perhaps irrelevant, under these circumstances. But stripping it of any function whatsoever seems extreme, inviting the impression that its presence in the system or in the body no longer has any meaning at all. But this is the result of a concept of function which is not about meaning or purpose but only causal analysis. As with malfunction

attributions, these results are internally coherent – but they are also problematic for the way that function is actually applied in the world.

There is one last, perhaps even more awkward result related to malfunction: in any given token, effects are produced both by what we would normally consider correct function as well as what we would normally consider malfunction. Thus, there is no apparent prohibition on saying that the function of a broken leg is to immobilize the animal. Without the break, the animal would be running, so the break in the animal's femur plays a causal role in producing the immobility. When function becomes solely about tracing the causal effects that produce a phenomenon, there seem to be nothing to prevent such perverse functional attributions. This can occur even on the type level. Davies (addressing an objection from Mohan Matthen) admits that under a certain analysis the unwieldy tusks of the narwhal whale (which were presumably selected for attracting mates) also perform the function of preventing the animal from moving quickly.[7] In nearly all the examples proffered by both Davies and Cummins, every effort is made to select systemic capacities which conform to our intuitions of function in terms of contribution and even origin, all while insisting that these things are not properly part of function attribution. It is true that the tremendous flexibility of systemic effects theory allows it to endorse almost any intuitive function attribution we would care to make – but

---

[7] Fascinatingly, here Davies adduces an argument that the immobilizing function of the trait may really exist at the population level – that particularly large tusks may have prevented the individuals that had them from winning mating fights, and that this check helps to explain "the capacity of the [large tusk] population to reach equilibrium" (Davies 2001, 97). However, since by hypothesis the trait was selected *against*, we are still left with essentially the same kind of unintuitive function attribution with which we began, and little has been gained in the effort. The function of the trait has become to select against itself. This is at least as strange as saying that the function of the trait is to slow the animal.

it is also true that the theory in and of itself provides no real guidance in making these attributions; and it is just as liable in principle to endorse function attributions which are unintuitive.

Therefore, while systemic effects theory is metaphysically conservative, on its own it does not seem sufficient to do all of things that we expect of a theory of function. It may be that our expectations of teleological function are metaphysically naïve, and require correction. If all teleological accounts of natural function are guilty of involving metaphysically unsupportable attributes, as Davies claimed, adopting systemic effects theory as the only naturalistically plausible account may be justified (indeed, necessary). But the disruptive nature of this should not be overlooked or underestimated, and the abandonment of so central a facet of function should be compelled by real necessity. It is, perhaps, too early to resort to this.[8]

Moreover, eliminating these basic teleological and normative aspects of function complicates the application of the theory to what is perhaps the paradigmatic, and normally the least problematic, area of function: artifacts, and intentional function in general.

**Dichotomy between Natural and Artifact Function**

As he begins the section describing systemic effects theory, Cummins is upfront about his area of concern. "Throughout this section, I am discounting appeals to the intentions of designers or users. … There can be no doubt that a thing's function is often identified with what it is typically or 'standardly' used to do, or with what it was

---

[8] Besides which, systemic effects theory does in fact have problems of its own, as we will shortly see.

designed to do. … Functional talk may have originated in contexts in which reference to

intention and purposes loomed large, but reference to intention and purposes does not

figure at all in the sort of functional analysis favored by contemporary natural scientists."

(Cummins 1975, 757-758, note 13). In other words, while artifacts are the prime

example of function attribution, their direct analogy to natural systems is inappropriate;

therefore an entirely different concept of function is required. For his part, Davies

likewise explicitly begs off the question of artifact function, admitting early on in his

seven-chapter monograph that "Extending the theory of systemic functions to the relevant

sorts of intentional and conventional factors involved in the functions of artifacts is a

project for some other occasion." (Davies 2001, 8).

Lacking such an account, natural and artifact function are divided into separate

spheres, as conscious intention is involved in the latter but not the former. One of the

claimed virtues for the etiological accounts of Wright and Millikan was that they served

to unify natural and artifact function, as both were explicable by appeal to origin. A

different sort of origin was involved in each case, to be sure; yet in both cases, the

function of the item is that effect it has which helps to explain how the item 'got there',

and this was appealed to in order to distinguish function from accident. But systemic

effects theory, having thrown off etiology and teleology, does not clearly apply to

scenarios where both are involved.

This leaves the formerly more secure picture of artifacts with a dilemma. Can the

theory of systemic effects account for the function of artifacts as well, or not? If it can, it

can only be one sense of function alongside some other, for function in artifacts is scarcely comprehensible without the possibility of malfunction, or without function being tied to the designer's intention in placing the part. This would leave us with two distinct senses of function in artifacts – one teleological, the other not. This hardly seems desirable, and it may not even be coherent, as they would directly conflict in attributing function in cases of malfunction. If systemic effects theory does not apply to artifacts at all, then there are still two different concepts of function at work in the world at large: one for intentional scenarios, and another for non-intentional.

According to Cummins, etiological theories fail because they neglect to "distinguish teleological explanation from functional explanation" (Cummins 1975, 747). And yet, systemic effects theory seems at risk of failing (if not outright, then at least in comparison to other accounts) by divorcing the two. This brings us to the third, and perhaps most revealing problem with systemic effects function: just when does it apply?

**Systemic Effects And The "Promiscuity Problem": What Things Have Function And What Do Not?**

A virtue of the etiological accounts, as documented in chapter 2, is that they are quite clear about the conditions function could be attributed under, and when it could not apply. In Wright's terms, they distinguished function from accident quite rigorously. Unless the item was somehow brought about due to its purported functional effect, no function can be attributed to it. Thus, these theories would never hold that simple natural phenomena – for example, rainclouds – had functions just because they were could be

construed as beneficial from some perspective. These systems have no function because their origin has nothing to do with their purported functional effects. To perhaps over-simplify, clouds arise via the accumulation of water vapor in the atmosphere, and they produce rain when that accumulation cannot be sustained further. When they produce rain, the moisture is returned to the soil – among other things, aiding the growth of vegetation. But we are not normally inclined to say that the function of clouds is to water plants; this violates basic intuitions of what function is about. Etiological theories honor this intuition by showing how clouds do not have this function because their origin is entirely disconnected from the effect at stake: even without the existence of plants to water, there would still be clouds and rain.

Such a distinguishing principle does *not* seem built into systemic effect theory. Systemic effects theory looks only to the explanation of observed systemic capacities. But, as was highlighted earlier, it is unhelpful in identifying or qualifying those capacities. The concern therefore is that systemic effects theory "licenses the attribution of function when in fact no such functions exist." (Davies 2001, 73). The example of the rainclouds is one Millikan raised in pressing this allegation: "according to Cummins' definition it is, arguably, the function of clouds to make rain with which to fill the streams and rivers, this in the context of the watercycle system, the end result to be explained being, say, how moisture is maintained in the soil so that vegetation can grow." (Millikan 1989b, 294).

The basic problem is one of distinguishing effects from functions, not in any specific context but in general principle.[9] If function is to retain any value as a concept separate from mere causal effect, we must know what separates them. Even though neither hearts nor rainclouds have purposes for systemic effects theory (as they are natural objects produced without intentions), intuitively, hearts have functions and clouds do not. Unlike the intuition of natural purposiveness, this distinction is an intuition which systemic effects theorists would like to uphold in some form, if only to avoid the embarrassment of attributing function to nearly anything. Indeed, the problem, which has come to be known as the "promiscuity problem", is the risk of attributing function so liberally that the term becomes useless. If every effect is a function from some obscure perspective, function attribution becomes essentially meaningless.

Trivializing the difference between a mere causal effect and a function is a risk which systemic effect proponents, including Cummins, recognize: "If the account I have been sketching is to draw any distinctions, the availability and appropriateness of analytical explanations must be a nontrivial matter." (Cummins 1975, 763). Without nontrivial criteria for application, function can be attributed to any chain of effects whatsoever.

The basis for Millikan's objection is quite clear: "it is quite true that, in the context of the water cycle, clouds function to produce rain, function as rain producers; that is their function in that cycle. But in another sense of "function", the clouds have no

---

[9] Note that this is a different issue than attributing function to traits which are disadvantageous, as explored earlier. This is about attributing function to entities which intuitively have *nothing to do with function* at all.

function at all-because they have no purpose." (Millikan 1989b, 294). Millikan takes

etiological origin to be constitutive of purpose, of teleology itself. We know that clouds

have no purpose because we know that they did not arise as a result of any of their

effects, such as plant-watering. And for Millikan and Wright, if there is no purpose, there

can be no function – merely effects. But, prima facie, etiology or teleology need not be

the operative criteria for distinguishing function from effects. Other principles may serve

just as well. Indeed, we already know that it cannot be etiology or teleology for systemic

effects theory, as even the most unimpeachable instantiations of function have no purpose

for it – that is, after all, the point of the theory. The question is: What principle within the

slim resources of systemic effects function theory can do this necessary job? Since for

systemic effects theory function is whatever leads to the relevant systemic capacity, the

answer would seem to depend largely on what makes something a systemic capacity.

**Cummins' Conditions**

Cummins suggests the difference must relate to our interest in producing a

functional explanatory analysis of a capacity. But this interest must itself have guidelines

for the promiscuity problem to be addressed – otherwise we could claim interest in

virtually anything. Cummins therefore provides three conditions which he says reflect

our interest in seeking functional analyses of the kind systemic effects theory provides.

"The explanatory interest of an analytical account is roughly proportional to (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 sophistication of the program appealed to, i.e., the relative complexity of the organization of component parts/processes that is attributed to the system. (iii) is correlative with (i) and (ii): the greater the gap in sophistication and type between analyzing capacities and analyzed capacities, the more sophisticated the program must be to close the gap." (Cummins 1975, 764).

This is the difference, he tells us (slipping again into convenient artifact examples to defend a theory intended to apply to natural systems), between the functional analysis a digital synthesizer's capacity to produce sound and a bell's capacity to produce the same sound. Even when both produce the same sound, the means by which the synthesizer does so are so much more complex and multi-layered that a functional analysis of this capacity is immensely interesting and useful. On the other hand, little can or need be said about the bell, which is simply a resonating piece of metal. In a similar way, presumably, the capacity of the heart to pump blood is complex and interesting, while the capacity of the cloud to rain is simple.

Having said all this, he notes that "It must be admitted, however, that there is no black-white distinction here, but a case of more-or-less. As the role of organization becomes less and less significant, the analytical strategy becomes less and less appropriate, and talk of functions makes less and less sense." (Cummins 1975, 764).

Two questions then arise: First, is this so? Does interest in explanatory analysis really track with organizational complexity? And second, if so, does it matter? To the first question, recall that the functional explanations provided by systemic effects theory

are not of the part, per se, but of the system in which the part appears.  Thus, in

attributing function to the cloud, what we are really doing is explaining the capacity of

the entire water cycle system.  And this, arguably, is a great deal more complex than a

cloud alone – indeed, it seems very much like the kind of thing which we might be

interested to explain, particularly in the details. [10]  Indeed, this is much of what the field

of meteorology is composed of.

So it is not the case that simpler entities necessarily imply simpler explanations.

But the second is the more important question: does it matter?  In other words, is our

explanatory interest the right yardstick for when function applies and when it does not?

Because systemic effects theory outsources the identification of the capacity to be

explained, this seems to be all that Cummins can appeal to.  Yet our intuitions of function

tend to be that function should be more about the objects or systems themselves than our

curiosity about them.  Clouds shouldn't gain function simply because we are presently

interested to explain rain.  It is here that Davies, who otherwise essentially endorses

Cummins' theory as it stands, offers his one major refinement.

**Davies' Hierarchical Constraints**

While Davies notes that he essentially agrees with the substance of the conditions,

he allows that "Cummins's (i) and (ii) are too abstract to give a clear answer in all cases.

… It is not always obvious when one phenomenon is simpler than or different in kind

from another." (Davies 2001, 79-80).  In addition to being an ambiguous judgment in

some cases, the lack of clarity in the criteria is also a problem: "the requirement that the

---

[10] Davies, somewhat oddly, directly denies this, as we will see.

analyzing capacities be simpler and different in kind prompts the question, Simpler and

different in kind *in what respects*?" (Davies 2001, 81).  Given these weaknesses, Davies

takes his task to be making the content of Cummins' conditions more structured and

rigorous.

Rather than focus on the distinction in capacities at each level, Davies takes a

wider view and characterizes the relevant systems as a whole.  "Of primary importance is

the *kind of system* we are trying to analyze in terms of systemic functions." (Davies 2001,

82).  In a word, he holds that they must be hierarchical.  "The sorts of phenomena to

which the theory of systemic functions properly applies are those that are hierarchically

organized.  A system is hierarchical if it exercises a capacity at one level by virtue of the

organized capacities operating at some lower level of organization." (Davies 2001, 82).

This, Davies claims, is all that is needed to address the promiscuity problem, because

"Restricting the theory to hierarchically organized systems provides resources with which

to distinguish the functional from the merely causal" (Davies 2001, 86).  This turns out to

be quite straightforward: "Unless the specified system consists of two or more levels of

organization, where the capacities at one give rise to those at the next, the theory of

systemic functions cannot be applied" (Davies 2001, 84).

Cummins justified the introduction of his conditions by way of our explanatory

interest; but Davies' justification for his hierarchy restriction is indeed more rigorous.

The theory of systemic effects is after all a theory of *analysis* – of breaking systems into

parts.  And in order for there to be an analysis, there must be joined parts to analyze, for

what has no parts cannot be broken apart. Essentially, all that Davies has done here is to clarify the prerequisites for systemic effects analysis. To analyze a capacity, the capacity must have stages or parts to analyze. The only thing outlawed, then, is monolithic systemic capacities without any analyzable levels of capacities.

By Davies' lights, clouds are just such an example. Rain is not a systemic capacity flowing out of a set of lower-level organized capacities. It is just different forms of water in a physical system. Thus, "the water cycle is not a hierarchical system. It is a mere cycle." (Davies 2001, 94). Therefore, rainclouds do not have systemic function from any capacity standpoint, and Millikan's charge that systemic function may attribute function to them is evaded. If we accept both Davies' prescriptions on hierarchy and his interpretation of the water cycle, then clouds do not have systemic functions. But even so, it is not because they are in any way metaphysically different than hearts: it is only because clouds and their capacities are too simple to be functionally analyzed.

The question is then: Is this an adequate response to the objection of promiscuous function attribution? In other words, is the line that capacity hierarchy draws a defensible one for the function/effect divide? Is it selective enough to do justice to our intuitions of what function is about? While there may be some overlap, note that the set of systemic-function-eligible hierarchical systems is not the same as the set of proper-function-eligible (implicitly) teleological systems.[11] Indeed, the former seems a great deal larger than the latter.

---

[11] I say "implicitly" because Millikan's focus is more on the functional parts themselves than their systemic context – this is part of difference between proper function, which looks at the origins of parts, and

**Davies' Hierarchy Constraints Critiqued**

We know that, according to Davies, the hierarchy requirement enables us to attribute function to the heart, but not to clouds. This litmus test is passed (again, at least according to Davies). But this is just one pair of items. How does the requirement fare *overall*? Clearly, it is not too restrictive in that it allows function to be attributed to the heart, and presumably to all the other significant systems and capacities of biological organisms. But is it restrictive *enough*? After all, a great many physical systems could still be characterized as hierarchical that we would not otherwise think of as functional.

And indeed, Davies admits that the hierarchical restriction "fails to prevent [the] proliferation [of functions], since even simple natural objects are composed of lower-level components (molecules, for example) that give rise to higher-level capacities." He then adds, "Some may take this as an objection, but I do not. … [for] … To say that systemic functions are ubiquitous is to say, at least in part, that hierarchically organized systems with higher-level capacities are ubiquitous." (Davies 2001, 85). This ubiquity – and its problematic nature – is demonstrated in the case of table salt dissolved in water. Salt crystals are only slightly more structurally complex than clouds – yet this incremental increase is enough to make the system "hierarchical in some minimal sense", and thus tractable by functional analysis (Davies 2001, 101). As a result, Davies admits

---

systemic functions, which explicitly looks to systemic context. But in attributing purpose to functional parts, a teleological systemic context is implied (or, perhaps, constituted).

that they can enjoy systemic function attribution, even when the system they appear in is no more complex than a glass of water.[12]

"Consider the capacity of salt to dissolve in water. We explain this capacity by appeal to the bonding capacities of certain kinds of molecules." (Davies 2001, 52). Because these capacities are organized at a lower level and collectively result in higher-level capacities, "We may analyze the capacity of salt molecules to dissolve in water in terms of the systemic functions of constituent molecules" (Davies 2001, 6). The system at issue is here is presumably Sodium Chloride in crystalline form introduced to a large collection of liquid water molecules. This is a somewhat odd system, for it consists of two separate entities brought together in a one-time encounter. Yet perhaps the analysis goes something like this: With respect to the capacity for the salt crystal to dissolve in water, the function of the ionic Sodium-Chloride bonds is to break under the strong polar attraction of the water molecules and thus to release the Sodium and Chloride ions into solution. In this system, with respect to the capacity of dissolution, the function of the ionic bonds is thus to hold the crystal together until water is introduced. Is this something we should recognize as a function?

At the very least, the function attributed to salt's ionic bonds is one that we would never think to make outside of such a specific and artificial functional analysis. It does not comport with our regular intuitions of function attribution. Moreover, it seems clear that this would not satisfy Millikan's original objection, which was that clouds had no

---

[12] It would be a quite different matter, as we will see in chapter 4, if the salt was seen playing a role within an organism. Thus, the objection is not simply that salt is attributed a function, but that it is attributed a function essentially in isolation.

purpose. Surely, neither do salt crystals in a glass of water (human intentions excepted). If anything, intuitively, clouds seem closer to having functions in the water cycle system than do the ionic bonds of the strange "salt-dissolving-in-water" system. Thus, it is only Millikan's choice of example that allows Davies (and by extension Cummins) to answer her objection. If Millikan had thought to object to the attribution of function to salt-in-water rather than water-in-air, according to Davies own analysis the charge would stick. But of course the real force of Millikan's objection was never one that systemic effects function proponents would recognize in the first place; she complained of a lack of natural purpose in the ostensibly functional (cloud) system, but Davies and Cummins deny that such a thing exists anywhere. Obviously, then, they cannot recognize it as a criterion for function attribution.

Teleology aside, however, given that some discrimination in attributing function is called for, is the hierarchical requirement sufficient? Is it a mode of function attribution we could learn to live with, even if it would take a recalibration of our basic intuitions of function? At this point, it seems an open question whether the structures Davies has erected to contain and direct the application of systemic effects theory are effective in producing an overall viable theory of function. At the least, we know it will not be one that our present intuitions entirely accord with. Systemic effects theory remains more about *how* systems function than what their functions *are*. And yet, as exquisitely sensitive to the definition of both the system and the capacity as its flexible function attributions are, it provides only the most rudimentary guidance on how to

define either.  Somewhat more guidance comes from the surprisingly close relationship

Davies' theory mandates between the two.

**Systems and Systemic Capacities**

As we have seen, one of the stock examples of the promiscuity charge is the

potential for function attribution resulting from the purported capacity of the circulatory

system to make sound.  Given this as the systemic capacity, one could presumably say

that the function of the heart is to thump rather than pump.[13]  Davies holds that

Cummins' original conditions likely, if not definitively, rule out the attribution of sound-

making as a function of the heart on the grounds that there is no difference in kind in the

causal contribution across its parts.  But, as we saw, Davies reinterpreted Cummins'

explanatory conditions in terms of the overall systemic structure in order to make them

more rigorous.  He must therefore show how his systemic hierarchy requirements

produce this prohibition more definitively.  Yet the circulatory system is clearly

hierarchical, and it obviously makes some sound.  Why then is its capacity for making

sound not legitimate for function attribution?  It is because, Davies says, "not every effect

of a system is a systemic capacity" (Davies 2001, 80).

Let us back up slightly and recap in order to appreciate this development.  Davies,

like Cummins, agrees that "a central burden on any theory of functions … is to provide

principled grounds for distinguishing between mere systemic [component] capacity and

---

[13] Indeed, as noted earlier, Cummins entertains this attribution at one point, at least under the condition that the ultimate capacity at stake is psychological comfort.

systemic [component] capacities that are functional." (Davies 2001, 85). [14] Not all effects

of an object are its functions, for otherwise function would lose its meaning. For

systemic effects function, "a mere [component] capacity is one that contributes little or

nothing to some more general capacity of the containing system" (Davies 2001, 85). "A

genuine systemic function [of a component], by contrast, is a capacity the effects of

which contribute to the exercise of some more-general capacity of the larger system."

(Davies 2001, 86). Functions are defined as those lower-level effects that contribute to

the systemic capacity at issue.

Thus, it is clear that the sound the heart makes is not a function it performs

relative to the circulatory system's capacity to circulate nutrients, because its sound

makes no contribution to that capacity. Yet it had seemed, based on our understanding of

systemic effects theory to this point, that to successfully make such an attribution all that

was need was a simple change in perspective of the capacity. The shift in focus from the

system's capacity to circulate blood to its capacity to generate a particular set of sounds

would seem enough to allow the attribution of systemic function of noise-making to the

heart.

Davies here denies that this is the case. Sound production, he says, though clearly

an effect of the system, is not in fact a *systemic capacity* of the circulatory system at all –

and thus no function can be attributed relative to it as part of that system. It turns out that

---

[14] It is sometimes difficult to know when Davies is referring to lower-level capacities of components *within* systems and when he is referring to the higher-level capacities *of* the system as a whole. He often slips between them using the same generic term to refer to each level. Thus, some quotes are offered here amended with my best interpretation based on context.

it is not enough for the system which produces the capacity to be structurally hierarchical. True systemic capacities are only those which are hierarchically produced – they must be the result of multiple lower-level capacities propagating up the levels of the system.  In fact, they must involve *every hierarchical level of the system* in order to qualify as a systemic capacity.  In other words, the systemic capacity must be the product of the entire hierarchical system operating as whole.[15]  This has rather dramatic consequences.

In the example at hand, sound is not a systemic capacity of the circulatory system simply because not every component in the circulatory system contributes to it. Capillaries, notably, are presumably silent, unlike the heart and blood flow through the major arteries.  Therefore, the circulatory system, defined as all those elements which together produce blood circulation, necessarily including the capillaries, cannot be characterized as having the systemic capacity of sound production as not every part in it participates.  In regards to the sound that is produced, "The relevant system under analysis is not the circulatory system but rather a small aggregate of noisemakers that also happens to be part of the much larger circulatory system."  As a result, "the theory of systemic functions does not warrant the attribution of the function of making noise in the context of analyzing the circulatory system.  At worst it warrants the attribution of making noise in the context of analyzing a quite different 'system' – a system that has as one of its components a small aggregate of noise-making traits." (Davies 2001, 92).

---

[15] There are many profound issues implicated in this sketch, including issues surrounding the definition of the discrete levels and the emergence of more holistic capacities as distinct from their components.  I overlook such potential issues here, as Davies does, and concentrate only on what the net outcome means for the broader function debate if we accept this system.

Thus, for Davies, the relevant system is implicated by the chosen capacity. If there are components of a system that do not contribute to the capacity at issue, they are not considered part of that system as defined by that capacity. Thus, when one attempts to attribute sound production to the noisy parts of the circulatory system, "the relevant system is not the circulatory system as a whole, but merely the heart, the arteries, and the blood" for only these parts of the circulatory system are involved in making sounds (Davies 2001, 92). What had seemed separate issues – systemic context and systemic capacity – turn out to be very closely joined. This codependency of system and capacity directly impacts the range of function attribution.

Ultimately, it means that for any given hierarchical system, there is one or only a few capacities that qualify as systemic and are thus viable for function attribution; indeed, a hierarchical system will typically have a single *characteristic* capacity that distinguishes it – the characteristic effect that results from contributions of all parts in every level in the system.[16] It turns out, then, that given a system, systemic effects theory provides more guidance in the selection of relevant capacities it had heretofore seemed.

**Circularity and Constraints on Function Attribution**

Yet are things really so simple? There is a circularity involved that reveals the latent arbitrariness still present in the theory. For any given system, we can now (in

---

[16] Though Davies made it explicit, in retrospect similar principles are evident in Cummins. "The biologically significant capacities of an entire organism are explained by analyzing the organism into a number of 'systems' – the circulatory system, the digestive system, the nervous system, etc. – *each of which has its characteristic capacities*." (Cummins 1975, 761, emphasis added). In a footnote, he further adds that "what makes something part of, e.g., the nervous system is that its capacities figure into an analysis of the capacity to respond to external stimuli, coordinate movement, etc." (Cummins 1975, 761, note 18).

principle) pick out the relevant systemic effect – the one (or few) which is the combined

effect of all the discernible levels of capacities within the system.  But at the same time,

we must already know what that top-level capacity is in order to identify that system.

This is not necessarily a vicious circle, but neither is it a strong constraint on seemingly

arbitrary function attribution.

For example, the function of the heart in the circulatory system is to pump blood,

because the systemic capacity of the circulatory system is to circulate the blood, and

pumping is the causal contribution that the heart makes to this systemic capacity.  This is

all very well.  But how do we know that the salient capacity of the circulatory system is

to pump blood?  Because this is what we empirically observe it to be currently doing.

And what is this 'it', the circulatory system?  Just those parts that contribute to this

capacity to circulate the blood.  Thus, identifying a capacity allows us to backtrack in

defining the system, making hierarchical constraints on attributing function relatively

minimal.

Indeed, the production of sound can still likely be attributed as a systemic

function of the heart – it just cannot be attributed that function as part of the "circulatory

system".  But as part of a hypothetical "sound production system", it might.  This would

require that "the production of noise contributes to the exercise of some higher capacity

… assuming that there is a higher capacity to which the production of these noises

contributes, then, crucially, the heart has the systemic function of making noise *relative

to that system.*" (Davies 2001, 93).  Yet if the system has already been defined according

to the sound capacity at issue, why should we question whether the higher capacity exists or not and hesitate in attributing function on its basis?[17]  After all, the flexible capacity-defining-system logic even allows the attribution of function to the ionic bonds in dissolving salt.

**Systemic Capacities and Contributions**

Although the constraint on systemic capacities may have relatively little effect in prohibiting function attributions, as long as we are willing to be flexible on the system that has them, for systems that are already well-characterized in terms of their top-level capacities, systemic effect function does seem able to attribute the functions we would expect to see within them.  Indeed, for complex systems where the characteristic capacity is confidently known, and which we intuitively think of as teleological, a systemic effects attribution looks very much like a contributions approach attribution.

The systemic capacity constraint has the effect of establishing a certain "special" capacity to be explained for each system – the capacity by which that system is defined. As may be recalled from our initial discussion of Cummins, what purportedly distinguished a systemic effects account from a contributions approach is primarily the postulation of a *special* capacity to be explained for any given system, which alone is eligible for function attribution.  Systemic effects theory has ended up far closer to a contributions approach than would have been expected from those remarks.  Yet the large difference that remains between them is that a contributions account is explicit that this

---

[17] Or, if this is interpreted to mean that the systemic capacity must itself contribute to some still-higher capacity, then we are truly back to a full-fledged contributions approach.

special high-level effect is the teleological "end" of the system, and that only teleological systems properly enjoy function attribution, whereas the systemic effects account, as we have seen, is considerably more liberal in its attributions, and more conservative in the nature of the function that is attributed.

Yet in nominally teleological systems such as organisms, systemic effects theory is nearly indistinguishable from a contributions approach, though without the teleological implications. In non-teleological systems, systemic effects attributes functions where a contributions (or, for that matter, etiological) approach would not apply, and where we would not be intuitively inclined to apply the term, such as the function of the ionic bonds in salt. The salt example thus remains as a signal reminder that the systemic capacity constraint is not a guarantee that the capacity will be one that we normally think of as the purpose of the system (nor, of course, is it intended as such). Systemic capacities do not by definition identify what the system is "for", even if those effects coincide in practice in nominally teleological systems. The systemic capacity, even in complex systems, is not in itself a function much less an end of the system – it is just something that the system does, something that happens. Thus, both hearts and salt crystals can enjoy function attribution from their own unique systemic capacity perspectives.

The question, finally, is whether, given the parallels, a systemic effects approach is preferable to a contributions approach. Is it important to deny teleological norms even when that creates problems with how function is usually applied? Is it helpful to be able

to attribute function to things like the ionic bonds of salt in water?  Perhaps the only reason it might be judged so is if the strictures of naturalism – of its apparent a-teleological-ness – force us to a theory of function that is likewise stripped of all teleological dimensions.  But might there be a way to make a teleologically-informed contributions approach compatible with both natural systems and naturalism?  To answer this, we must more closely investigate the contributions approach, which has only been sketched in a most rudimentary form here.  This is the task of the next chapter.

## Conclusion

For any intuitive function attribution we may be inclined to make – whether natural or artifactual – systemic effects function stands ready to endorse it.  We have merely to specify the appropriate system/systemic capacity, and if the function is currently being performed, and the system is minimally complex, systemic effects theory will endorse the attribution.

Yet the function that is attributed will be of a limited nature: there is no purpose implied, and no possibility of malfunction in a systemic effects attribution.  This makes the attribution itself of relatively less explanatory value, and forces us to rely on extra-theoretical concerns to speak intelligibly where purpose and norms of function are presumed important.  This complicates the theory's application and relationship to medicine and even more so to engineering.  In addition, systemic effects theory stands ready and willing to attribute functions to things which we would not normally be inclined to attribute function to at all.  Indeed, any system which can reasonably be

characterized as having at least two stages of causal effects or "capacities" becomes

eligible for function attribution – even something as simple as a salt crystal.  The

constraint that the systemic capacity reflect the entire hierarchy of the system turns out to

be relatively minimal, as the system is redefined on the fly based on the selected capacity.

So long as the capacity exists, a system can be found to reflect it, and so long as that

capacity can in some way be characterized as having at least two interacting levels,

systemic functions obtain relative to it.  This is what allows the dissolving salt crystal to

have function attributed within it based on nothing more than it ability to dissolve.

Systemic effects theory as articulated by Cummins and Davies is explicit in

avoiding teleological implications because the systemic effects theory is a reaction to the

metaphysical anomalies of the etiological theories explored in the previous chapter.  In

essence, the theory is the result of highlighting the problems with etiological theories of

function (Cummins reacting to Wright, Davies to Millikan), and excising the offending

etiology and teleology as problematic, unnatural and ultimately unnecessary.  But in

doing so, the theory loses more than just the resources to attribute malfunction; it also

loses a principled way of distinguishing function from mere effect.  While its attempts to

make up for this lack are valiant, and partly successful, in the end they seem inadequate.

If a simple, natural salt crystal can be attributed function in the act of dissolving in water,

something has gone wrong – or at least it has diverged substantially from our intuitions of

what function is primarily about.

As a result, in supplying what we typically expect of a theory of function – independently identifying function from among other effects; attributing, or at least conceptually accounting for, malfunction; providing additional explanatory insight – we find that it either denies the legitimacy of the intuition or forces us to appeal to resources outside of the theory itself.  Functions can be distinguished from effects, but only relative to the systemic capacity at issue, and only if that capacity is currently observed. Malfunction attributions can perhaps be made, but only outside of and unrelated to the function attributions of systemic effects theory.  Function attribution can tell us what a part is contributing to a systemic effect, but not what it is for, what it means, or why it is there.  In the end, systemic effects theory simply seems too pared-down to do the job we would otherwise expect.

Is this an adequate general theory of function?  Consider the theoretical criteria applied to proper function in the previous chapter.  It was suggested there that part of the criteria when choosing a theory of function is its ability to 'get at' what is meant by the term as it is already used, to offer greater explanatory depth and precision in doing so. Does systemic function do this?  Considering that systemic function has difficulty engaging with the areas where function attribution otherwise seems most secure – artifacts – it is hard to argue that systemic function stands ready to serve as the overarching philosophical account of function.  Undoubtedly, it does get at an important dimension of function – causal contribution.  But in order for this kind of attribution to

proceed, the systemic effect to be contributed to must first be specified, and systemic effects theory provides only schematic guidance in doing this.

We earlier noted that the challenge for any theory would be to strike "a balance between breadth of coverage of phenomena (i.e., the range of function attributions which the theory can make) and explanatory depth (i.e., the depth of what it means to attribute a function)." Etiological theories such as proper function prioritized depth over breadth, we saw, and were found wanting partly as a result. We can now see clearly that systemic function theory does the reverse, prioritizing breadth over depth. Systemic function can attribute any intuitively appealing function imaginable, if only at a paper-thin level of explanatory depth – but also, problematically, many more less intuitively desirable functions besides.

This analysis seems to indicate that it would be preferable to find a way to retain some central teleological aspects of function attribution while finding a smarter way to address the problems they incur. We next turn to the contributions approach, beginning with the most established examples in the literature, and moving to a novel, elaborated version as this candidate.

CHAPTER FOUR

THE CONTRIBUTIONS APPROACH TO FUNCTION

**Introduction**

**Finding a Place for Teleology**

By this point in the analysis, it is clear that an important task of function theory –

perhaps its central task – is to coherently articulate the teleological dimension of the

concept of function. The key difference between the two theoretical approaches explored

so far (the etiological and the systemic effects) was seen to be their treatment of

teleology. Etiological theories consider teleology to be so central to the concept of

function that they require each functional part to have its own teleological origin story

(i.e., the part having come about due to its ability to perform its function). On the other

hand, systemic effects accounts, which attribute function based on the causal role a part

has in producing some more general systemic capacity, exclude on principle any

conception of teleology whatsoever.

Each of these extreme approaches to teleology in function was seen to incur

serious negative consequences. Etiological theories were found to be riddled with

difficulties resulting from their rigid insistence that the origin of each and every

functional part be explicable through reference to its function. The systemic effects

account, by contrast, was found liable to attribute function to nearly anything that could

be linked to a systemic capacity, even if the relevant system seemed arbitrarily defined

and without purpose; at the same time, it could not attribute the robust normative function required in clearly teleological systems such as artifacts.

In searching for a philosophical account of function that is judicious and consistent in attributing function, comports with our intuition that function somehow has to do with what a thing is "for", and applies equally to artifacts and natural systems, a third way seems required. This third way would seem to need to allow teleology to inform function attribution without predicating all function directly on teleological origin. In the previous chapter, such an alternative made a brief appearance: the contributions account. Cummins noted the existence of such approaches only to dismiss them (although in the end his systemic effects account, as reinterpreted by Davies, ironically came to closely resemble a contributions account in important respects). But there the nature of a contributions account was only broadly sketched. To see whether such a theory holds promise as a middle way, we must study a fully-developed example.

Contributions accounts are somewhat uncommon in the function literature. A relatively primitive version of a contributions approach revolving around "usefulness" was outlined for natural systems by John Canfield in 1964, while Morton Beckner expanded this to include artifacts in 1969 (Canfield 1964, Beckner 1969). More recently, Tim Lewens sketched a rough account he calls the Naïve Fitness (NF) theory, which is as follows: "the function of a trait t is F iff traits of type T, of which t is a token, make a significant contribution to fitness by performing F." (Lewens 2005, 102). However, Lewens' scope of concern is strictly limited to evolutionary biology. Therefore, we will

study in detail what is likely the most prominent and well-defended contributions

account: Christopher Boorse's Generalized Goal Contribution theory, or GGC.

In GGC, function consists in contributions to the goal of any goal-seeking system.

GGC is designed for efficiency and simplicity: any effect that contributes to the goal of a

system which has them can thereby be said to have function. This applies whether the

function was intended or not, and whether the system is natural or artificial. The only

complexity Boorse allows in this schema is whether the function is stronger or weaker,

corresponding roughly to how frequently such a function is performed by an item.

Obviously, such an approach relies on having goal-directed systems in order for there to

be contributions within them. For this, Boorse appeals to cybernetic accounts applying

equally to natural and artificial systems. While a contributions approach will be seen to

have much to recommend it in principle, problems with such a stripped-down version of

a contributions account will also be encountered.

Following the exposition of the GGC theory, we will engage in a critique focusing

on three main areas of weakness. First, it has a problematic relationship with intention in

artifacts resulting from its conception of function as depending on use only, without any

consideration for design intent or object properties. Second, and relatedly, it will be seen

to have profound problems in how it recruits external objects into teleological systems.

Finally, any contributions account must provide an account of the goal-directed systems

within which functions are found. GGC seems to suffer from a lack of a successful and

consistent account of goal directedness; a cybernetic account is recommended by Boorse,

but it is not clear that this kind of account is up to the job, or whether, indeed, he is even truly abiding by one.

Certain of these weaknesses will lead us to revise the GGC account in substantial ways. This analysis begins with a fresh attempt to reevaluate the relationships between etiology, systemic effect and contributions to a goal through an accumulating series of scenarios involving artifacts. This analysis will lead us to a crucial insight: the introduction of a distinction between the performance of a function *by* an item and the assignment of the function *to* the item. While function performance will be seen to consist in contribution to a goal, the assignment of such a function to an item will be seen to occur through knowledge of design etiology, selection history, or typological preponderance. Any or all of these three modes provides reasonable expectation of function performance from a token item; this expectation is in turn what grounds the assignment of the function to the item. This account of function assignment also makes possible an account of malfunction, something that was not fully possible under GGC. Malfunction, I will argue, is simply the disappointment of that expectation for function once it is assigned to the item. A psychological explanation of the normative dimension of malfunction based on the projection of the function expectation onto the item itself will also be offered. Finally, a superior account of system constitution will be offered, one based on the explicit top-level intention behind the use of an artifact or other external object. This heavily revised contribution account will be referred to as the Assigned Contribution theory.

These revisions leave only one major issue in GGC unanswered: attributing teleology to natural systems such as organisms. This, in turn, will propel us to the final set of considerations in chapters 5 and 6.

## Boorse's Generalized Goal Contribution Theory

**Generalized Goal Contribution**

Boorse's basic idea is simple: function is contribution to the goals of a system. He is led to this concept based on the observation that in "every context where functional talk is appropriate, one has also to do with the goals of some goal-directed system." (Boorse 1976, 77). We have already observed how this seemed true when Davies' willingness to attribute function to the ionic bonds in the dissolving salt crystal clashed with our intuitions of what kinds of effects deserve to be called functions. This intuition was also the basis for Millikan's original criticism that systemic effects theory would be forced to attribute functions to clouds. This was a criticism in her mind precisely because clouds are, we take it, *not* goal-oriented systems; they produce rain, but not because that is what they are for. Thus, part of the difficulty of systemic effects theory – the reason why it requires us to adjust our concept of function if we are to accept it – is its refusal to recognize teleological contexts as vital for function. For Boorse, by contrast, such a context is what makes possible any function attribution in principle: "Any goal pursued or intended by a goal-directed system may serve to generate a function statement. Functions are, purely and simply, contributions to goals." (Boorse 1976, 77).

Goals are able to "generate" function statements in precisely the same way that systemic capacities generated them in systemic effects theory: by attributing function to the lower-level effects that help to bring them about. Carburetors have the function of mixing air and fuel not simply because that is what they were designed to do, as Wright and Millikan would have it, and not simply because that is the causal contribution they in fact make to the systemic capacity of the engine to make mechanical power, as Cummins and Davies would have it. For Boorse, they have that function because that is the causal contribution to the *goal* of the engine as it is used: producing mechanical power. The logic is analogous for natural systems. Hearts have the function of pumping blood not because that is what they were selected to do, as Wright and Millikan would have it, and not simply because that is the causal contribution they in fact make to the systemic capacity of the organisms to respire, as Cummins and Davies would have it. For Boorse, they have that function because that is the causal contribution they make to the *goal* of the organism to stay alive. Like systemic effects theory, the contribution at issue here is in terms of efficient cause. Functional contributions are those effects that causally enable the attainment of the system's goal (or, as appropriate, at least its goal-seeking behavior). Yet unlike in systemic effects theory, the effect being contributed to is not simply a hierarchically-produced system-wide capacity – it is the (or a) *goal* of the system. Because it is about the furtherance of goals, function in GGC is expressly teleological – a distinction that has major implications.

First, by hypothesis the only systems in which GGC attributes function are those that have goals. Thus, GGC does not seem to suffer from the same kind of promiscuity problem seen in systemic effects theory. Clouds and salt crystals on their own are simply not proper systems for function attribution in GGC because in themselves they lack goals. Organisms, on the other hand, are goal-directed systems and as such are rife with functions, from their internal physiology to their behavior. Artifacts are goal-directed as well, insofar as human beings supply intentions relative to their use. GGC is thus able to both attribute functions with more discretion than a systemic approach, and also to fulfill our intuitions that function has to do with purpose. Moreover, involving teleology in function attribution seems to at least provides the possibility of an account of malfunction – something, as we saw, systemic effects could not do. A contributions account like GGC thus seems a promising middle ground, answering many of the more serious difficulties of the non-teleological systemic effects account without, perhaps, incurring the overreach of the etiological account.

Yet while a contributions account provides a promising schema for incorporating teleology in function, it does raise some profound new questions even while it addresses many of the old ones. Two questions stand out as most obvious to begin with: First, what counts as a contribution? And, second, what qualifies a system, whether natural or artificial, as goal-directed, and why?

**What is a Contribution?**

We have established that the functional contributions in GGC theory are essentially causal, but this does not entirely clarify their nature. Under what circumstances and how reliably must a contributory effect occur to count as a function? Must they be consistently performed? Must they be intended or selected, or can they be wholly accidental? And must the effects be necessary for producing the goal to count as function?

In short, according to Boorse, any effect which advances a goal within a goal-directed system – even just once, accidentally – counts as function in GGC. Indeed, he explicitly notes that "functions may be performed only once and by accident" (Boorse 1976, 80). Here, by way of example, he offers the story of a soldier who is shot in his pocket Bible, which, by stopping the bullet, saves his life. This effect was a freak, unexpected occurrence. It seems unlikely to recur. And, it was not intended either by the publisher in producing the book, or by the soldier in carrying it. However, preserving his life was most certainly a goal of the soldier.[1] For that reason, and because the Bible contributed to this goal on this one occasion, Boorse considers it to have performed a function in stopping the bullet.

Whereas for Wright one-time accidental incidents could not be functions precisely because they occurred unintentionally, Boorse sees no principled difference between accidental effects and effects that were intended or selected; so long as the effect contributes to a systemic goal, it is a function. This neatly sidesteps difficulties

---

[1] We will consider how his carrying of the Bible related to that goal when GGC is critiqued in section II.

encountered in Millikan's proper function such as the paradox of a function needing to

exist before it could be selected and thus count as a function. Yet this is not to say that

all function attribution is precisely equal in GGC. Boorse establishes a sliding scale of

function, from weak to strong, based, very roughly, on how routinely the contribution is

made. An accidental function, like the Bible stopping a bullet, is considered a 'weak'

function due to its rarity. Yet the only thing separating it from strong function is the

frequency of its occurrence.

Judging the strength of a function turns out to be fairly complex – more complex,

perhaps, than Boorse allows. Roughly, function frequency may be judged in two ways:

either by what is typical within a given token system, or by what is typical for the system

type. In the Bible example, we can presume that the function performance is unusual

enough that it is not typical for this individual soldier, much less for soldiers and Bibles

in general. That is, of course, what makes the story notable at all. This explains why this

function performance, under these conditions, is considered "the weakest of all functional

attributions" (Boorse 1976, 80). Yet in other examples Boorse offers, we see cases where

a token system has a function routinely performed, yet by an element which does not

perform it in other systems of its type. For example, one of the hypothetical

function/accident cases Wright introduced was an assembly line conveyor belt that was

unsettled by an earthquake, and thereby developed a ripple that somehow passed only

'good' transistors (Wright 1973, 165). In this one facility, at this one point in time, what

was merely a conveyance became, in effect, a quality-control sorter.

Wright, of course, raised the examples in order to point out that this is still not a function of the conveyor, since that is not what it was designed to do (though he allowed that we might say it "functioned as" a sorter). Yet because this effect contributes to the goal of the overall line, Boorse *does* consider this to be a function the conveyor performs. However, since this is not something that other conveyors of its type do, he agrees with Wright that calling it *the* function of the conveyor would be going too far – even though the ripple induced in *this* given conveyor belt contributes this functional effect with complete regularity. Therefore, Boorse considers sorting to be a weak function of the line, though perhaps a stronger one than the one-time bullet-stopping Bible.

The strength of a function, then, is a judgment call related to frequency of function performance over time, typicality over type, and various pertinent contextual factors. As evidence of the ambiguity and complexity of this judgment, it seems that Boorse himself is not entirely consistent about what constitutes strong function. In the assembly line example, he readily admits that the sorting effect is not "the" function of the line, which implies that it a weak function. Yet Schaffner raised a structurally similar example of a human born with a fully functional bursa of Fabricius, an organ in birds which protects against viral infection, due to a radioactivity-induced mutation (Schaffner 1993, 388). Even though no other human has such an organ, Boorse states that "a strong function statement could be appropriate" in such a case, because the organ reliably produces the functional effect of viral immunity (Boorse 2002, 87). Why in this case the function would be strong while in the assembly-line case the function remains weak,

when both features were accidentally generated, both routinely contribute to the goal of the system, and both are anomalous for their type, is not explained.

Ultimately, for Boorse, the "difference between weak and strong function statements seems to be simply a loose matter of how often or consistently the function gets performed" (Boorse 2002, 71). In general, "[w]hat converts a function X performs into "the function of X" is our background interests in the context in which the function statement is made." (Boorse 1976, 81). Yet, he cautions that "[t]here are no rules for converting functions performed into functions possessed." (Boorse 2002, 71). Not even regularity of contribution, or comparison within the system type, can conclusively establish a function as one that the item merely performs, or a function that belongs to the item.

Yet this ambiguity is easily tolerated, for, philosophically, Boorse claims that not terribly much separates the two. Indeed, "the distinction between weak and strong function statements is illusory to this extent: there is no important conceptual constituent of the idea of "the function" or "a function" which is missing in "performing the function."" (Boorse 1976, 81). Function is about contribution to the (or a) goal of a system; whatever does this, and whenever this occurs, a function has been performed. In many cases, the contribution is so frequent or regular or somehow significant that we consider it "the" function of the thing that makes the contribution. Notably, this may be the case even if the function performance was essentially accidental each time it was performed by the item. To return to the Bible example, Boorse asks, "if a battle-weary

veteran looks back over a long career in which he was shot thirteen times in the pocket Bible, can he truly say that stopping bullets has been one of the functions of the book?" (Boorse 1976, 81). Even if the soldier never intended this, Boorse thinks that he may justifiably claim that was its function. Indeed, according to Boorse, on the first such occurrence the only reason that it "sounds false to say: "This Bible performs the function of stopping bullets," [is that] the simple present [tense] suggests that the function gets performed more often than it has been." (Boorse 1976, 80).

Function itself is a clear principle for Boorse: contribution to a systemic goal. Yet the "possession" and "strength" of the function is a judgment, one often subject to vague appraisals of the regularity, typicality and significance of the connection between the item and its functional contribution. At the weakest end of the spectrum is a function that is merely performed, but that cannot be called the function of the item. At the stronger end is a function that the item possesses, almost in the sense that Millikan's proper function attributed functions to items. This is the case in particular when not only does an item routinely perform a function in an individual token, but a token of its type routinely performs that function in systems of the same type. To take our stock example, "it is not only the function of [a] token heart to pump blood, but also the normal function of that organ type in the human species" (Boorse 2002, 87). Thus, we can comfortably say that pumping blood is *the* function of the heart in humans. Ultimately, then, "the contrast between weak and strong function statements is solely an effect of the English articles "a" and "the."" (Boorse 1976, 80).

It should also be noted that even in cases where a function is attributed as the function of an item, this does not imply that the function is exclusive in either direction – that is, that this is the only function that the item has, or that this item is the only thing that performs this function. An item may have more than one function, and it may be the case that more than one item in the system is capable of performing the function. Referring to a function as the function of a thing is contextual to the relevant systemic goals at a given point in time. This allowance is particularly important when a complex system may have multiple goals it seeks at various times (as is the case with the behavior of higher animals). The full set of conditions Boorse sets forth are as follows: ""A function of X is Z" means that in some contextually definite goal-directed system S, during some contextually definite time interval t, the Z-ing of X falls within some contextually circumscribed class of functions being performed by X during t-that is, causal contributions to a goal G of S." (Boorse 1976, 82).

This leaves just one last ambiguity to address in the nature of causal contribution: Must the contribution be necessary for the goal? According to Boorse, it need not be: "The pumping of the heart may be a contributory cause to the circulation of the blood without being essential to it, since circulation can occur by artificial means. But … we are clear enough what it means to say that the heart is helping to cause the circulation of the blood, even if a heart-lung machine is ready to switch on at a moment's notice." (Boorse 1976, 78). The contribution that GGC looks for is any active aid to the system, whether essential or not, and counterfactuals notwithstanding.

**Natural Goal-Directed Systems**

We now turn to the more difficult questions surrounding goal-directed systems, focusing on natural systems first and artifacts in the next section. How do we know which systems have goals and which do not? What makes a system goal-directed? First, consider the structural relationship between function as goal-contribution and the elucidation of goal-seeking systems. While the latter is undoubtedly a prerequisite for the former, they remain formally independent. In other words, while Boorse propounds a theory of function as goal-contribution, he claims that this alone does not commit him to any particular account of goal-seeking. Given a goal, GGC allows us to attribute function; but it is no part of GGC itself to attribute goals. That task can be fulfilled in a number of ways. As a result (and assuming that intentionally-sought human goals are unproblematic), for the purposes of GGC as a theory of function, "the notion of goal-directed behavior, when applied outside the realm of intentional action, looks more like a theoretical concept of biology to be explicated according to convenience." (Boorse 1976, 78). Thus, whatever the account of natural system goal-seeking we choose to employ, function as contributions to goals remains secure. The unspoken exception, obviously, is if there is no satisfactory account of goal-seeking in natural systems at all because natural systems (humans and higher animals excepting, perhaps) have no goals. This would certainly be Davies' and Cummins' contention, and this remains the core of the difference between the two approaches.

In any case, the account of goal-seeking which Boorse endorses is essentially cybernetic. In a cybernetic account, a system is considered to be goal-seeking if and only if its behavior can be characterized as being in apparent pursuit of a goal. This means that not only is the system's behavior compatible with attaining a given goal, but that when the conditions for attaining the goal change, the system adjusts in response. More specifically, "To say that an action or process A is directed to the goal G is to say not only that A is what is required for G, but also that within some range of environmental variation A would have been modified in whatever way was required for G." (Boorse 1976, 78).[2] The paradigmatic example of cybernetic goal-seeking is the common thermostat. It attempts to regulate the temperature, adding or withholding heat as required in pursuit of the set temperature. But the approach can be generalized not only to homeostatic physiological systems, but to the goal-seeking behavior of organisms. "For example, 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" (Boorse 1976, 78-79).

Yet the inability of such a simplified account of goal-seeking behavior to account for all cases of ordinary goal-seeking are well known. For example, pursuit of one goal may be interrupted by the opportunity to more easily acquire another, incompatible goal – yet this does not mean that the first goal was never sought. A system may also have a goal but lack the means to pursue it; it would act if the chance arose, but perhaps it never

---

[2] Boorse bases his version of cybernetic goal analysis on the work of Gerd Sommerhoff, who held that "on the phenomenal level … life is nothing if not just this manifestation of apparent purposiveness", namely its "end-serving and integrating activities" (Boorse 2002, 76).

does.  In short, there are various ways in which a system may have a goal, but not give the kind of overt behavioral evidence that a cybernetic account demands.  For example, a cat may stalk a mouse only to be presented with a bowl of cream, whereupon the original goal is abandoned.  While the cream is equally a goal for the cat, its observed behavior was not consistent with seeking it; meanwhile its behavior to that point indicated a goal it was not ultimately committed to.  It thus seems perilous to posit functions relative to any singular specific goal of organisms based only on their observed behaviors.  As Boorse admits, "there is no way of finding a unique goal in relation to which traits of organisms have function" (Boorse 1976, 79).

There are other problems as well; systems can malfunction, and seek the wrong target.  Nagel mentioned the homeostasis that the kidneys strive for in terms of water content in the blood as an example of a cybernetically observable goal.  But, as Melander pointed out, if the kidneys began targeting the wrong concentrations – concentrations that would kill the organism – a cybernetic analysis would seem forced to embrace this new target as their goal, given their actual behavior (Melander 1997).  A cybernetic analysis alone, in other words, cannot distinguish function from malfunction.  To fix this, we must appeal to higher-level goals towards which the kidneys themselves contribute.  After all, if the malfunction threatens the animal's life, it can hardly be a contribution function on that level.  The cybernetically observable phenomena, in particular of the organism's internal physiology, must thus "terminate in the highest level goals of the organism" (Boorse 2002, 76).

Both of these difficulties push Boorse to account teleologically for the ultimate goals of organisms. Boorse, endorsing Sommerhoff, says that in general, "organisms are centers of activity which is objectively directed at various goals-for example, survival and reproduction" (Boorse 1976, 79). We appeal to such apical organism goals particularly in attributing internal physiological functions: "In physiology the goal-directed system S is the individual organism and the relevant goals its own survival and reproduction. Whatever contributes to these goals reliably, throughout a species or other reference class, is assigned a physiological function" (Boorse 1976, 84). These, after all, are what we normally think of as functions in organisms, and for good reason. If indeed "life is just a kind of natural goal-directedness" as Boorse and Sommerhoff take it to be, then "it is not arbitrary to take life's continuation, individual survival and reproduction, as the apical goal. And what a specific organism's next-highest goals are – that is, what activities its life consists in – is, on the cybernetic analysis, a fact." (Boorse 2002, 76).

But Boorse does not limit organisms to only their own life-goals, and postulates higher-level goals that transcend the organisms, such that traits and individual organisms may have functions within ecosystems: "the individual bird is only one system S. At the level of a population or species, new goal-directed processes seem to appear … and one can ascend still higher by viewing a whole ecosystem as a teleological unit." (Boorse 1976, 83). He considers it a strength of GGC theory that it is able to adapt itself to such a wide range of contexts. All these goals can exist simultaneously: the goal of catching a mouse, the goal of attaining cream, even the higher-level goals of perpetuating the

species, or maintaining a trait.  When we attribute function in GGC, the goal against which the function is attributed will be determined by the context of our analysis.  Thus, "there is no sense in asking which goal G is the goal at which [a given organism's] behavior aims, and in respect to which its parts have functions, except in so far as this goal is clear from the context of discourse." (Boorse 1976, 83).

**Artifactual Goal-Directed Systems**

In terms of goal-directedness, two clear differences between artifacts and organisms emerge.  First, unlike organisms, most artifacts do not function autonomously, which means that a cybernetic analysis of their goal-directedness as self-contained systems is rarely appropriate.  Second, artifacts typically involve a human operator in their use.  These two differences happen to be complementary; although most artifacts would not be found to be goal-seeking on their own on a cybernetic analysis, we can appeal to human intentions in their use for goal-directedness.  Of course, a small number of artifacts are paradigms of autonomous goal-seeking, such as heat-seeking missiles and thermostats.  On a cybernetic analysis, these artifacts have their own internal goals that they seek independently.  Yet most artifacts are not "goal-directed in and of themselves", as Boorse puts it (Boorse 1976, 79).  Most are inert and require humans to operate or use them, like chairs, pens, cars and screwdrivers.

However, in order for a contributions account to include human goal-directedness in the use of these objects, their human users must be included as part of the system.  The system as a whole must be goal-directed for a contributions approach to apply.

Therefore, in these more typical cases of artifacts, we attribute systemic goals, and thus functions, "by taking the artifact together with its purposive human user as a goal-directed system.  Chairs have functions because they contribute to the goal-directed human activity of sitting, fountain pens because they contribute to the goal-directed human activity of writing. Such objects, which lack the appropriate organization to be independent centers of teleology, must inherit their functional features from people's use of them." (Boorse 1976, 79-80).

This devolution of the teleology of goal-directedness from the users' intentions to the artifact is actually representative of how teleology is always propagated in a contributions approach.  In general, a contributions approach "propagates" the teleology from the top level of the overall system to the contributing system components.  Just as we treat the organism as a system as directed to the goals of survival and reproduction, and credit the function of its parts, traits and behaviors that contribute to these goals, the goals of the user inform the function of the artifact.  Thus, "the role of intention is merely to determine the over-all goal of the use of the artifact. Once that is established, the function of a part is its actual contribution to this goal" (Boorse 1976, 80).

It is, of course, natural to appeal to intentions to explain artifact functions.  We saw Wright rely directly on it, albeit in terms of the intentions of the designers of the artifact.   And the inevitability of involving intentions is one reason why systemic effects theory had difficulty engaging with artifact function at all.  Yet only a contributions

approach is forced to recruit the artifact user as part of the analyzed system. This significantly complicates the definition of the system, as we will see.

It may seem that constituting systems that include conscious users and their intentions opens up a schism between artifacts and organisms in how function is attributed to each. Yet here we must recall the formal independence described above between GGC as a theory of function, and the method of goal attribution for the relevant systems. According to Boorse, "As long as artifact functions are determined by human purposes, I see no disunity in the analysis. Human purposive behavior is but one example of goal-directedness. For the analysis to be unified, there is no reason why all goals must be determined in the same way." (Boorse 2002, 76).

If non-autonomous artifacts must include their human users in order to be included in a goal-directed system, are there analogous systems consisting of other animals and external objects? For example, does a rabbit hiding in a bush convey function to the bush with respect to the rabbit's goal of evading predators?[3] The logic of GGC to this point would strongly suggest that it does, and indeed Boorse endorses this conclusion – provided that the animal and the external object are conjoined frequently enough that the goal contributions they make qualify as 'regular'. "I suggest, then, that external objects have functions in relation to goal directed systems, such as organisms, when they make regular causal contributions to their goals" (Boorse 2002, 82). Even if it was considered to be regular in the case of a given rabbit and a given bush, such that the bush has the function of protecting the rabbit, this function is not, of course, exclusive to

---

[3] This example was raised as an objection to GGC by Lowell Nissen in (Nissen 1997).

the bush – many other organisms may use the same bush to advance their own ends simultaneously, and the same rabbit may hide in many other bushes.  But the fact that GGC attributes function to external objects even when used by non-human animals is notable.  While we might consider a rabbit to be minimally conscious, this is not essential to the function attribution – the logic holds even when the organisms involved are insects, as Boorse explicitly indicates.  The logic of GGC is clear: anything that regularly contributes to the goals of a goal-directed system (which all living organisms are) can be attributed function in relation to it.

This means that non-teleological systems, such as clouds, might actually be attributed function in relation to goal-directed organisms – even plants.  After all, rain makes quite regular contributions to the life goals of plant life.  Thus, Millikan's objection that Cummins' systemic effects theory is liable to attribute the function of rain to clouds could also be made against GGC.  This is ironic since one of the expected advantages of contributions approaches in general is that they limit themselves to teleological systems, thereby sidestepping the promiscuity problem entirely; Boorse indeed claims that GGC "is not, like [Cummins'] function, hopelessly overbroad" (Boorse 2002, 73).  Yet Boorse insists that this degree of broadness is not problematic, just so long as we are explicit about what this really means: "there is no harm in saying that rain has a function in the lives of plants … as long as we are clear that these functions have no tendency to explain … the rain" (Boorse 2002, 83).  Whether or not this is problematic will be considered closely when we critique the theory.

**GGC and Malfunction**

As may be recalled, one of the major aspects of function attribution that is lost when teleology is excluded from a theory of function, as with systemic effects theory, is the possibility of attributing malfunction. As we saw, allowing this possibility was in fact the primary motivation for the development of Millikan's proper function, and accounts for its strongly teleological nature. Can the more indirectly teleological nature of GGC account for malfunction?

For reference, compare Wright's description of the windshield washer switch that, due to a manufacturing defect, fails to activate the washers according to the intentions of the designer. For Wright, the fact that those intentions brought the switch into being (or at least into the system) means that they alone determine its function – a function it thus has regardless of its success in actually performing it. Because parts have their functions etiologically, should they not perform it, they would still *have* the function they are not *performing*. This is the classic definition of malfunction: a thing not (or not adequately) performing what is nonetheless its function.

Yet in GGC, and in contributions approaches generally, function attribution is made not on the basis of origin, but on actual contributions to a systemic goal. This presents a quandary in terms of malfunction: it is not possible for a part in GGC to have a function but not perform it, because functions are had *via* performance. To be attributed function, a contribution must be in the making; yet malfunction would imply that this

does not occur. Though he is never explicit, it is likely for this reason that Boorse avoids using the term "malfunction".

Nonetheless, he does allow that "The fact that a part may fail to perform its function deserves some explanation" (Boorse 1976, 83, note 14). According to Boorse, in the case of the disconnected washer button, "activating the washer is not the actual function of this token button; it has none until the button is fixed" (Boorse 2002, 89). After all, if intentions alone were enough to confer function, Boorse argues, then failed designs, no matter how badly they fail, all have their intended function. But, referring to early unsuccessful attempts at airplanes, he notes that "Objects that cannot fly do not have the function of flight. They have only the intended function of flight, which means, precisely, that they are intended to have the function of flight but do not have it." (Boorse 2002, 89). For Boorse, then, functions are either performed, or not performed – but there can be no such thing as malfunction per se.

But he remains sympathetic to the intuitive claim that the button is not doing something that it should; that it is somehow 'broken'. The question is: if such a claim cannot be made on the basis of intention, on what other basis can it be made? Obviously, the lack of function performance is far more interesting and relevant in some cases than others. The Bible that fails to protect its bearer from bullets is far less remarkable than an armored vest that fails to. Besides intention, what is the difference in such cases? A clue is provided by Boorse's best attempt at reinterpreting the malfunction claim of the washer switch. A malfunction attribution of the switch "might mean that activating the washer is

the function performed by this (kind of) switch in most cars of this model." (Boorse 1976, 83, note 14). This could certainly have caused us to expect the unperformed function to have been performed; and as Davies demonstrated, expectation is an integral part of the concept of malfunction. Generalizing, this would mean that the significant variety of non-performed functions are exemplified by "a function typical in the trait type which does not exist in the token." (Boorse 2002, 89). Even here, however, we must refrain from calling it a malfunction, and the normativity involved is strictly descriptive rather than prescriptive. Thus, we can say that a part is not doing what is typical for a part of its type; but we cannot quite say that it is not doing what it "should".

It may seem, from his rather piecemeal and cursory treatment of this issue and the desire to avoid malfunction claims, that Boorse considers unperformed functions relatively unimportant. In fact, nothing could be further from the truth. Indeed, the bulk of Boorse's philosophical output over a long career concerns theories of health and disease (Boorse 1977, 2011). In his own extensive theory of biological disease, which he calls Bio-Statistical Theory or BST, roughly, a part is considered diseased if its function performance falls below the range that is statistically normal for the appropriate system class (generally determined by species, sex and age). Such a typological approach may seem to invite the circular classification problem we encountered in our analysis of proper function's normativity. Yet Boorse is not locked in, as Milliken was, to defining the trait type by its function. Traits such as organs can be defined morphologically, and compared functionally, without any risk of circularity or class membership confusion.

Indeed, he points out that in the fields of anatomy and physiology, as they are actually practiced, classification is in fact made on a morphological basis.

## Critique of Boorse's Generalized Goal Contribution Theory

**Intentions and Goal Attribution**

While formally distinct, a contributions approach to function is, ultimately, parasitic on an account of goal-directed systems. Such accounts are likely to be quite different for artifacts and for organisms; yet while he appeals to the general outlines of a cybernetic account, Boorse never provides a fully fleshed out account of either case.

We know that intentions are central to the goal-directedness of artifacts; this is why Boorse is forced to bring the user into the goal-directed system. Yet we also know that certain kinds of intentions – namely design intentions – are not as relevant for GGC as we might have expected. Indeed, design intentions are relegated to secondary status, useful only as a resource for determining the contributions likely made by a given part of a given system. As Boorse says, we can easily imagine that, rather than asking about design intentions, "In asking our mechanic for the function of the carburetor, we [are] asking simply for whatever it standardly contributes to our goals in driving." (Boorse 1976, 83). Appeal to design intention is in fact merely a "convenient contextual device" (ibid.) for getting at salient contributions to normal use.

Yet while design intentions are not salient as such, other intentions are central and indispensable to establishing the goal of the artifact system, and thus for function attribution in GGC. After all, "the function of artifacts seems obviously determined by

human goals". These goals are their intended goals *in use*. This conception of "pure use-functions suggest that neither design nor selection is required for artifact function, as so many writers assume, but only an object's role in a human-goal-directed activity" (Boorse 2002, 68). Indeed, even "objects not artifacts at all can still acquire functions via people's purposeful use of them", such as the sharp stone used as a knife in chapter 2 (Boorse 2002, 68). GGC thus relies directly and exclusively on what the current user is actually doing, i.e., intending to accomplish, in order to attribute purpose to the overall user-artifact system, and, via that, function to the artifact(s) involved.

Yet consider some of the untoward implications latent in this approach. According to Boorse, the function of a thing is always only whatever it is used for. If my toolkit consists entirely of a single large screwdriver, which I therefore also use as a pry bar, a chisel, an awl, even a hammer – then those are among its functions. Unlike the weaker "functions as" construal, these are functions just as genuine as those each proper tool would have were it used instead.[4] Driving screws may not even be among the functions of this all-purpose tool if I never use it thus. Likewise, a pair of scissors still in their package languishing in a drawer have no function according to GGC, until and unless they are used for something. Or, to adapt an example that Boorse offers, if I use my phone only as a paperweight, then this is its function.[5] Yet surely such scenarios are

---

[4] The only distinction possible would lie in the strength of the function, which it might be argued is weaker in the case of the screwdriver-as-hammer, for example, due to the fact that other screwdrivers are not often used this way. But we have already seen how malleable and ambiguous function-strength judgments are in GGC theory. In isolation the screwdriver would seem to have the hammer function strongly so long as it is frequently used as such.

[5] Boorse asks, "is holding down papers one of the many functions of my telephone? Again the answer does not seem to depend on my intentions in acquiring the telephone, but in some vague way simply on how

ignoring something central to the nature of the artifacts involved. Why should the

screwdriver have a shaped tip and a round handle if its function is not somehow related to

this? Even more, why should a telephone have complex internal circuitry and a handset

with sound transducers? Is the mass of the phone the most salient of its feature from a

functional perspective?

Boorse might argue that the function of the phone as paperweight is merely a

weak function, given that more items like it are used as phones than as paperweights. But

what if this was not true? We can easily imagine a future civilization discovering phones

and not knowing what their function is (or was). Functions in use can change, of course.

But is there nothing about function which relates to either the etiology of the artifact or its

detailed properties? What does it mean that we would still recognize the object as a

"telephone" even if it was only used as a paperweight, and yet another object next to it

might be recognized only as an actual "paperweight"?

Moreover, not only are all design intentions irrelevant in GGC, but even use

intentions may be irrelevant in circumstances where contribution to a goal occurs without

them. That is, the user's intentions, their goals in their activity, need not even relate to or

include the artifact to be attributed function relative to these goals. Just so long as the

artifact does happen to contribute to these goals, it thereby has function. The main

problem Boorse takes himself to be solving by this arrangement is unknown function

mechanism, where an item contributes to a goal in an unknown or misunderstood way,

---

often I use it to do so." (Boorse 1976, 81). In my use of the example, I imagine that it is not used as a
phone at all, leaving only its function as a paperweight.

and so could not have been intended by design. As Boorse explains, "parts of artifacts may have functions wholly unknown to their makers. Many ancient mechanisms achieved their desired goals without being understood by the people who built them… As currently understood, the actual function of yeast in fermentation is to produce enzymes which catalyze the conversion of sugar to carbon dioxide and alcohol. Presumably, then, that has always been the function of yeast in brewing devices. It did not suddenly acquire this function with the advent of chemical theory. But brewers with no knowledge of enzymes cannot intend their yeast to produce them." (Boorse 1976, 73). By making only the final goal intentions relevant, and attributing function to all the elements which contribute to it, Boorse solves this problem neatly. And yet, as we saw most poignantly with the return of potentially functional rainclouds, and to a lesser degree with the bullet-stopping Bible, making design intentions entirely irrelevant for function attribution and counting only contributions to a goal compels the GGC account to attribute function rather liberally to what we would normally consider items unrelated, or at best tangentially related, to the goals at issue.

**Potential Functions**

Consider the action of the knock sensors in a modern engine. These sensors (in practice, essentially small microphones) detect the vibrations characteristic of misfires. When misfires occur, the engine computer receives these signals and retards spark timing in an attempt to attain more complete combustion at lower compression. The point is that if everything is working smoothly, a knock sensor does precisely nothing. It is only

when a misfire occurs that the sensor is triggered, and makes a contribution to the goal of

the system: producing smooth rotational power.  In a perfectly running engine, therefore,

according to a strict interpretation of contribution, the knock sensors have no actual

function.  But they do have a critical *potential* function.  Likewise, the classic biological

example of potential function is gamete cells, the vast majority of which do nothing, but

some small fraction of which are the indispensable basis for sexual reproduction.

This prompts the question of how GGC handles such "potential functions".  GGC

struggles with this because for it function is ultimately founded only on what contribution

an item currently makes to the goal of its system.  By definition, function performance is

about actual contribution.  But in cases of rarely performed, contingent function, no

contribution occurs in most cases.  Indeed, if that is the case often enough, such

functions, if they are attributed at all, would seem to be exceedingly 'weak', according to

Boorse's schema.  Yet surely ovum and knock sensors have their respective functions as

strongly as do carburetors and hearts, even if they do not have the opportunity to make

their unique contributions as often.  Yet GGC struggles to account for this.

As we saw, Boorse does allow that some functions may go unperformed due to

defects; this is when functions typical for the type are not present in a token.  Accounting

for functions that go unperformed for lack of opportunity is a somewhat similar problem.

Boorse therefore returns to his model of "species typical function on species-typical

occasions" (Boorse 2002, 93).  But what does it mean to say that the function or the

occasion is 'typical' for the type?  The typical spermatozoon does not, nor does it have

the occasion to, fertilize an egg; yet surely this is its function nonetheless. The best that GGC seems able to do is to say that this is its function *when it has any at all*: "fertilizing an ovum is the only function ever performed by a sperm … so [this] must be the normal function of [the] item if it has any" (Boorse 2002, 92). But this seems loose, unsatisfying and in conflict with routine function attribution.

**Cybernetics and Goal Attribution**

Given that the goals of human users are essential in constituting a goal-directed system involving artifacts, it is unclear whether, as Boorse implies, a pure cybernetic account really applies to artifacts, or whether user intentions factor in directly, obviating a cybernetic analysis. After all, intentions are not observable as such, and so do not seem amenable to a cybernetic analysis. In analyzing artifacts as used by humans it may be quite difficult to tell, though observation alone, what the goal being sought is, or even when it has been reached. As with the apical goals of organisms, it is unclear whether Boorse stays true to his claim of relying solely on cybernetic analyses to establish goal-seeking for GGC.

More generally, can a strictly cybernetic analysis, which looks only at actions in response to conditions, really detect or reflect all the goals determined by intentions? A cybernetic analysis can only attribute goals by induction. Furthermore, it can only detect goal-seeking through reactions to difficulties encountered in their pursuit. If we presume that a cybernetic analysis is passive, and does not actively intervene to present such difficulties, it may be of little use. If no difficulties are encountered, then a cybernetic

analysis might not be able to reach a firm conclusion, for if the goal is attained smoothly and immediately, might not look like goal *seeking* behavior at all to a cybernetic approach. How would a strictly cybernetic analysis determine that a goal was in fact attained as opposed to being abandoned? How can it recognize what ultimate goal is being sought, as opposed to necessary intermediate goals?

A cybernetic approach also seems to presume a repertoire of responses that constitute a valid path to the goal that may not exist. For example, we can assume that plants, as living organisms, have the basic goals of survival and reproduction. They even have a limited repertoire of responses that a cybernetic analysis would recognize, such as heliotropism, as Boorse recognizes (Boorse 2002, 68). But if a plant is placed in a dark room, how could any response to this frustration of its goals be detected? Does this invalidate the concept of it having a goal? If so, there are many comparable situations where even human goals would be cybernetically invisible. Boorse acknowledges something similar in what he refers to as the "missing object" problem, where a system cannot attain its goal because the goal does not exist; but he merely notes in response that a concept of the "internal representation of the goal state" may offer a solution (Boorse 2002, 69). While this would clearly help, it also seems to directly contradict the founding principle of cybernetics, which is that goal-seeking *status* consists entirely in observable goal-seeking *behavior*. Moreover, while conscious intention is an obvious example of the internal representation of the goal state", it is hard to see what the correlate would be in a plant or other non-conscious organism. As Nissen notes of such attempts, "there is a

remarkable looseness and uncertainty in identifying the component or property doing the representing … It is equally difficult to determine what is represented." (Nissen 1997, 66).

For all these reasons, a cybernetic analysis alone seems inadequate to reliably attribute goal-seeking status to a system, and seems liable to miss much that is otherwise considered teleological.  Equally serious questions exist around the constitution of the goal-seeking system.

**External Objects and Goal-Seeking System Constitution**

We know that the goal-seeking system to which an artifact belongs must include the human user, unless the artifact seeks goals autonomously.  And we know that that use need not conform to the artifact's design, nor be routine or directly intended to qualify as function.  Perhaps the paradigm example of this was offered above in the case of the bullet-stopping Bible.  Yet a closer study of this example reveals some fundamental difficulties in how systems involving users and external objects are constituted, beyond the liberality noted above.  The problem is, at what point does the teleological system involving the object begin to exist?  This is important because an inert object only participates in goal-seeking through connection to another system that is goal-seeking on its own (like a person). If it's not yet a part of that overall system, it can't make contributions to it.  And if it becomes part of the system only *by* contributing to it, this is circular and leads to promiscuity. [6]

---

[6] It is important to note that in his much later "Rebuttal on Function", Boorse entertains the possibility that external objects need not actually be considered *part* of a goal-directed system in order to have function

As we saw, if a Bible in a soldier's vest pocket saves his life by stopping a bullet, according to Boorse, "the system S is soldier plus Bible, and bullet-stopping contributes to its goals by saving his life." (Boorse 1976, 80). The Bible therefore performed a function, we are told. But just how was the system which had this as is "its goal" constituted? Clearly, not getting shot was a genuine goal of the soldier, independently of him carrying a Bible. But the relevant goal-seeking system, we were just told, is the soldier *and* his Bible.

A plausible interpretation of the origin of this goal-seeking system is that it was constituted when the soldier intentionally strategically placed the bible in that pocket *as armor*. In that case, function of some strength can be unproblematically attributed; the Bible is fulfilling its user's intention for its use. But we know that this is not the case here because the example was offered as a case of a function being performed "once and *by accident*" (Boorse 1976, 80, emphasis added). It is accidental just because the effect was unintended, and unforeseen even by the soldier himself. The soldier-Bible system with the goal of bullet-stopping, therefore, does not seem to predate the identification of the goal, but to follow from it.[7] It is both ad hoc *and* retroactive. Because the Bible protected the soldier, it is inducted into a goal-seeking system, whose goal was the protection of the soldier. It became part of a system that did not previously exist; or, at best, it joined a system to which it did not previously belong, but which did already have

---

attributed to them via their contribution to one (Boorse 2002, 82). However, such an arrangement seems only to exacerbate the problems considered here regarding the promiscuity of function attribution, rather than ameliorate them. Moreover, it is not clear whether Boorse himself is committed to this revision. Therefore, I do not pursue this defense further.

[7] There are echoes here of Davies' reverse identification of systems according to their systemic capacities.

this as its goal. This can be seen more clearly by considering a parallel example where the protecting object would never have been considered in any way part of a system involving the solider until the fortuitous effect is observed. Consider, for instance, a rock outcropping which deflected the bullet rather than an item carried on the soldier's person. Boorse's logic applies to this just as much as to the Bible. By contributing to the soldier's goal of survival, the rock would have performed a function and been incorporated into a system with the soldier in order to satisfy the requirements for such an attribution under GGC. Boorse must say, in strict parallel with the Bible example, that the rock performed the function of bullet-stopping for the soldier. For that matter, a defective cartridge which misfired but which otherwise would have struck the soldier had it fired also seems to have performed a function and would be inducted into a system with him to justify this attribution. Indeed, anything and everything that has ever served to protect the soldier could be attributed function as they contributed to his goal of surviving. Something seems clearly wrong with this arrangement

**GGC Critique Summary**

GGC, because it is concerned only with contributions to goals, bases function attribution exclusively on actual events of contribution. This works well enough much of the time, since in most cases similar objects are used in similar ways to contribute to similar goals. This is particularly true for organic function, because unlike artifacts and other external objects, function in organisms (particularly internally) generally occurs without the complications and variations introduced by intentions. Perhaps ironically, so

long as they are allowed as goal-seeking systems, GGC is able to attribute function within natural organisms more efficiently than to artifacts.

But in artifacts, this approach fails to capture important aspects of function attribution. It fails to recognize what makes an object have the functions typically attributed to it even before they are performed. GGC does not seem to adequately embrace the intentional aspects of artifact function, including why objects were made and how that reflects what they are good for. Finally, GGC struggles to capture what is at stake in the intentional use of objects, no matter their origin or nature. Not only may these use intentions conflict with design intentions, but what might be called unintentional use, such as the bullet-stopping Bible or even rock, are counted equally with intentional use.

The only scale for function in GGC is frequency of goal contribution, either in token or type, yet there are no real guidelines for what makes a function weak or strong. This does not seem adequate to attribute function with sufficient precision or discretion.

Then there are the still-unsolved problems any contributions approach must confront in attributing overall teleology – goals – to both natural and artifact systems. With respect to organisms, the challenge is to rigorously attribute goals to them while remaining within the confines of naturalism. The problem with the solution Boorse has offered is that a cybernetic approach is fallible, and must ultimately appeal to apical goals of survival and reproduction. Yet Boorse offers no firm account of these goals, other than them being endemic to life itself. With respect to artifacts, the challenge is to not

only attribute goals to the artifact-user system, but to rigorously identify the goal-directed systems that have them so that they are not ad-hoc and post-hoc. As it is, systems involving artifacts seem to be constituted on an ad-hoc and indeed post-hoc basis in GGC, enabling dubious function attributions, and preventing other intuitive function attributions until they occur.

Yet GGC also demonstrates why a contributions approach nonetheless holds great promise. By involving teleology via contributions to higher-level goals rather than directly through the individual etiology of functional parts, a contributions account allows function to be about purpose without insisting on the impossibility that each and every functional element account for its presence in terms of that purpose. It is well worth seeing if the problems noted with GGC can be rectified in a revised version of a contributions approach.

## Revising the Contributions Approach

A number of important questions for a contributions-based theory of function have now been raised. How do we account for the attribution of function to objects based on what they were made for, or what they are good for, even if they are not being used? What intentions best determine the ends or goals of artifacts: design or use? Is there any difference between function by design and function by accident? Is there any relationship between etiology and contribution? Is malfunction anything more than the merely descriptive non-function that Boorse allows? Finally, what constitutes a teleological system and determines its goals?

The task now is to devise an improved contributions approach that can better address these issues. I will proceed by first constructing a series of scenarios based on a case introduced by Wright, intending for it to demonstrate that there can be no such thing as accidental function. We saw above that Boorse rejects this conclusion. But there is much to be learned about the relationship between intention and contribution by studying imaginative variations on this case. These scenarios will explore the complex relationships between effect, intention and contribution, allowing us to return to the issues in GGC with fresh insights. This study will serve to clarify why a contributions approach is in fact the most promising candidate of the three considered to date. More importantly, it will help guide us to a revised approach that will avoid many of the complications encountered above, such as the post-hoc teleological system constitution problem. It will also lead to the elucidation of a more flexible and sophisticated conception of what Boorse calls strong and weak function, and unpack what each implies for malfunction.

The scenarios will primarily focus on the notion of design in artifacts, and what this should mean, if anything, to a contributions approach. Although Wright predicated all artifact function on design intent, he never closely examined the notion of design itself. And while Boorse held that design is essentially irrelevant to function, he too never attempted to define design or understand what might be at stake in it. He simply assumed that whatever contributed had function, without regard for origin or intention; but is there really no difference between design and accident? While the case and its

variations are of artifacts, due to the need to explore the issues presented by intentions in GGC, the results can and will be generalized to natural systems as well.

**Wright's Loose Nut in the Engine**

Wright introduced an intuitively compelling example in order to argue against the basic intuition of contribution approaches to function, at the time as articulated by Canfield (Canfield 1964). Canfield's version revolved around what he called the "usefulness" of the functional effect for the system. To counter this idea, Wright conceived a scenario in which what a thing does is useful to the system, but its presence is entirely accidental.

A brief primer on the system under study will prove helpful, as it will be referred to repeatedly in what follows. As depicted in Figure 1, in a reciprocating internal combustion engine, separate spring-loaded valves regulate fuel-air mixture flow into, and exhaust flow out of, each cylinder. When not pressed down on from above, as during the combustion and compression strokes, the valve springs seal the valves tight against the top of the cylinder. During intake and exhaust strokes, a rocker arm presses down on the 'hat' of the appropriate valve to open it; the rocker arm 'rocks' on a central pivot like a see-saw. In an overhead cam design, the other end of the rocker arm is directly actuated by a camshaft lobe. The tolerance, or 'lash', between the rocker arm and the valve top (depicted as dimension 'A' in Figure 1) is critical. If 'A' is too large, the valve will be too forcefully impacted by the rocker arm, creating noise and wear; the valve may even fail to open completely, creating power loss and misfires. If 'A' is too small, the valve

may be continually depressed, and so kept slightly open even when it should be seated, causing rough running, power loss, misfires, and burnt valves.  Valve lash is thus a critical parameter for engine performance.  To achieve the correct lash, an adjustment screw is provided at the end of the rocker arm (Item '2' in Figure 1).  The proper value for valve lash, 'A', is determined by the engineers and published as part of the specifications for the engine.
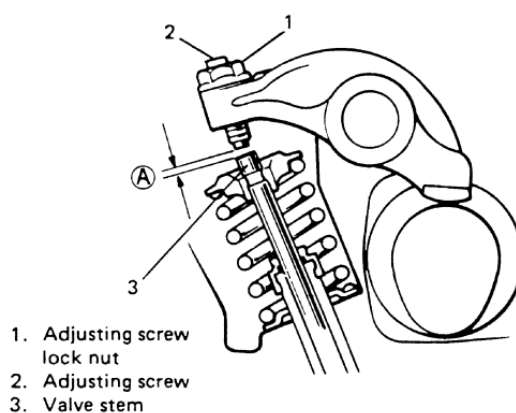


1. Adjusting screw
   lock nut
2. Adjusting screw
3. Valve stem

Figure 1: Mechanically adjusted overhead cam valve system[8]

In Wright's example, he imagines that if, in an internal combustion engine, "a small nut were to work itself loose and fall under the valve-adjustment screw in such a way as to adjust properly a poorly adjusted valve, it would make an accidental contribution to the smooth running of that engine." (Wright 1973, 152).  We can picture, in the diagram above, the nut falling over the top of the valve stem, reducing the lash such that what was too large a gap is now within tolerance.  As we saw in chapter 2, Wright holds that "conscious [design] intent is what determines the function an artifact

---

[8] From http://www.autozone.com/autozone/repairguides/Suzuki-Samurai-Sidekick-Tracker-1986-1998-Repair-Guide/ROUTINE-MAINTENANCE-AND-TUNE-UP/Valve-Lash/_/P-0900c15280087e4f

has got" (Wright 1973, 146). Therefore, he claims that we "would never call the maintenance of proper valve adjustment the function of the nut. If it got the adjustment right it was just an accident." (Wright 1973, 152). The nut is indeed helping the engine to run smoothly – it is making a causal contribution to what seems to be the teleological goal of the engine as it is used. But no one put the nut there, and no one even knows the nut is there.

As it reflects the specific intentional origin, design satisfies the etiological concept of function requiring that the presence of the functional item be accounted for by the fact that it has that function. The adjustment screw, unlike the nut, doesn't just make a teleological contribution to the engine – its very presence and origin are *explained by* this contribution. But the nut is an accident; its origin is not explained by the effect it is having on the system. It is an 'open loop', and not part of the engine's design. Therefore, Wright claims, it has no function.

But in this example, how do we *know* that the nut is there accidentally in the first place? As the author of this fictional case, Wright avails himself of knowledge he could not really have were he just looking at the engine: he *tells* us that the nut "worked loose" and ended up under the rocker arm screw. There is nothing philosophically suspect about this – Wright simply constructed a fully-formed scenario to demonstrate what he considered to be a clear example of non-function due to accident. Philosophers do this sort of imagining all the time, as we did above and will do extensively below. Yet in point of fact, if Wright were to really pull the valve cover off the engine in his car and see

such a thing, he might be a great deal more puzzled.  The fact that Wright seamlessly supplies an unseen backstory may seem a minor point; but it actually has serious implications, for it gets to the core of the question of what function centrally is, and whether we can know it when we see it.

**Original Scenario: Unknown and Accidental**

Let us now return to the nut example deprived of the certain knowledge that it is there entirely by accident.  We should first recognize that if the nut were to suddenly shift places or go missing, we would notice and miss its contribution: the engine would suffer power loss, run noticeably less smoothly and the noise and wear of the valvetrain would increase.  Note that precisely the same thing could be said if the adjustment screw, which is there by design, went missing.  In this particular engine, with the fortuitously placed nut, there is no relevant formal difference in the contribution that each – the accidental nut on this valve and the designed adjustment screw on adjacent valves – makes to achieving the designed-for and intentionally-desired end of the top-level system as a whole (the overall engine).  Any claim that the nut does not possess function while the adjustment screw does cannot rely on distinguishing their actual causal contributions to the system.  Indeed, according to GGC, they have identical functions, and it is likely that they even have them equally strongly.  In order for etiological theory to distinguish between them, we must have knowledge not of the formal contributions each makes but of their detailed history.

**Scenario Variation 1: The Accidental Nut Is Discovered – And Left in Place**

Now let us suppose that the owner of the car in question decides to poke around the engine to educate himself on its workings. He removes the valve cover and spots the nut, which has now seized in place. He notices how it shims the valve, allowing the rocker arm to make smooth, consistent contact with the valve top. What is he to think of this?

We know, by the stipulation of the case, that the nut worked loose from elsewhere inside the head, and found this resting place entirely by accident. The owner, however, presently has no way to know this. So let us imagine that he surmises (albeit incorrectly) that the nut was deliberately placed there by someone to shim the valve. Perhaps the owner believes it to be the originally designed function of the nut. By Wright's theory, we would in that case have had an example of genuine function rather than merely an accidental effect. Of course depending on precisely what the owner takes to have happened, the case may still be ambiguous. Wright insists that the presence of the functional item be explained by design. To believe that it was placed there deliberately is not yet to say that it was placed there by design, if design means the original engineering plan for the engine. We shall consider other possibilities shortly. For now, we imagine only that the owner supposes that the nut should be there, whether by original design or otherwise, and so he decides to leave it in place.

Note that once this occurs, even though the owner was mistaken about the origin of the nut, the intentional picture around the placement of the nut has now changed.

Simply by noticing it and deciding to leave it there because of the useful effect he saw it to contribute – even if his belief as to its origin was mistaken – he himself has thereby introduced an actual intentional aspect to the causal story of the nut's presence. Human intention is now a part of its etiology even if it was not before, because the owner is now aware of the nut and has *chosen* to leave it in place, due to the positive effect he observes it to be having on the engine (he can see that it properly shims the valve). The causal contribution of the part has not changed. But if there was no intention in its etiology before, there is now. There is at least a passive intent – an acquiescence – for the item to play the role in the system that it does. It could have been removed, but was not due to the effect it was observed to have.

Without any change in form or contribution, the same item seems to have shifted from non-intentional to intentional etiology (albeit not design) essentially on the basis of observation. Following the lessons of chapter 2, this indicates how fragile a basis intent is on which to base all artifact function attribution, particularly in comparison to the reliable empirical indications of causal contribution to the system.

**Scenario Variation 2: The Nut Was Placed Deliberately at Some Point**

In the variation above, we abided by the premise of Wright's original scenario: the nut came loose from somewhere inside, then ended up where it did by fortuitous accident. The only novelty introduced was in imaging what an actual owner might know or suppose or even change if he saw it. If the owner believed that it was supposed to be

there, he might supply an intention that it remain there, in the process unwittingly changing the intentional aspect of its origin.

But the belief of the owner need not have been erroneous. Perhaps, in other variations of the story, the nut really *was* placed there deliberately by someone. Perhaps a slow assembly line worker has been placing this nut in this location because he found that the speed of the line's movement did allow him enough time to properly adjust the lash on each valve with its screw. Rather than be penalized for holding up the line, he finds that it is faster to simply slip a nut under this valve, where there happens to be a small depression that will accept it. He does this with every engine that passes through his station.

Or perhaps the engineers have determined from field reports that this particular valve (assume it is an exhaust valve) is prone to overheating. In analyzing engines with this problem, they noticed that while the valve itself remained intact, the adjustment screw was prone to cracking. As a precaution, and because they cannot cost-effectively redesign the entire system, they decide to have a nut installed underneath the rocker arm to shim the valve in the event of a problem with the adjustment screw, and to provide a thermal buffer. Despite appearances, the nut in this case *is* there by design, even if it was not part of the initial design. Or, perhaps the car is used, and a former mechanic placed the nut there because he did not have a screwdriver handy with which to adjust the valve via the screw, or perhaps because the screw was stuck or broken, or perhaps because it had reached the end of its adjustment travel.

In each of these situations, the nut is present by deliberate intent, though perhaps not always by an intent that would qualify as design. If attributing function requires design intent as Wright maintains, the propriety of function in some of these cases will remain ambiguous – neither entirely accidental *nor* fully by design. And once again, the formal system is identical to the one in which Wright declared that the nut had no function.

**Scenario Variation 3: Selection for the Nut at the Factory**

We might also imagine a slightly different variety of accident. Rather than coming loose from somewhere else during use, let us imagine that the nut was accidentally dropped into the engine head during manufacture, without anyone noticing. In this case, if the nut had immediately worked its way between the rocker arm and the valve to effect a proper adjustment, then the nut's function could have been selected for.

Its causal role there would have been unknowingly selected for if we assume that every engine is checked for power output and smooth operation before it leaves the factory, such that any engine with a noisy valvetrain or rough running would not have passed quality control. In that case, the placement of the nut would have enabled that engine to pass inspection. There was no design intent for the nut to serve this role, and yet, all the same, its etiology is accounted for by its functionality just as etiological function theories require.[9] Had the nut lodged in a place where it negatively affected the

---

[9] As noted in chapter 2, for whatever reason Wright only considers design as the etiology for artifacts, and Millikan implicitly does the same. Why they accept selection etiology for natural systems but not for artifacts is an interesting question. It is clear that in any context both selection and design satisfy the function-etiology requirement.

engine's operation, it would have been discovered and removed.  It is only by adopting a positive functional role – by 'pitching in' and making itself useful – that it was able to pass unobserved.  And it is only by the nut adjusting the lash properly that the engine was able to pass performance testing.

Like the original case, an analysis of the engine alone could not conclusively demonstrate, in principle, which parts were functional by design intent or even by selection.  Such a determination could only come through ancillary historical investigation, such as interviewing engineers, mechanics and factory workers, analyzing QA records, or, ideally, consulting design records and specifications.  But within the system, neither design intent nor selection leave any definite identifying fingerprints.  However, if we were determined to discover whether the well-placed nut was possibly there intentionally, and did not have access to the people and records of the manufacturer, could we perhaps gain insight by comparing the system with other tokens of its type?

**Scenario Variation 4: Token Comparison**

Imagine again that the owner has removed the valve cover and is investigating the engine's valvetrain.  He spots the nut, sees how it properly adjusts the valve lash, and is puzzled.  He considers the possibilities we have just run through.  Did this happen by accident after the engine was built?  (The precision of its placement makes that seem unlikely).  Was it placed there by an assembly line worker too hurried or careless to adjust the valve with its screw?  (Its precision placement makes this seem somewhat more likely).  Did it happen by accident at the factory, and because it worked it was able

to pass inspection? (However unlikely its placement, this would at least explain it still being there). Or was it, perhaps, actually part of the design? (Perhaps there was a good reason, such as it being a workaround, developed during production to compensate for a design defect discovered after production had begun).

This time, the owner decides that he would like to know what is going on, because he would like to know if he should remove the nut and compensate for its loss with the adjustment screw, or leave it in. In other words, he wants to know if the nut really "belongs there" or not (he defines "belonging" as presence via intent with sufficiently good reasons that still apply). He has no recourse to the designers, the design plans, or the factory records or personnel. But he recalls that he does have a neighbor with the same model and year vehicle. If the nut is supposed to be there by design, he reasons, it should be there in his neighbor's car, too. His neighbor allows him to take a look, and he sees no nut under the rocker arm of the corresponding valve. But this is a small sample, and the odds are no better than 50/50; strictly speaking, and not taking into account the unlikelihood of the placement, either one could be the accident. Wanting a larger sample in order to be more certain, he joins an online owners forum, and requests that everyone there check their corresponding valve for a nut adjusting the lash. Dozens and dozens of owners helpfully check and report back, and every single one, including those built around the same time as his, says that they have no nut there. He now concludes that whatever else may have happened to cause the nut to end up where it did in his engine (he is still unable to rule out a lazy worker, a mechanic or any form of

accident), he believes that the nut was not part of the design, and thus was not originally supposed to be there.  So, seeing no further reason to leave it, he decides to remove it and readjusts the valve screw.

This story seems to show that type comparisons can at least distinguish design from random improvisation, accidental selection or pure accident.  Of course, this is reliant on the engine in question being a token of a uniform type, such that the statistical norms of the type can speak to design intent.  Yet even when this condition is met, there is both a minor and a major difficulty with the reliability of this sort of evidence.

The minor difficulty is that it must be specified what the proper thresholds for the statistical norms are, and there are no obvious standards for doing this.  What constitutes a sufficient sample size, and what is a sufficient preponderance?  What if only one online owner had responded, for a total sample of three and a no-nut predominance of 2-to-1?  Would this be enough to conclude something about design?  These statistical questions are inevitable, but they are not particularly troubling philosophically.  There is, however, a more serious way that this could go wrong.  What if the engines, which are after all largely identical tokens, all suffer from the same defect?  What if, for example, as in the original scenario, the nut worked loose from somewhere else in the head and accidentally lodged under the rocker arm?  And what if for the same reason, whatever it may be, half the other engines in the sample had similar things happen? [10]  It would then be impossible

---

[10] It would not even need to be the case that all of them end up properly adjusting the valve.  As we have stipulated, the unlikeliness of that occurring is what has prompted the investigation in the first place.  Yet any loose nut could be interpreted as having come from under the rocker arm, where it was originally

to distinguish design from defect. There is also possibility of the rushed line worker scenario, who put the nut under that valve in all the engines he touched. There, too, improvisation cannot be reliably distinguished from design on the basis of token comparisons. Once again, we are forced to appeal to information from outside the system itself, and even outside the system as a type.

The conclusion must be that while it is often a useful guide, type preponderance of a trait is not a perfect substitute for information regarding intention generally, and still less for known design intention. For artifacts, type standards can at best serve as *prima facie* evidence for design intention, all things being equal. A final scenario variation will drive these points home.

**Scenario Variation 5: Custom One-off Engine**

The typological points made above obviously apply only if there is an established type. Let us continue the same basic scenario of the nut adjusting the valve, except that now, instead of it being a mass-produced engine, imagine that the engine is the one-off creation of an eccentric lone engineer, made by hand in her own shop. The token comparison strategy explored above is now inapplicable. There are no others like it. If we had no other examples with which to compare it, when and how could we dismiss the nut as an accident? Certainly, there must still be a designer, and we can still look to design intention as she expressed it. For that, we could only turn to the account of the maker herself. We could ask the creator of this one-off engine if she intended the nut to

---

placed by design. It is also worth noting that Boorse confronts both these problems in his defense of his type-based BST theory of disease, though this is beyond the present scope.

be where it is, and to do what it is doing. Suppose she says that she did not originally intend the nut to go there, but that in tuning the engine she noticed the nut and decided to leave it there because it seemed to work. What then? Does it *now* become a function by design, or is it merely improvised function? Which standard must it meet to qualify as etiological function? Can it be both?

Worse, if the lone eccentric engineer had died upon its completion, the ambiguity of the nut under the rocker arm becomes even more intractable. We will never know if the nut was by design or by accident, unless there are records to appeal to. But even these could be unreliable or ambiguous. If there is a blueprint which shows the nut, we might conclude that it was designed. If there is a blueprint that does not show the nut, we might assume it was improvised (if not an accident). But what if the blueprint itself shows no nut, but a nut is penciled in on the print? Is it design or improvisation? Does the penciling-in indicate a last-minute design change, or a reminder that the nut was spotted and left in place?

In such a case, the very line between design and improvisation becomes ambiguous. What it comes down to, without additional units made or planned, is perhaps at what point in time the intent behind the nut arose. Was there a prior design phase, distinct from the actual construction stage? If so, at which stage was the nut added? While accident, implying a complete lack of intent, may remain a distinct category, the line between design and improvisation seems to be ultimately a matter of degree, of points on a spectrum – and in any given case it may be quite ambiguous. This does not

bode well for an etiological approach.  If function requires, as it does for Wright, an intentional etiology that explains the presence of the item in terms of its role in the system, we will never know if the observed effects of the nut qualify as function without information from outside the system attesting to its history.

On the other hand, once again, the contributions of the part are empirically evident – so long as we have some idea what the end of the system is.  Yet therein lies the complication for a contributions approach: what fixes the end of an artifact?  Boorse held that the end was basically set by the immediate intentions of the human user, no matter how disconnected from the intentions of the designer, or the suitability of the part for the task at hand.  Yet this broad approach proved problematic, attributing function liberally and on an ad-hoc, and even post-hoc basis.  Likewise, under this interpretation, artifacts crafted and suited for one explicit purpose would not ever have it until and unless they were actually used as such, and would still be attributed different functions if they were used differently.  This situation is what prompted us to look at the role of intention and contribution in artifacts afresh.  Let us now examine the issue of systemic ends more closely in light of the scenarios above.

**Intention and Ends**

We began by looking to the design intent of parts, as it is the most obvious form of intent relevant to artifacts, and yet the one which Boorse dismisses as essentially irrelevant.  But we also saw other ways that intentions became relevant for function attribution.  For example, the intentions of the engineers and the quality control

technicians are what informed the standard for the performance of the top-level system of the overall engine. In the case above where the nut was selected for at the factory, this set of intentions, in the form of performance specifications for the engine, selected for the presence of the nut without any awareness that the nut was there. There is no etiological design intention for the nut, yet top-level intentions were involved in its selection.

Thus, a confluence of intentions may intersect on a single artifact – not only at different scopes of analysis (parts, subsystems, top-level system) but also of different parties (users, designers). Which of them is relevant in attributing function, and under what conditions? If function is when a thing contributes to the top-level end of the system of which it is a part, then, logically, what counts as a contribution will depend on what counts as an end.

Imagine if the car owner had noticed that the engine was running a touch rough, and had resolved to do something about it. Before she can, coincidentally, the nut works itself into place under the rocker arm, and cures the problem. The owner recognizes that the problem has now 'fixed itself' and her intervention is no longer necessary. On the other hand, what if a less attentive owner had not noticed the slight rough running in the first place, and also did not notice or care when the nut shifted into place and resolved the rough running? In this case it might be difficult to justify a functional role for the nut at all. From the owner's own perspective, the intended ends of the system were *already* being met. By that user's own standard, the effect of the nut was imperceptible, and so did nothing to contribute to the intended end of the system.

Of course, from another perspective, had he brought the car to a mechanic for something else (before the nut shifted into place), that mechanic may well have noticed the rough running.  Or, for that matter, if a designer of the engine had observed the valve tolerance before the nut, he surely would have noted it as a defect in adjustment compared to published specifications.  Here again, which of the various possible intentions are the *relevant* standards for function is debatable.  It might be argued that the owner's inattentiveness is irrelevant because the only applicable standards for the engine's performance are those dictated by its design specifications.  Yet just as with intentional etiology of the parts, the specifications of the designers are only evidence of envisioned use intentions which may or may not entirely apply to the actual use of the system.

In other examples, design intentions in the form of specifications are even more clearly irrelevant.  Sound quality in hi-fi is a good example.  Acceptable music reproduction is an inherently qualitative and thereby subjective standard.  What meets designer specifications may not meet customer expectations and vice-versa.  Vacuum tubes, for example, by their nature add significant amounts of higher-order harmonic distortion, something designers of conventional solid-state amplifiers generally strive to avoid at all costs.  Yet the objectively less accurate sound of tube amplifiers is aesthetically preferred by legions of audiophiles.[11]

---

[11] The conventional wisdom is that while vacuum tubes produce far higher levels of harmonic distortion, most of this occurs at even orders, which human psychoacoustics interpret as realistic timbre.  While solid-state amplifiers produce far less overall distortion, what they do produce is largely odd-order harmonics, which sound unpleasant.

Suppose then that there is a solid-state amplifier designed to reproduce analog signals as accurately as possible according to the usual quantitative metrics of harmonic distortion, frequency response, noise level and other such measurements.[12] Now imagine that a capacitor within the amplifier breaks down and introduces distortion in the signal path, a kind of harmonic distortion that mimics the profile of an expensive vacuum-tube amplifier. By the standards of the intentions of the designers of the capacitor, the capacitor is malfunctioning (that is, not operating according to design intentions). By the standards of the intentions of the designers of the amplifier, the amplifier is malfunctioning. But by the standards of the right prospective buyer, this amplifier may sound audibly superior to its competitors. By his standards, the capacitor is not malfunctioning, but making a very valuable functional contribution to a desired end. Who is correct? That depends on which end, and which resulting functional perspective, is considered most relevant or most legitimate.[13]

When any artifact is designed, it is done so with certain predetermined "use cases" in mind: contexts for which the artifact is intended, and standards for meeting the user needs of these contexts. This extends to specific market segments which the product is intended to fit, and niches which it is intended to fill. Thus, a hi-fi enthusiast is unlikely to be satisfied using an old cassette Walkman as his source component, and yet

---

[12] Note that we cannot simply assert that the design intent is to reproduce the signal perfectly because, besides being impossible, there must be specific finite sets of measurements by which to judge actual performance.

[13] It is convenient to focus on physical artifacts; however, perhaps the most common example of the design/use conflict is in software: as any computer user knows, there is often a fine line between feature and bug.

he could not accuse it of malfunctioning; it is simply not the right device for the task, although it is close in its basic function of music playback. The intent on the part of the designers is that the users employ the artifact within the range of designed-for "use-cases", so that there is a fit between the designer's envisioned intentions and user's actual intentions. This is when success in achieving the desired ends of the user is most likely. Yet the two intentions are formally independent of one another and we cannot ensure that they overlap in any given instance. The use of an artifact may always depart by degrees from that envisioned by the designers, and if the departure is great enough, at some point we might characterize it as improvisation.

**Expectation and Malfunction**

So it seems that according to the ends of the engine in terms of its overall design specs, both the accidental nut and designed-for adjustment screw perform the same function. Removing either one results in the same deficiency in attaining the ends of the system as a whole. Suppose now that after many years of the accidental nut reliably performing the same function as the adjustment screw, something changes. Perhaps the car hits a pothole particularly hard, causing the nut to be jolted out of position. The nut no longer properly adjusts the valve; it thus no longer performs a function. But is the nut *malfunctioning*?[14] If the nut was only there by accident, and it was only benefiting the engine by accident, on what basis could we say that the nut is malfunctioning? The nut either happens to be there, or it does not. There is no apparent basis to say that it "should"

---

[14] Note that "malfunctioning" is being interpreted very broadly here to mean "not doing what it should". Thus, shifting in place or breaking in two would both be examples of malfunction, even though we might not normally choose this term to describe these conditions.

be there, or that it "should" be properly adjusting the lash. If its function was serendipitous, its loss is not a failure but simply luck running out. To be sure, we might regret that the nut no longer helps achieve the intended end of the engine. But to accuse the nut of malfunction is more than an expression of regret – it is to say that it fails to do what it somehow *should* do. The accidental nut offers us no apparent basis for such a charge.

Compare this scenario to a comparable instance with the adjustment screw. Suppose the valve adjustment screw on the adjacent cylinder were to suddenly fracture from repetitive stress. The screw no longer properly adjusts the valve; like the nut, it no longer performs a function. Is the adjustment screw malfunctioning? Intuitively, we would readily acknowledge that the screw is indeed malfunctioning. It is not performing the function that it should. Unlike the nut, the broken screw *does* seem to deserve the accusation of having failed to behave properly.

Boorse was hesitant to attribute malfunction because to do so would imply having a function without performing it, which is impossible in GGC, just as it was for Davies and Cummins in systemic effects theory. Why, then, in these sorts of everyday examples, does it intuitively seem not only possible but ubiquitous? The core principle at work here seems to be expectation. We did not expect the nut to adjust the valve; that it ever did was a surprise. We do, however, expect the screw to adjust the valve. Indeed, that is precisely why it was designed, manufactured, and installed in the engine. As etiological theory would remind us, that is its reason for being.

Davies, in seeking to explain why etiologists believe so strongly in the intuition of malfunction, reached a similar conclusion. Speaking of a "malfunctioning" natural organ, he noted that "[t]he feeling we have that this token ought to be formed differently is an expression of our expectations concerning token components within this type of system … Precisely these expectations … are the source of our inclination to mistakenly see nonfunctional tokens as malfunctional." (Davies 2001, 177). According to Davies, when we attribute malfunction, "We are simply expressing our expectations of the token in light of our knowledge of the larger system" (Davies 2001, 178). This is true of artifacts as well, he believes, and is why, as in the case of the screw, we "harbor expectations of the components of artifactual system" (Davies 2001, 178, note 7). And yet, not all components of a system have this expectation associated with them; the nut, if we know it to have been accidentally placed, carries none.

Again, the difference between the nut and the screw is not one of function performance – it is one of etiology. In the case of the screw, it is apparent that the main source of these expectations is that it was designed to perform this function in the engine. The nut was not. Etiology may not be inherently involved in attributing function. But it, or something like it, does seem implicated in forming the expectations by which we attribute *malfunction*.

<div align="center">**Function Assignment**</div>

The scenarios show that, while GGC is on the right track in terms of what function is essentially about, it has missed something crucial. Function does indeed seem

to consist in causal contribution to teleological ends.  That common thread stands out in our analysis of all the scenarios – it is the only thing that each one has in common.  But even so, in not every case is that function performance intended or expected.

The apparent connection between malfunction and design points to a more fundamental distinction between the nut and the screw.  Even if both the adjustment screw and the accidental nut are in fact performing the same function, it seems that the functional claim for the screw is somehow deeper: it enjoys the specific expectation, even obligation, of performing the function in a way that the nut does not.  We might try to articulate this by saying that while both the nut and the screw *perform* the function, only the screw *has* this *as* its function.  The screw has effectively been *assigned* this function.  The nut *performs* the same function.  But it does not *have* this *as* its function.

The screw was intentionally designed and manufactured so that it would perform this function.  That etiology binds the screw to its functional role whereas the accidental nut is not bound to its role in the system at all.  The potential strength of such a bond between a part and its role is evidenced by the possibility of malfunction – for malfunction is just failing to adequately perform the function which the item *still has*, even when it does not perform it.  This function assignment is persistent, while function performance may be transitory; in cases of malfunction our expectations for function performance persist even after its actual performance ends.[15]

---

[15] Further evidence that malfunction is about expectation lies in the fact that we do not call a burned-out light bulb or a depleted battery malfunctional if they performed for a normal service life.  We expect that these devices will fail after a time.

As malfunction is simply function-having without function performance, it is not itself a novel concept. The theory of proper function, in particular, was conceived of to provide an account of this. Yet in this conception performance still implies having; just not vice-versa. In contrast, what is being proposed here is quite different: it is the divorce of function-having from function performance such that *either* is possible without the other. This is an important and novel principle, without direct precedent in the thinkers we have studied here (or, it seems, in the function literature generally).

There have been some parallels, of course. Boorse comes close to articulating a distinction similar to having and performing in his schema of weak and strong function; but this distinction is never spelled out, and no firm criteria for judging between strong and weak are offered (indeed, we are told there can be none). Moreover, even strong function is not eligible for malfunction attribution, according to Boorse. Millikan's proper function is an example of a sort of function assignment; yet for her, this is all there is to function attribution. There is no true function performance apart from actually having the function. Millikan and Wright also approach the distinction in noting the difference between 'function' and 'function as'. However, 'function as' was treated as a second-rate imitation of real function, a mere figure of speech; genuine function performance was taken to go hand-in-hand with having proper function. But if function is really about contribution, function performance by an item not having that as its function is just as genuine as that performed by an item that does have that as its function. Just this is the lesson of the nut and the adjustment screw. The nut, when in the right

position, makes just as real and regular a contribution to the system of the engine as does

the screw.  There is nothing ersatz about its contribution and thus its function

performance.  No euphemisms are called for.  Nonetheless, the designed adjustment

screw has that *as* its function in a manner that the loose accidental nut does not.[16]

No thinker, it seems, was willing to fully separate the concepts of performing a

function and having a function.  Yet once postulated, it seems obvious that this is a vital

difference, and, moreover, one that solves many puzzles in function theory – including

the real difference between the nut and the screw and why we still think of them

differently even when we realize they are doing the same thing.  It seems that a thing can

perform a function without actually *having* that *as* its function.

The reticence in embracing this deep distinction may stem from a perception of

inconsistency in saying that an item can perform a function without having that as its

function.  Yet properly understood there is no difficulty here.  This simply means that an

item is making a real contribution to the ends of the system with which it is involved,

even though it has no particular "pedigree" to do so (or at least, none that we know of).

Such circumstances are not common for obvious reasons: the world is complex enough,

and functional systems sophisticated enough, that it is unlikely for an item to drop in

randomly and play a valuable role.  More often, random intervention in a system will

---

[16] It is true that we might colloquially say that the nut is "functioning as" an adjustment screw.  But this can plausibly be interpreted as meaning that the nut is performing the same function it would if it were an adjustment screw, rather than impugning the function itself as lesser.  In contribution terms, the function is just as real; it simply is not expected of the nut, as it is of the screw.

result in interference with it attaining its goal, not assistance.  But this does not mean that accidental function performance is impossible, or incoherent.

Indeed, if we think back to one of the more compelling arguments against proper function in chapter 2, it was seen as problematic that function could not be attributed to first-generation organism traits that were plainly beneficial but that had not yet been selected.  Though the traits were beneficial, they arose randomly, so Millikan could not yet attribute them proper function, despite their seeming to be functional in all other respects.  With the assignment/performance distinction, this scenario becomes far more tractable.  We may say that the traits have not yet been assigned function due to their not yet having been selected for – there is no lineage on which to base any expectations for function.[17]  But we can still recognize them as fully functional based on their contributions.  Every biological function, of course, began as an unselected, randomly-generated trait.  To say that no trait has any function until it is selected is to miss what the concept of function itself seems to be about – contribution to a goal.  And to say that there is no difference between having been selected and being accidental, as Boorse does, is to miss something equally significant.

Distinguishing 'performing a function' from 'having a function' allows us to avoid conflating the performance of a function with the etiology of the functional item.  Wright was wrong to attribute function based entirely on etiology; but Boorse was perhaps wrong to entirely ignore it, for origin does have relevance for function

[17] Though Millikan was referring to a selection etiology and we have so far based function assignment only on design, recall that design and selection were both considered functional etiologies by Wright and Millikan.  As we will soon see, such assignment can come about in more than one way.

attribution. Etiology cannot explain what function *is,* but etiology (if known) conditions our *expectations* of function. If we know that something was brought about in order to serve a function, then we have good reason to expect that function to be performed by the item, and we interpret that function as being assigned to the item as its "job".

We have seen that for artifacts, human design intention is an important basis for such an expectation and assignment. For Wright, indeed, it was the only basis for function in artifacts at all. But our panoply of examples above paints a different and more complex picture.

**Selection and Function Assignment**

In Scenario Variation 3 above, we imagined that the nut, rather than coming loose at some point after manufacture, had instead been accidentally dropped into the head at the factory. Because the nut had assumed a functional role within the system, if only accidentally, the engine, with the nut, was able to pass quality control screenings. Thus, the nut was essentially selected for via its function: if the nut had not performed a function as it did, it would have been detected and removed. There is no specific intention behind the functional role of the nut itself, but the intentions for the ends of the overall system resulted in the nut's functional contributions being selected for. Thus, the presence and position of the nut are explained by the function it performs, much as they would have been if the nut had been intended by design.

Obviously, selection of accidental parts is not typical of artifact manufacturing. Typically, each functional part is specially designed and assembled to fulfill a specific,

preconceived role, and these roles cumulatively fulfill the design-intended end of the system. For example, as we have noted, the adjustment screw present on the rocker arm above each valve is intended to allow adjustment of the gap between the rocker arm and the valve top. Yet even engines which have only their intended parts (and no stray nuts) pass through the same quality control inspections to ensure that all parts successfully perform their intended functions.

The fact is, then, that even the intended, designed parts playing their intended, designed functional roles within the system are there not *only* by intention – they are also there by selection. If any parts had failed to play their intended roles, they would have been eliminated in favor of parts that did.[18] This selection principle is not limited to mass production scenarios – it goes all the way back to the original prototype of the engine. When that first prototype of a new design was assembled, the same standards for part performance would have been applied. Parts which failed to carry out their intended function would have been removed, redesigned, or replaced. Design explains intended function; but, assuming there is any effort to ensure success, only selection explains *successful* function performance. Likewise, design intention explains original part *creation*, but selection explains *retention* following creation. Selection depends on prior phases of design or accident to provide materials against which to perform selection. But

---

[18] It should be noted that this construal of selection is broader than the one typically understood to apply in Darwinian natural selection. There, most philosophers of biology, following Sober, insist that natural selection is only possible given actual variation in traits from among which to select (Sober 1984, Lewens 2005). In my interpretation of selection for function, I am defining selection only as a filter that rejects non-functional or sub-functional system variations, whether or not multiple variations actually are presented to the filter.

given such materials, selection, in the sense we are concerned with, involves picking out just those systems which conform to the predetermined criteria for success in the system and rejecting the rest. In the present example, the criteria are a smooth-running engine producing adequate power. All parts which successfully contribute to this end are selected for when an engine passes through quality control based on these metrics.

But selection in artifacts clearly does not involve or imply omniscience of everything that is being selected. The accidental nut is just as selected for as the designed screw. Both perform the same function, and help the system achieve its end. As a result, both are selected for, even though one is unknown and unintended and the other known and intended. The blind selection of functional elements whose particular function is unknown is not limited to accidents with particular tokens. It can apply just as well to entire technologies, and often has, as the brewing example Boorse offered above demonstrates.

This is in fact a far more common scenario than it may seem, or than even Boorse allows. Even modern engineering relies on processes and technologies which may not be fully understood (or may be mistakenly conceived of by the standards of as-yet future science) to nonetheless achieve reliably successful ends. Technology and engineering could not operate otherwise. To assert that function in artifacts only ever occurs via specific intention for each and every part is in fact to suppose that we are omniscient, arranging matter with complete clairvoyance. Actual engineering design is a far messier and haphazard proposition, even as scientific understanding continues to advance.

Science has played catch-up to technology perhaps as often as the other way around; for example, the invention of semiconductors preceded their theoretical understanding, just as the steam engine preceded the science of thermodynamics. Semiconductors and steam engines to some significant degree were built on the basis of doing what had been observed to work, meaning what had helped to achieve systemic ends – not from pure preconceived design intent alone. Of course, there may be some ambiguity here with the nature of the concept of design intent, specifically the degree to which it implies an intention to utilize the precise mechanisms by which a given part performs its function. But by any reasonable definition, design intent requires at least understanding the specific contribution a given part will make to the ends of the system – and thus of the reason why it is necessary. This is not always available. To the extent that it is not, only selection is operative.

It should be apparent that selection can operate not only on tokens, as with the accidental nut, but also (and in fact more often) on entire types. In a prototype engine, for example, the token parts are treated as representatives of potential types: when the design is put into manufacture, those tokens, having undergone selection, become the first of their type (hence the etymology of the word "proto-type"). Further tokens can then be selected not only on the basis of their contribution to the system's ends (actual function performance), but also by conformity to the parameters established for the type – for example, the length and thread pitch specified for the adjustment screw. The pragmatic value of this is obvious: selection can be performed outside of and prior to the

assembled system (for example, the screws could be manufactured and selected for based on their specs at an entirely separate factory, rather than test-fitted to an actual engine on the line). However, this is possible only if the part itself is intentionally designed – it would be impossible to perform token-level selection outside the overall system if the specific contributions of the token are unknown. One would simply have to introduce the token part into the system and see if the desired ends are achieved.[19]

**Design vs. Selection**

Design is not the only means to close the "teleological loop" in artifacts. Wright invokes design intent to explain the presence of the part in terms of the function it performs; but we now see that in a somewhat different way, selection does this as well. Design explains how the *original* creation of the item is due to the intention that it perform its function. Selection explains its *retained* presence in terms of the successful performance of that function, with or without design explaining the original genesis. Both effectively close the teleological loop by showing how the presence of the item in the system is in fact due to the fact that it performs its function, as well as vice-versa.

Because both design and selection close this teleological loop, both are reasonable grounds for the assignment of function to items which have these etiologies. Just as design intent allows us to say that the adjustment screw possesses a function while the random nut does not, selection seems to allow us to say that the random nut that passed through quality control possess its function as well. But if both design and selection are

---

[19] The exception would be the use of proxy indicators in the form of incidental cues, for example the smell or color of historically successful yeast in Boorse's example.

each independent bases on which to make function assignments, what happens when they conflict?  The valve adjustment screws in the engine are there by (and have their function assigned via) design intention.  They also, we have said, have their function via selection as well, because they succeeded in making necessary contributions to the predefined top-level ends of the system in a context that would have eliminated them if they had not done so (i.e., they helped the engine pass QC).  Both design intent and selection apply to both these particular screw tokens and their type as whole.  On the other hand, the accidentally deposited nut which passed through QC does not have its function assigned by design; at best, it only has it via selection alone.  So should we say that this nut has an assigned function or not?

In the discussion of selection, we had already noted that intentional design is not always present in every part of an artifact.  Often, there may be incompletely understood aspects to the technology, and in such cases selection is relied upon even in the absence of design.  If this is so, then the mere absence of design is not disqualifying for function assignment.  Yet it is also the case that this was a freak occurrence.  Unlike the examples of technological mystery, which are deliberately repeated, the selection of this nut was pure happenstance.  That is to say, it is not *typical* of the engines coming off the line.  We touched on this in the scenarios when the owner compared his engine with others of its kind.  Therefore, while it is true that in *this* token engine *this* token nut was selected for its function, in most token engines of its type, no such part token exists.  In most engines,

the design-intended screw is selected for this functionality. If the nut has any selection-assigned function, then, it has it only very weakly.

This reminds us that there is a difference between token-level selection and type-level selection. For our purposes, a type is essentially a set of like tokens. Therefore, the difference between token-level selection and type-level selection comes down to whether the token undergoes replication to become a type. When a token serves as a prototype, and is replicated (in the artifact case intentionally via manufacturing, and in the natural case via organism reproduction), then we can say that it is the *type* which has been selected and not merely the token. In these cases, a *selected* function also becomes a *typical* function. This in turn points us to a third and final basis for function assignment: typology.

**Typology and Function Assignment**

What if the epistemic challenges of etiology prevent us from knowing the true origin of a part? Or what if neither design origin nor selection seems to cleanly apply? What can we conclude about function assignment in such cases?

Imagine now that we stumble upon a crashed alien spaceship. We find no aliens aboard, but we do find what seems to be an armory, with 100 identical devices that look to us like ray guns. Each features rudimentary pictograms on their use – the equivalent of safety instructions – allowing us to conclude that their design-intended use is as a weapon. On the top of each of the devices is an oddly glowing purple crystal. As an experiment, this crystal is removed from one of the devices. The device is found to no

longer fire its beam. We now know that in this token, that crystal performs a function, contributing to the intended end of the device.

Of course, we lack all records of design or production. The etiology of the crystal part is thus unclear. We cannot conclude that the crystal, which we observe to be functional, is there either by design or by selection, or even some other possibility. The crystals might be like semiconductor chips – meticulously designed and manufactured with relatively complete understanding of their operational principles. Or, for all we know, the crystals might be naturally occurring on their world, and merely need to be picked up off the ground and fitted to the devices, which perhaps merely crudely harness the energy they naturally emit. Yet because there are many tokens, we can compare them against each other.

Suppose that we find that in 100% of the token devices, all found to be in working order initially, when this crystal is extracted the devices no longer fire their beams. Because the technology is alien to us, we cannot immediately discern the precise role of the crystal in the system. Perhaps it is the energy source. Perhaps it focuses the energy. Perhaps it is just a key. We also cannot know whether or not the aliens themselves understood its operational principles, or to what degree, or even whether it is itself a designed part or not. While we know that the crystal performs a function, our ignorance makes it impossible to assign it a function via knowledge of its etiology, for we have none.

The only thing that we can be completely confident of is the fact that this crystal has a *typical* function in the artifacts of this type. Further empirical investigation could tell us much more about precisely what that function is; but even at this point, we know – assuming only that we are not mistaken that the tokens are identical and collectively constitute a type – that the part has a standard function across the system type.[20] Therefore, even not knowing how the part works, or whether the designers themselves understood precisely how the part worked – in other words, knowing nothing about etiology – we could still be sure that the parts *as a type* have a function. The only thing required for this conclusion is the confidence that the performance of the function by the part is standard across the devices. That is, we can reasonably assign the part a function (once we determine what its function actually is) on the basis of *typology* alone. Typology is essentially *de facto* function assignment – it is simply the *recognition* that, regardless of etiology, the part type in question already does in fact regularly perform a certain function uniformly across tokens in a system type. Thus, functions can be assigned to part tokens on the basis of being of a type whose tokens are seen to generally perform the same function in like systems.[21]

---

[20] A potential confusion should be avoided here: we are assuming that the tokens are in fact identical and constitute a type because if we did not know this, it is logically possible that the crystal plays a different role in different ray guns. It could be a power source in one, an energy lens in another, a key in yet another. We would not be able to discern these differences from our simple test of removing the crystal, but further empirical testing could detect these differences in functional role without making any appeals to etiology. Therefore, for the sake of simplicity, we assume that we are correct that the ray guns that appear identical are in fact functionally identical.

[21] It is perhaps worth noting that, as the quotes referenced at the end of the previous section reflected, Davies considered typology alone as the source of function expectations generally, whether for natural systems or artifacts.

In general, to assign a function to a part is to *assert* that the part has the performance of a certain function as its "vocation" – that, from our perspective, the token part can be expected and even be considered "obligated" to perform the function in question. The part has been assigned this function as its "job". This assignment of function applies even when that token is unable to perform the function – indeed, that is just when function assignment becomes important. For, so long as the function is in fact performed, it seems to matter little whether or not it is being performed by parts which have it as their assigned role. We worry more about such role assignment when something goes wrong, for then we often want to know which part has failed to perform its assigned function so we may correct it.

The expectation behind function assignment thus involves two subtly different components, one epistemic and one normative, which are conflated in the semantics of the word "expectation" but which are important to distinguish. The first, epistemic, aspect consists in having a reasonable basis on which to expect that the token part *actually will* perform a certain function. The second, normative, aspect consists in having some basis on which to claim a *deficiency* in the token part if that function expectation is not fulfilled. These two aspects are conceptually independent. In the next section I examine more closely how typology provides the epistemic basis, and in a following section I will address the normative, first with typology and then more generally.

**Epistemic Expectations and Function Assignment in Design, Selection and Typology**

As we have seen, design intention and selection (whether of type or token) each have strong and explicit etiological ramifications. When either applies, by definition, the presence of the part in question is explained in large part by reference to its function. The part is therefore expected to perform the function because its very existence is connected with this function performance.

In design, this is true even if no such part has ever performed that function, as in a failed prototype. Even then, we are justified in expecting function from design because function is consciously intended by design. Indeed, this is almost, though not precisely, tautological: our design intentions are our hopes for the functionality of the part, although we know that those hopes could be disappointed. In any case, the part carries expectations for actual function performance by virtue of the intentions which created it.

In selection, the presence of the part in question (either as the token itself or as a type or both) is explained by the fact that it or another like it (i.e., the token itself as or its prototype) *did* in fact successfully perform the function on previous occasions. This precedent provides the epistemic basis for the expectation of function performance from the present token part. An example of a selected type is a successful prototype that is put into production; an example of a selected token is a token part which helps the token system pass quality control screenings. A selected token part may or may not also be a designed part, and so have design-intended assigned function in addition to token-selection assigned function. Furthermore, a part may have a type-selected function yet

not a token-selected one, as would be the case with the successful prototype that enters

mass-production without the benefit of any Quality Assurance to ensure that each token

is functional (and the part in question fails to be). Again, all these modes of function

assignment (design, token-selection and type-selection) account for the function of the

part in terms of the etiology of the part, in turn providing a reasonable expectation for it

to actually perform a function.

Typology is different, however, in that no particular etiology, whether design or

selection or any other can be assumed – though any may still apply. Hypothetically, the

actual origin could even be accidental. The exclusively accidental etiology of functional

artifact parts would usually be an implausible scenario because with artifacts human

intentions always come into play. Even if a token part's origin were accidental, it would

usually be followed by selection of some form, depending on whether the intended ends

(in design or use) were met or not. In other words, widespread functional traits in

artifacts are rarely if ever both un-designed and un-selected, because the use of the

artifact is ultimately intentional, and this intentional use enforces a form of selection.

This may on occasion happen randomly, such as with Wright's original nut example. But

such aberrant cases do not usually apply to entire artifact types precisely because they are

random, and so unlikely to occur uniformly. Therefore, in artifacts, we can generally

presume that any part type found to play a function across an entire system type is likely

there intentionally, either by design or by selection.

Thus, type function assignment in artifacts, while extremely useful to appeal to in practice, is in effect a stand-in for incomplete knowledge of etiology. Because intention (and with it design and selection) is usually so easy to appeal to with artifacts, we indeed must go to some imaginative lengths to create a scenario where typological function must be relied upon instead (such as the alien spacecraft). In practice, we do so only when etiological information is unavailable or unreliable, such as when the owner in Scenario 4 resorted to comparing his engine with others.

With natural organisms, however, typologically-assigned function takes on much greater significance, as accidental origins are more relevant, and perhaps not uncommon. The paradigmatic example of natural accidental function-origin is of course an advantageous first-generation trait produced by random mutation that has not yet spread through natural selection. However, in such cases typological function also cannot be appealed to for obvious reasons. But as we saw in chapter 2, (e.g., with skull partitions) the fact that an organism's trait is typical cannot guarantee that it was selected, even if the trait is functional. Among the many possibilities, the beneficial trait could, for example, be a genetic or anatomical accompaniment of another trait which was selected for (which Stephen Jay Gould termed a spandrel); or the trait could have been selected for quite different reasons in the distant past than their present function (likely the case of the skull sutures); or the trait could simply be a by-product of basic chemical or physical limitations.

In other words, in organisms, type-level selection (i.e., natural selection) does indeed give rise to typical function, just as it might for artifacts. But the reverse connection is not as assured. Natural selection, besides being often difficult to demonstrate and difficult to interpret, is thus not the only possible route to typical function. For these reasons, in organisms, typical function can be a more reliable basis for function assignment than appealing exclusively to selection history. Typological function assignment is thus as important as the etiological varieties, and essential for the overall goal of a unified account of artifactual and natural function.

In typology, epistemic function expectation is based on the simple empirical fact that the vast majority of tokens in an available sample do in fact perform the same function.[22] Like selection, then, typology is primarily about precedent. In selection, this precedent takes the form of verified examples of like tokens which successfully performed this function in the past (this could be either a prototype ancestor, or the token itself when it was selected – or, typically, both). Rather than finding precedent in preceding instances, typology looks to the contemporary "precedent" set by like tokens. In a loose sense, typology is the "inductive" counterpart of "deductive" type-selection. In type-selection, because we know that the type has been selected, we reasonably expect that all tokens of said type will exemplify the same functional characteristics. In typology, we know nothing of selection history (or any other history), but we observe that

---

[22] As we saw in the scenario examples, the statistical guideline for what constitutes a sufficient majority is a practical challenge when applying this mode of function assignment. No general theoretical guidance seems possible, as any threshold must depend on the context of the case, and the interests of the investigators.

like tokens do in fact predominantly perform like functions in like systems and therefore, inductively, we then reasonably expect any given token to also perform this function.

Finally, note that assigned function in this revised approach still admits of varying degrees of strength, although this is not as important as the weak-strong function scale was in GGC. Nonetheless, some function can be said to be assigned more strongly than others, with the degree of integration judged by how many of the three potential assignment criteria are met. The adjustment screw in the engine, for example, is intended by design, selected by quality control, and typical for the system type. It has as strong a function assignment as is possible.

The accidental but selected nut token in Scenario 3 has only a weak token-assigned function. The improvised nut by a mechanic in Scenario 2 has a similar token-selection history, and also has a design-assigned function, if only weakly. The accidental nut without selection in Wright's original example has no mode of function assignment, and therefore performs a function it was not assigned.

Arguably, design intention is the weakest of the three assignment modes, since it is the only one that by itself does not imply function performance, but only the intent for it. Thus, a pair of human-powered wings might have design-assigned function, but be unable to perform the function. Selection, by contrast, implies successful precedent, as does type function.

**Potential Functions Revisited**

In the scenarios above, it was stipulated that the screw and the nut made the same contributions to the engine system – maintaining the correct valve lash.  The difference between them was seen to consist not in the function they perform but in the expectation that they would perform it.  Because the screw was designed (as well as selected, and typical) while the nut was not, only the screw was seen to incur the *expectation* that it function, even if its actual function performance is no different that the accidental nut.

Yet there is a reason why, in the design of the engine, a screw was chosen to set the lash and not a nut.  The fact is that a screw has external threads which allow its vertical position to be easily modified; this, in turn, makes the valve lash easily adjustable.  Thus, as the valvetrain wears, the adjustment screw, as its name implies, allows the lash to be adjusted to compensate. The nut lodged on the top of the valve, however, is of fixed dimensions and position.  While it may initially adjust the valve just as well as the screw, its dimensions and position cannot be adjusted, and so it cannot compensate for wear.  Eventually, the lash set by the nut will go out of tolerance, and at this point the nut will no longer adequately perform the function it once did, while the screw still can, if readjusted.  The screw, because of its range of adjustment, provides additional positions; this adjustability can itself be thought of as an additional function; unperformed at the moment, but one which the nut is entirely unable to perform.  This is thus a more subtle example of an issue we had already encountered, and noted as a weakness of GGC: potential function.

Recall that in the examples of knock sensors in engines or gametes in organisms, GGC was unable to satisfactorily account for their potential yet unperformed functions. The difficulties posed by potential functions can now be seen as another result of the inability to distinguish performing a function from having a function. Cases of potential function can be thought of simply as having an assigned function without currently performing it. The screw is properly adjusting the valve (a function it has and performs), but it also stands ready to compensate for wear by shifting up or down, even though it is not currently doing so (a function it also has but is not performing). It has this latter function assigned to it in all the same ways as the former, including all of the three modes we have elucidated: it was designed to be adjustable, it was selected to, and this is typical for its type. For all three reasons, we *expect* the adjustment screw to be able to compensate for wear. This expectation assigns the function to the screw even when it is not currently performing it.

This basic distinction even helps explain the previously-mentioned example of the scissors in a drawer, which make no actual contribution to anything, and so perform no function, and yet still can be said to have the assigned function of cutting (again, presumably via all three modes, design, selection and type).

**Normativity and Function Expectation**

We have seen an explanation of how typology, as well as design and selection, can provide the epistemic basis for an expectation of function. If a given part token is of a type of which the majority of members in fact perform a certain function, then

statistically the chances are excellent that a given token will also perform the function in a like system. But, as noted earlier, there is a substantial difference between the epistemic expectation that a function *will* be performed, and prescriptive normative expectation that a function *should* be performed. Simply because something can be reasonably expected to do something is not in itself reason to say that it normatively should do the thing, and deserves blame if it does not. And yet, as we saw with Millikan, the instinct for this is kind of normativity is quite strong and seems an innate part of malfunction attribution. To the extent that this is true, where does this normative prescriptivity come from, and how does it relate to the more rational epistemic expectation of function?

An explanation of this intuition seems particularly important for functions that became assigned without the involvement of intentions; i.e., via selection for natural organisms, or via typology for any system. In the case of etiologically-based function assignments in artifacts, whether via design or selection, there are always human intentions (either for the part or the whole, respectively) which can be appealed to to support prescriptive functional characterizations (either directly or indirectly, respectively). Yet even if the assignment is made without the involvement of intentions, we still tend to make normative malfunction attributions – for example, a "bad" heart. Davies argued that this is simply a metaphysical mistake that we must disabuse ourselves of. But a study of how we come to make such attributions in light of assigned function and the expectations it engenders reveals a more innocent explanation

If I throw a rock at a window, I may have a reasonable epistemic basis (based on past experience and a general understanding of the materials involved) to believe that it *will* break the glass. There is, however, no apparent basis for the belief that it prescriptively *should* break the glass – that the rock is *supposed* to. If the rock breaks apart rather than breaking the window, it could not reasonably be accused of malfunctioning, even if it was my intention to break the window with it. The rock may be performing the function of breaking the window, insofar as it is carrying out my intention towards the window.[23] But the rock does not have this function assigned to it by any basis for function assignment we have conceived of. It has no assigned window-breaking function via design intention, for it was not designed for this by anyone in a way that would explain its creation or properties; indeed, it was not designed at all. Neither does it have such a function assigned by selection. Even if I had chosen it based on size, it cannot have been truly selected even as a single token unless it was tested such that it would have been discarded if it was found unable to perform the desired function. We can imagine a proxy selection, for example, by weight and hardness. But even this would only qualify as selection if there was in turn an empirical basis for the selection parameters.[24] Finally, the rock was not performing a function which most other tokens of its type (if it has one) have frequently been observed to perform. Thus, while the rock

---

[23] The rock is performing a crude function only if the intentions of the thrower are premeditated and explicit. If they are, the rock could be considered an improvised tool performing a function inasmuch as it contributes to the goal. It would still lack any assigned function.

[24] We could imagine such a case, where a series of tests have been conducted to determine the minimum mass required to break the type of window in question when thrown at a given distance with a given force. If a rock were really selected on the basis of this data (meaning it was tested and would have been rejected had it weighed less), then it is possible to imagine that the rock could then be assigned the function. We would, after all, have good evidence to expect the rock to break the window under these conditions.

can perhaps be fully expected to break the window, this expectation is unconnected with the assignment of any function to the rock.  What is different about the expectations of function involved in assigned function?

If I take a glass cutting tool to the window, rather than throw a rock, I not only believe that it will cut the glass, but also that it is supposed to.  Indeed, I believe that it will primarily because I believe that it is supposed to.  The glass cutter is different from the rock in that it is represented (e.g., by its labeling and packaging) as having a certain history, which has – or should have – invested it with the desired functional capacity.  On this basis, we take it to have an assigned function via all three modes: we believe that it was designed for function, tested and selected for function (as a type and to some extent as a token), and that most examples of this tool are used and perform likewise.  With such a strong basis for function assignment, if it fails to perform it is liable to be accused of malfunction.

What is curious, and key for understanding the meaning of the normativity implicit in malfunction, is that even though the function came to be expected of and thus "assigned to" the device primarily by means of the intention of its makers (or at least beliefs about this), once assigned, it is the device *itself* which comes to bears primary responsibility for function performance.  Likewise, it is the device itself which we blame when it fails to perform.  Therefore, in malfunction it is the device which is held liable – not its originators (at least not exclusively).[25]  Thus, if the cutter fails in its task, we

---

[25] We may of course blame the makers as well.  But only in the most egregious cases would we blame *only* the makers and *not* the tool.  Notably, these are just the cases where there is insufficient evidence to

would say that the tool is poor, that it is a bad cutter. In general, whatever the source of function assignment, whether design, selection or typology, the assignment of function involves a subtle shift in the location of responsibility. In malfunction the locus of failure becomes the part, rather than the relevant makers. We say that *the part* failed, *the part* is bad – not (or not only) that the designers or manufacturers failed in their intentions.

It turns out, then, that explicit intention is not in fact the direct foundation for normative function judgments. Intentions are a major source of function assignment in artifacts for obvious reasons. But once that function is assigned, its source becomes less relevant. Once assigned, the item itself is seen as bearing the function – that is, after all, the meaning of function assignment. This is why typologically-assigned functions, as well as selection-assigned functions in organisms, are intuitively seen as eligible for normative function evaluation even without being able to appeal to any intentional etiology. They are, in fact, eligible in precisely the same way as designed and selected functions in artifacts are.

The normativity of malfunction is thus essentially a mental projection of functional obligations onto the functional part itself based on our own conception of its assigned function. This is true regardless of the mode of function assignment. In all modes of function assignment, whether from design intention, selection, or typology, the

---

reasonably expect (and thus assign) the function in the first place. For example, if the package marked "glass cutter" contained a plastic spoon, we would accuse the providers of fraud, not blame the 'tool' for malfunction – because there is no good basis on which to assign the item this function at all (neither design nor selection nor type assignment apply).

justifiable expectation that it *will* function becomes, in the human mind, the belief that it *should*.

If this is so, then Davies need not have been concerned that a normative concept of malfunction necessarily transgresses naturalism – but we must understand the nature and source of the perceived norms. Non-function becomes malfunction because *we* contribute the prescriptive normativity ourselves, when we assign a specific functional role to the part. The prescriptive normative dimension is contributed by our own psychology when this epistemic expectation is "transferred" to the functional item itself, such that the item is imagined to be normatively obliged to perform its function. Thus, to take a design assignment example, "I expect this screw to properly adjust the valve lash" becomes "this screw is supposed to (is *itself* obliged to) properly adjust the valve lash". This same principle of "responsibility transfer" explains the prescriptive dimension inherent in malfunction no matter the mode of function assignment.

This account of functional norms may not satisfy Millikan's high standards; she requires a function normativity that is somehow independently real. Yet Davies' objections to this strong interpretation are compelling. It is deeply unclear what the "reality" of such normative standards for natural systems could even mean. Metaphysically, a psychological account such as the one sketched above would seem to be all we can hope for with natural systems; plus it has the added virtue of applying equally to artifacts. Moreover, this account of malfunction seems to comport with all of our ordinary intuitions of what it means to attribute malfunction.

**Intention and System Constitution**

Finally, in light of these significant refinements, we turn to reexamine one of the

less satisfying of the original GGC examples above: the case of the bullet-stopping Bible

or rock outcropping. Did the Bible, much less the rock outcropping, really perform a

function by deflecting the bullet away from the soldier? According to a general

contributions approach, this is possible only if the Bible and the rock are part of a goal-

directed system. But in Boorse's GGC, this inclusion seems to occur only *after* they have

made their fortuitous contributions. The problem is thus that the systems seem to be

constituted dynamically on an ad-hoc and post-hoc basis: anything, no matter how

unforeseen, that happens to serve a goal of a human at the center of the example, gets

absorbed into a "system" with him or her and attributed function within it.

A slight variation can make this difficult case even more problematic: instead of a

soldier in a war zone, who would doubtlessly have protection from gunfire as an explicit

goal in his mind, we can just as easily imagine the same scenario occurring with a hiker

in the woods and a stray bullet from a hunter. In this case, it would be difficult to say that

such protection was ever even an intended goal of the hiker, as it would be unlikely to

have ever occurred to him that he would need such protection. Yet it is clear that GGC

would not treat this scenario any differently – it would still insist that whatever protects

the person thereby serves a function as it contributes to their inherent goal as an organism

of survival. This shows that intentions are fully irrelevant – not simply with respect to

the role of parts within a system, but even with respect to the goals of the system or its

constitution. The only slim reed grounding this "system" is the presumed goal of the person to remain uninjured.

Notably, what we do not see here is any explicit intention for the use of the artifact or objects at issue. Recall from the above, notably the study of the standards of performance for engines and stereos, that what was seen to determine the goals of an artifact system was the intention of the user in their use. But with the Bible, there is no intentional use postulated relative to the function being attributed. The example is notable just for this reason.

A main strength of a contributions approach is that it allows systemic elements to be attributed function even without direct intention, such as with the accidental nut in the engine, or the unknown mechanism of the yeast in brewing. This avoids messy etiological issues while highlighting what function performance seems to be centrally about. But while a part within a system need not enjoy an assigned function to perform a function, an inanimate system cannot be goal-seeking by accident. An artifact or inanimate object has no goals of its own.[26] It is only goal-seeking to the extent that it joins with and participates in seeking the goals of another system which *is* itself inherently goal-directed. The question then is, what does it mean to "participate" in this goal-seeking? Boorse holds that in this respect serendipitous contributions to the goal are no different than intentional ones.

---

[26] Boorse does argue that some artifacts, such as thermostats, are goal-seeking in and of themselves based only on their behavior. However, we have already seen how error-prone a strictly cybernetic account of goal-seeking can be. Furthermore, even such systems only become goal-seeking in their limited fashion via intention.

But only in the latter case is the item brought into the system *before* the goal is actually attained. In other words, only with intention is there a system including the object which is goal-*seeking*. In cases of accidental contribution to a goal, like the rock or the Bible, the items were not by any account part of any goal-seeking system – they simply received credit for having happened to contribute to the goal of an independent system, and were thus retroactively characterized as being part of that goal-seeking system the whole time. But the item is either part of a goal-seeking system *while* the goal is in fact being sought, or it is not. It cannot simply become part of a goal-seeking system by having happened to help reach a goal for a system of which it was never part until, ostensibly, that moment. [27]

There is only one phenomenon capable of recognizing the goal being sought and recruiting inanimate objects to aid its actual pursuit, thus constituting a true goal-seeking system: some form of conscious intention. Only by joining the system through intention can any inanimate, external object participate in goal seeking, and thus enjoy function attribution by contributing to the goal in question. What makes the Bible and the rock case problematic is the lack of intention. Even if we assume that the soldier did have protection from bullets as an explicit goal, it remains unclear how the Bible or the rock outcropping were part of any system participating in this goal without the benefit of any intention on his part that they do so.

---

[27] The problem of system identification here is in a sense formally similar to the selection-class problem encountered earlier with proper function. In that case, some way of identifying the type to which the object belonged independent of its successful function performance was needed, otherwise there could be no defective tokens of any type. In the present case, we need some way of identifying system membership other than successful contribution to the top-level goal.

In short, function in artifact systems is assigned at its top level by human intentions. Or, put another way, if the object is thought of as being absorbed into a goal-seeking system *with* the user, then that affiliation can only come about via intention. Given this top-level goal, accidental functions within the artifact system itself are possible. This is why the nut in the engine is performing a genuine function. But it is not possible to have a human-artifact goal-seeking system itself constituted by accident. At the top level, an artifact system participates in goal-seeking only through the intention with which it is put to use – and in no other way. To neglect this would compel us to attribute function to innumerable objects under innumerable circumstances, just so long as some serendipity can be ascribed with respect to some goal of some organism. Indeed, far from a mere detail, this would make function nearly ubiquitous: few objects on earth are so isolated that they cannot be said to benefit the life goals of some organism through some causal chain.

This principle of inanimate external objects having goals only via intention also helps make better sense of other organism-object systems, such as the rabbit hiding in a bush, or plants being watered by rain. These sorts of cases represent significant sources of promiscuity of function attribution in GGC, as we saw earlier. But in light of the above, it becomes clear that the organisms are not 'using' the objects in the requisite intentional sense that would convey goal-directness to the objects. There is a qualitative difference between animal behavior involving an external object (e.g., a rabbit hiding in a

bush) and a human or higher animal deliberately using an artifact.[28]  The rabbit's use of

the bush (probably) does not convey function to the bush – the function remains within

the animal's own behavior, by which it seeks to favorably align itself with its

environment; the bush is simply part of the context in which it does so.  Rabbits are not

tool users, and neither are plants; rabbits do not use bushes, nor do plants use the rain.

No function performance is involved for the bush or the rain in these cases.  But a

human's use of an artifact or object does convey function performance to that item. This

difference is entirely due to the deliberate, premeditated intention with which the object is

put to use.  Intention is what causes the system, consisting of the user and the object, to

arise and cohere.  Intention alone can thus impart function performance to inanimate

objects.

  Finally, note that the above discussion pertains to inanimate objects external to

organisms, and not to items which are physically part of goal-seeking systems such as

organisms.  The issue above is what makes something part of a goal-seeking system *other*

*than* actual physical integration.  Skeletal bones, for example, are not external objects but

internally integral parts of the goal-seeking system that is the animal.

---

[28] There are cases of non-human animals which may also qualify as intentional use – for example, chimpanzees have been known to use small sticks to extract termites from a nest.  Such cases could certainly qualify as intentional use, so long the animal is being somehow deliberate in its use of the stick.  Other cases, such as the insect example Boorse describes, clearly would not qualify.  By my lights, the animal must be able to represent to itself its use of the tool to seek the goal in order for the tool to properly be considered part of a goal-seeking system.

**GGC Revisions Summary: Assigned Contribution Theory**

The main insight of GGC, and contributions approaches in general, is that the primary characteristic of function is making contributions to an end. This is what constitutes function performance. But despite this promising core, GGC left much room to improve. With several important tweaks and elaborations, a more defensible, flexible and reliable contributions account of function can be had (and an account superior to any of those studied to date in important ways). These revisions, as described above, begin with the introduction of a distinction between performing a function, and having been assigned a function. Making function assignment a distinct concept from function performance enables us to make far greater sense of the multiple confusions previously seen when analyzing cases such as the accidental nut in the engine.

The rational psychological expectation of function performance, on which function assignments are based, seems to come about via three distinct modes: by having knowledge of the design origin of a thing; by having knowledge of the selection history of a thing (either as a type or a token); or by having knowledge of the typical contribution of a like thing in like systems.[29] Function can be assigned to (i.e., expected of) a part by any or all of these modes, with increasing degrees of strength possible when two or more of these modes overlap. Mass-produced artifacts, for example, typically have functions assigned by all three modes at once. Prototype devices will at least have function through

---

[29] It should be noted here that these three modes are constrained by the stipulation of a contributions approach to function. Thus, all three modes are tied to actual or intended instances of functional contribution. Another apparent mode by which function performance might be expected of an item is the nature of the item's actual properties. However, appealing directly to properties is at odds with basing function on contribution effects. An approach based on properties is the subject of the following chapter.

a design assignment, and, if successful, a token selection assignment as well. Natural organisms will typically have both selection- and typologically-assigned functions of their parts.

Introducing the concept of function assignment also allows us to make better sense of cases where a function is expected but not performed, whether because of a lack of opportunity or need, or a defect in the item. The former are cases of potential function; the latter malfunction. In both cases the function is assigned, or "had", but not actually performed. Thus we can understand malfunction as the disappointment of the expectations resulting from the mode of function assignment. For example, we expect function performance from the parts of artifacts because we believe them to have been designed, selected and to be typical of their type. Their internal functions having been thus assigned to them, via our reasonable expectations, we are liable to accuse their parts of malfunction when these expectations are disappointed. Malfunction is thus merely a projection of human expectations; yet once function is assigned, we transfer those expectations for function onto the parts themselves. This in turn accounts for why malfunction seems to be strongly normative regardless of whether the part in question is natural or artificial.

Finally, the constitution of teleological systems involving artifacts and external objects was refined such that these items are only included in the system via the conscious intention of their users. This eliminates the strong propensity towards promiscuous function attribution we observed in GGC without this requirement. Under

that arrangement, even the rain could be attributed function in relation to the contribution it makes to the life of plants.

I will refer to this heavily revised contributions approach as the "Assigned Contribution" theory.

## Critique of Assigned Contribution Theory

As described in detail above, the Assigned Contribution theory is superior to GGC in a number of ways, and also is not liable to the basic criticisms encountered with etiological theory in chapter 2 or systemic effects theory in chapter 3. Nonetheless, it does have certain weaknesses. Two of the more substantial ones will be explored here: blindness to the important interaction of form and function characteristic of functional parts, and the ongoing lack of a fully satisfactory account of natural teleology.

### Function and Properties

A contributions approach attributes function to the parts of a goal-directed system based solely on their contribution to the end goals of that system. We have examined what constitutes a goal, what constitutes a system, and what constitutes a contribution. But what, precisely, constitutes a part? And how does this relate to the nature of the properties it may have?

For example, there seems to be a difference between a complex, specialized macro-part like a heart or a carburetor and the lowest-level elements of which each is composed. Carbon atoms are found in all organic matter, but the function of carbon as a type is not to compose organic matter. Iron is found in most carburetors, but Iron as a

type does not have the function of composing carburetors. The difference is most easily seen between what we usually think of as the parts of systems, and those low-level elements which compose its structure. But the distinction truly consists between parts that seem to have function inherently as part of their type, and those that have function at best contingently and only as tokens. If this is an important distinction, it is not clear that a contributions approach, including the more sophisticated version developed here, can account for it. A closer study of another example may make this principle, and what is at stake in it, clearer.

Chlorophyll is a complex organic molecule which makes photosynthesis possible in plants. This is its essential contribution to the life goals of plants, a role which furthermore it seems to have been naturally selected (via the plant species which produce it) to serve. Promoting photosynthesis is thus the function which chlorophyll performs, and which, on the basis of selection and typology, it is assigned (i.e., expected) to perform. Chlorophyll itself is anchored by a single atom of Magnesium at its center. Magnesium, and its precise atomic structure, make an essential contribution to the structure and function of chlorophyll, and through it to the life goals of plants. Thus, Magnesium also would seem to perform a function in plants, and to have this function assigned to it, in the same way and to the same degree as chlorophyll's is.

But there is a large difference between saying that chlorophyll has a function in plants and saying that Magnesium has a function in plants. Notably, chlorophyll is not found outside of the plant life in which it plays its role, while Magnesium clearly is.

More importantly, for related reasons, chlorophyll has the form that it does only *because* of the role that it plays within the plants which synthesize it.  As evidence of this, consider that the relevant genetic code was selected precisely because it produces chlorophyll.  Chlorophyll would not even exist otherwise.  Magnesium, on the other hand, existed long before chlorophyll or any plant life was even possible. We might say that Magnesium was selected from among available alternatives to anchor chlorophyll; but this selection did not result in the creation of any Magnesium atoms – only in chlorophyll molecules arising around some of them.  Chlorophyll "is what it is" – has the properties that it does – because of its function and vice-versa.  What chlorophyll is is thus virtually inseparable from its function.  In principle, the same is true of hearts and carburetors as well.  Yet this is not true of Magnesium since it exists as itself outside of whatever functional role it may occasionally play in plants.

To a contributions account, Magnesium enjoys a function attribution indistinguishable from chlorophyll's (the details of their respective contributions excepted).  And yet the properties of chlorophyll are shaped and constrained by their function – in short, customized to their role in plant life – in a way that Magnesium is not. Chlorophyll's function is essential to what it is; Magnesium's function is incidental to what it is.  This seems like a significant, even crucial, distinction closely connected with the idea of function.  But it is one which a contributions approach, by its very nature, is unable to recognize – all that matters to it is that both contribute to the life goals of plants.

In general, we seem to need some way of recognizing the difference between a part which as a type only exists in the form that it does because of the systems to which it belongs and parts whose types exist independently of any role they play in any system. The former are exclusive to the systems in which they function; the latter are not.  The form and function of the former are interdependent, while the form and function of the latter are merely contingently related.  For the latter, their form enables their function (as in Magnesium anchoring chlorophyll) but their function plays no role in determining or constraining their form.  In the former, the interaction goes both ways.

Finally, note that this is closely related to, but not the same as, the etiological concept of function.  In etiological function, function is attributed to the type if and only if the existence and/or placement of the part is due to the effect that it has there.  The function/properties connection explored here is not about origin, however.  It is about the two-way connection between form and function, independent of origin – though in practice these are often related.

**Lack of a Firm Account of Natural Teleology**

As noted earlier, a contributions account of function relies on first having an independent account of the goal-directedness of the systems in which functions are found.  The cybernetic account of goal-directedness which Boorse appeals to was seen to enjoy at best limited success.

The account above of the goal-directedness of artifacts and other objects is a substantial improvement – but it makes them reliant on the goal-directedness of their conscious

intentional users. Arguably, the deliberate, conscious intention of human beings requires no further defense as teleological. If anything is teleological, surely it is conscious premeditated intention; indeed, this is typically taken as the paradigm of teleology. To the extent that there are problems with consciously held goals, it is beyond the scope of this work to consider them. As we saw, artifacts become part of goal-directed systems, at least in Assigned Contributions theory, though being intentionally used.

But what of non-conscious natural functions – namely, of organisms, including humans? In order to attribute function within organisms – for example, to their physiology and their non-conscious behaviors – they must first be characterized as teleological systems. The remaining question, then, is what makes natural organisms goal directed? What grounds their teleology? A cybernetic account was seen to be less than satisfactory, yet some serviceable account is still required.

## Conclusion

A contributions approach, of the three approaches to function studied to date, seems to strike the right balance in terms of teleology, which has shaped up as the greatest challenge for theories of function. In a contributions approach, teleology informs the content and direction of function. But teleology is not expected to directly account for each and every instance of function, as with etiological approaches; and neither is it entirely ignored as irrelevant, as with systemic effects function. However, a bare-bones contributions approach such as Boorse's GGC leaves too many loose ends – among them an inability to distinguish between what an item was made for, or good for, and how it

was put to use, as well as lacking an account of malfunction. With external objects like artifacts, the goal-directed systems were also liable to be defined loosely, and even post-hoc, leading to extremely promiscuous function attributions. Moreover, Boorse struggled to articulate what made systems, whether natural or artifactual, goal-directed, a prerequisite for any contributions account.

GGC can be significantly improved with two substantial modifications: First, differentiating actual function performance from the human expectation of function performance. Assigned function explains why we expect function performance from some objects and not others, while still recognizing that this expectation does not in itself constitute function performance. It is the possibility of their divergence which explains malfunction. Focusing on expectation also accounts for how designed or selected things are different from accidents. The second refinement is to reform the constitution of goal-directed systems involving external objects such that they are constituted by intention in use and not serendipitous accidents. This curtails the attribution of function to external objects and prevents extensive promiscuous attributions. These refinements and others lead to what I have called the Assigned Contributions theory, which I consider the strongest and most flexible overall of those examined to date.

Nonetheless, some problems do remain, including the ongoing need for a satisfactory account of natural non-intentional teleology, and a perceived need to distinguish between parts whose form and function are deeply interrelated and those whose are not. These remaining two difficulties, though they seem unrelated on the

surface, in fact together point the way to another distinct and relatively unexamined

approach to function: through the nature of the properties of the parts of a system, which

can in turn mark the system itself as teleological.  Thus we are led to consider a final,

tentative and entirely novel alternative account of function – one based not on etiology

and not on causal effects or contributions – but only on the formal properties of systems

and their parts.

CHAPTER FIVE

THE PROPERTIES APPROACH TO FUNCTION

**Introduction**

From the foregoing chapters, it has become apparent that the basic task for a

philosophical theory of function is to articulate answers to essentially two deceptively

simple questions. First: *What, if anything, makes a thing 'for' what it is does?* And

second: *How do we handle cases where a thing does something it is not 'for' or does not*

*do something it is 'for'?* In these two simple questions lie the roots of most of the

convoluted issues we have examined at length. The first question pertains to the nature

of the innate relationship between function and teleology; the latter to the possibilities of

imitations of function and of malfunction.

As described at length in chapter 2, etiological theory holds that a thing is 'for'

what it does *only* when the presence of the thing is causally explained by what it does –

for example, a device is designed and manufactured expressly so that it might perform its

function, and the effects which explain the selection of the traits of an organism are their

functions. But this proved problematic, for many things seem as if they are for what they

do without having such origin, while others have such origin without doing anything

useful; moreover, knowing origin is much harder than knowing what something is doing.

Reacting to such failures, we saw in chapter 3 that systemic effects theory held,

essentially, that a thing simply *never* is 'for' what it does – with the possible exception of

249

consciously intended functions, which it declined to treat. But this robbed function of any substantial meaning, and allowed almost anything to count as function; moreover, it had no obvious application to the paradigmatic context of functions, artifacts. Most recently, we saw in chapter 4 that contributions theory, at least as developed there, held that a thing is 'for' what it does when we have reason to expect that the thing will make a contribution to the ends of the system to which it belongs; but nonetheless things can still make such contributions even when we do not expect them. The former it called assigned function, meaning that the thing is indeed 'for' what it does; the latter it considered simply the performance of a function, without the item being for what it does. But contributions theory is complex, and requires the independent establishment of the system as teleological so that there are ends to be contributed towards. For natural systems such as organisms, this serious challenge is still as yet unresolved.

While etiological theory was ultimately found unworkable, it did have the strong virtue of directly and succinctly addressing this central question with a simple principle. A properties approach offers an entirely different but similarly succinct way of answering the primary question of what makes a thing for something – a way that moreover may also allow us to simultaneously identify systems as teleological. This presents the possibility of a stronger and more unified answer to the question of when a thing is for what it does. The approach in this section does not have a parallel in the existing function literature, though it inherits many of its ideas from emergence theory. Largely as a result, the approach in this section will remain more tentative and schematic than

those of the previous three chapters. But the possibilities inherent in the approach make it worth pursuing.

**Properties and Systemic Context**

What is it that makes a heart a heart? We have encountered this question before, when we needed to know what made a token qualify as a member of the type, particularly if it could not perform the function typical of the type. Boorse favored a morphological view, which allowed him to include sub-performing tokens as members of the type. Millikan, on the other hand, seemed forced to take a functional view, which complicated her story of malfunction, since a dysfunctional heart could no longer count as a heart at all. But how we recognize a token tells us something fundamental about the definition of the type. So let us pose the question anew, straightforwardly, on a pragmatic, even naïve basis.

What makes a heart a heart? A heart is an organ that pumps blood. But that is only to state what it typically *does*, not perhaps what it *is*. If the heart were removed or the organism to die, the heart, we would ordinarily say, remains a heart. If a heart remains recognizable even on the counter of a butcher shop, actually pumping blood cannot be what makes it a heart. Clearly in this scenario all we can appeal to are the physical, formal properties of the object. We see that it is a heart because of the configuration of the ventricles, the arrangement of its veins and arteries, and the nature of its tissues. Of course, the more anatomy and cardiology we know, the easier this recognition is and the more precise we can be. But even at a gross level, we recognize a

thing as a heart because of its physical properties – indeed, this seems to be what "makes" the thing a heart. If these properties are how we recognize a token heart, what does this mean for the type "heart"? Why are these properties essential to the concept of a heart?

Let us imagine that we do encounter a heart out of context. To identify it, we first look at its gross morphological characteristics – the presence of muscular pumping chambers, valves which regulate fluid flow, arteries and veins which receive and expel blood. In addition, there are many more subtle properties which we cannot immediately detect – such as the configuration of the muscle and nerve cells – all of which are necessary for the heart to function. Note how each of these properties present themselves not only as features of the object itself, but as clues to the kind of system to which it belongs. The arteries and veins testify to further connections; the valves to the flow of blood; the nerves to a nervous system with which they interface. We see not simply what is there as properties of the organ, but how these indicate what is not there – what is missing but must be there (and presumably was) to give identity to the organ. We see in these properties how they connect with and relate to the system to which the heart conceptually belongs – even when physically removed. A heart is for the pumping of blood in a cardiovascular system.

Conceptually, then, though obviously not physically, there can be no such thing as a heart in isolation. A heart is a heart not in itself alone but as a meaningful part of a functional system. A heart is an organ that is integral to the cardiovascular system, and

as such only "makes sense" in the context of this system.  It cannot be properly defined without reference to the cardiovascular system, or indeed the entire organism to which it belongs.  A heart is what it is because of its place in the whole of which it is a part. "Heart" is thus a *contextual* concept – a heart only "makes sense", only "has meaning", in the context of a cardiovascular system in which it plays an essential role.  Just as a language provides the context within which a word can have meaning, a cardiovascular system provides the context within which the concept of a heart has meaning.

This contextual, properties-based perspective seems generalizable to any functional item.   In this view, functional objects are inherently contextual – with the crucial caveat that the item be somehow *specialized* for its task.  In contrast to the other three theories of function examined previously, this properties approach picks out functional items from among all items not according to their origin or effect, but according to their nature *as* objects.  It distinguishes those which have no inherent contextual dependence or relationships from those which do.

In fact, in the biological realm at least, this is hardly a novel observation.  As James Blachowicz points out, "Most philosophers who have sought to identify the distinctive principles of living things, including Aristotle, Kant and Hegel, have noted the fact that the organs of living things have no real meaning or function apart from the organisms of which they are parts. An oxygen atom is a self-sufficient, individuated physical entity even when not a component in a water molecule, but a hand or heart or eye is not, when taken apart from the organism for which it serves a specific function."

(Blachowicz 2012, 264). In the properties approach to function, we expand this principle from biological organs to functional items in general.

This conception of function picks out certain kinds of highly specialized entities, like hearts and carburetors. While these sorts of parts undoubtedly function by making contributions to the goals of the systems of which they are a part, there are profound differences between this properties approach and the contributions approaches of the previous chapter. For a contributions approach, function performance consists entirely in causal contribution to the goal of the system. In contrast, a properties approach looks not directly to the causal contribution itself, but to the properties of the part which allow it to make the contribution. In a properties approach, a specialized part "is what it is" only because of the system to which it belongs and which (in a general sense) it is "for".[1] The properties of a part make it what it is (and what it is for) insofar as they attest to the context to which it belongs, and the role that it plays there.

It is worth contrasting this identification of the function of a part via its properties with the way that function was assigned to parts in the Assigned Contribution theory of chapter 4. There is a superficial similarity in that in both theories function becomes tightly bound to the part, such that the function becomes expected of it. But there are profound differences as well.

---

[1] Blachowicz terms such parts "Self-Differentiated". The coinage refers to the way in which the constraints of the overall organism enable the differentiation of its own parts (as will become clearer below). However, the use of this term could be interpreted to imply that the *origin* of highly differentiated parts is always the result of such constraints. On the contrary, in contrast to etiological theory, a central premise of a properties approach is that the origin of the functional parts is irrelevant, and that their function and purpose is inherent in their properties alone. I will therefore refer to these functional parts simply as "specialized" rather than Self Differentiated.

Assigned Contribution theory supplemented a basic contribution principle by articulating why it seems to us that some things really are "for" what they do, while others make the same contributions without being "for" what they do. It postulated that such assignments are made by us psychologically based on our reasonable expectations of function performance for some parts due to our knowledge of precedents of or design intentions for the function. But because contribution is still the coin of this realm, any entity which is seen to make contributions to the system is still considered functional. This includes specialized objects as well as those that are entirely unspecialized. We saw this in the example of chlorophyll towards the end of chapter 4. Because the chlorophyll molecule makes a vital contribution to the goals of plants, it has a function in plants – and indeed, has that function assigned to it through our awareness of its selection history and its typicality. But if this is so, then by the very same logic the magnesium atom central to the molecule must have a like function and assignment. Yet chlorophyll is unique to plants, while magnesium is not. Moreover, the structure of chlorophyll is constrained by its function in plants, while the structure of magnesium does not depend on its role in chlorophyll. Indeed, magnesium is more common outside of chlorophyll than within it, and of course magnesium existed long before chlorophyll or plants did. Thus, chlorophyll's fit within the cellular biology of plants seems to make chlorophyll customized for plants, while magnesium's fit within chlorophyll does not seem to make it for anything. Chlorophyll is what it is due to its function; magnesium just is.

In spotlighting specialized objects alone as functional – indeed, denoting specialization as constitutive of function – the properties approach is making a new distinction, one which, as we saw at the conclusion of the last chapter, a contributions approach is unable to make. This new approach is based on a fusion of identity, purpose and function: A functional thing "is what it is" due to its function in a specific context; as such, the thing is not even comprehensible as what it is outside of that systemic context. Specialized parts are what they are due to their role in the systems in which they are found, unlike non-specialized parts which are what they are regardless of systemic context.

Loosely speaking, a properties approach to function can be thought of as looking for the connections in an overall pattern, like pieces of a jigsaw puzzle. The space defined by a missing piece constrains what pieces goes there – only one with the requisite properties (in a puzzle, in terms of both shape and image).[2] We can say that a piece that fits is "for" that space, and only that space. In the same way, a specialized part is "for" the functional roles it fulfills, and only those roles. Thus, a screwdriver is the perfect complement, or link, between a human hand and the turning of a screw. This is the function it has. It's shape, its grip, its tip shape, and its materials all attest to this. It fits the space defined by the pieces it connects with. Outside even the potential context of a

---

[2] While helpfully evocative, the disanalogy in this simplistic metaphor is that the constraint is provided merely by a collection of other pieces, and as pieces they are all on the same logical level. Strictly speaking, there is no higher container system involved, of the kind detailed below. Moreover, while puzzle pieces are distinct, specialized parts may be refinements in the form of an existing 'piece'.

circulatory system, a heart is merely bulk biological matter, with meaningless properties. Inside one, its complex properties find meaning and the object becomes recognizable.

**Container Systems**

The above articulates the concept of specialized parts. But what makes a part specialized, rather than an object with essentially arbitrary properties, is its belonging to a system in which it has function and meaning. What then of these systems? These can be thought of as the systems which contain the specialized parts, in the sense that a whole contains its parts. But more is at stake in these than mere integration or composition. A container system not only comprehends the part but presents the context within which the part becomes comprehensible. In this context, parts can also undergo further refinement, attaining new features which themselves become more specialized for their own particular role (e.g., paws supporting the action of the legs in running, or becoming specialized into hooves). There is thus a nesting of general to specific roles in specialized parts and their container systems, a hierarchy which seems unique to functional relationships.

Container systems provide a platform, a context, on or in which the specialized object finds a fit, and becomes functional. But it does this by presenting constraints on the kinds of parts that *could* become meaningful within it. The nature of the container system sets limits around the kinds of things that would prove valuable within it. In a sense, the container system provides the problem for which specialized parts can become solutions. We saw this in the example of the heart, which was seen to gain its identity

through its role in the circulatory system. The circulatory system is the context within which the heart becomes comprehensible as having a purpose because of the way that the heart's properties fit the needs of the system.

This container system influences the nature of the parts which can become part of it, and find meaning within it, by presenting a platform for the development of ever-more specialized parts which contribute to the overall system. But container systems can also nest within each other, for the circulatory system itself is a highly specialized part of the overall organism.

In biology, we can think of one the most generic and basic container systems as the "gut tube": a primitive system which ingests matter and digests it for energy. Once this container exists, it then presents multiple opportunities for specialized parts which could be fit within it to aid its activities. For example, the gut tube may admit of parts which help it to digest more kinds of matter, or of parts which allow it to move around the environment, or better sense the environment, or avoid being consumed by other gut-tube based organisms. In short, all of the pressures which natural selection operates on are here expressed as opportunities for parts to take up roles within the gut tube container system. Moreover, there is a hierarchy of such parts as they are organized into systems, each of which can be considered as a container system in its own right, with its own set of "job opportunities" and problem definitions, all of which allow highly specialized parts (and only such parts) to be integrated into the system.

As Blachowicz described it, the gut tube is "a primitive whole whose distinctive properties could then influence subsequent top-down differentiation of its internal parts (through Darwinian principles, of course). Thus, once an organism with a gut tube was formed, mutations of its internal structure that provided enhanced chances for feeding that gut would be favored (not eliminated) over those that did not. The differentiation of subsystems relating to locomotion, sensory detection of objects, protection, etc. were all selected because of the higher-level structure that is the gut tube itself." (Blachowicz 2012, 264).  The gut tube (as the ideal of a container system) demonstrates how container systems constrain the properties of the parts which become part of it while at the same time giving meaning, identity and purpose to those parts.

But note that this gut tube wouldn't have served a purpose without a metabolic system in its cells to support.  If the gut tube did not provide material required for the metabolic activity of the cells, or if cells did not require such an input, the gut tube would not have served as a container system at all.  Here the details of cell biology become implicated, as cells themselves are container systems in their own right, and we are led further into the biological details.  Organisms have long series of nested container systems and their respective specialized parts: the heart is for the circulatory system, which serves the gut tube (apart from which the circulatory system may have no meaning), while the gut-tube is in turn for the replenishment of the metabolic system of the cells, and so on.  The teleology is nested; there is container constraint and part specialization at every level.  Identifying distinct container systems in an organisms is not

necessarily a precise exercise and their interactions are not a neat functional stack, for

many container systems interact mutually and have multiple dependencies. The eye, for

example, serves the gut tube container system by helping to identify food; but nutrients

from the gut tube also nourish the eye. Moreover, the eye has its own internal levels of

container systems, such as the retina (which even has its own circulatory system).

Ultimately it is the job of physiology, not philosophy, to map out these relationships and

hierarchies; the biological and schematic details are of less importance here than the

underlying principle of container systems and the way that they constrain the properties

of the parts which will "work" within them, if the parts have suitable properties.

The container system thus does much more than merely contain the specialized

parts – it is what allows them to be recognized *as* specialized. In most cases, it is also

what allows the specialized parts to actually *arise*, whether through natural selection or

intentional design. But the nature of the *origin* of the specialized parts is not what makes

the parts specialized. This is just how a properties approach differs – profoundly – from

an etiological approach. Specialized parts are specialized by virtue of their formal

(typically, spatial and physical) properties, and how they fit into a given container system

which gives them functionality (whether or not that also played a role in bringing them

into existence). They are not specialized simply by virtue of having been generated a

certain way by or through the container system. Neither are they specialized simply by

virtue of being contained in just any system, as magnesium is contained by chlorophyll,

and electrons are contained by atoms, yet these components are not specialized. They take on no new meaning or identity when they belong to such systems.

A container system constrains its parts, but so does any system, almost by definition. A container system also constrains the properties of the parts which can join the system – but, again, so do many generic systems, like molecules and atoms. What makes a container system unique is that it constrains the nature of the parts such that the parts have a new and unique identity – they become recognizable as a thing whose identity is joined to their function. The parts of a container system become new kinds of entities, not merely bulk matter objects, through it. Indeed, for them to be recognized in this way *requires* positing the appropriate container system. A heart requires a circulatory system to be a heart, and a nut requires positing a bolt to be a nut. In this, the container system is unique – so unique that it constrains its parts in this way even when they are not physically a part of it, just as we recognized the heart through its own properties alone, or would recognize a stray nut on the sidewalk. The container system is required as a context, but the influence of that context is so pervasive that it can remain imagined rather than actual.

The concept of a container system is thus essential to the properties approach, because it is only within such a system that specialization becomes possible. Note also that even if the specialized parts appropriate to it have yet to arise, we can still characterize a "container system" as such if it has the *potential* to accept specialized parts and make them meaningful. To do so, this "overall" structure must be logically "higher"

– more general than – the specialized parts which it can come to serve as the container

for. The container provides the initial opportunity for specialization by providing the

more general platform on which specialization can occur, even if it has not yet.

While many of the examples given above have been biological (as that is the more

challenging area), the principles above apply equally to artifacts. For example, the

container system of the car presented opportunities for the development of the engine,

while the container system of the engine presented opportunities for the development of

carburetors and valve adjustment screws. Unlike organisms, human intentions shape the

ultimate goals of artifactual container systems at their highest level. However, as with

any container system, the teleology of the parts is based on their close fit with the system.

Those intentions, therefore, as essentially a form of origin, need not be directly appealed

to: we can tell the purpose of a part from its properties and the way that these fit into and

find meaning in their corresponding container system. Artifacts do differ in one

important factor: the number of nested levels they present will generally be smaller.

Because their container systems are directly imposed from outside, artifacts are not

nested to the same degree or in the same way as organisms. This is because there is no

corresponding system for the basic biological unit of the cell in artifacts.

**Container Constraint and Etiology**

While the container system does (by definition) constrain or limit the nature of the

parts which can join it and thus find meaning within it, as was noted above, this

constraint is not in itself etiological – that is, this constraint cannot be expected to

necessarily explain the *origin* of the parts themselves.  It is worth briefly exploring the impact of this principle.

One problem for the etiological approach, it will be recalled, was that while it appealed to selection for the origin of functional parts in organisms, it could not recognize the initial appearance of a part type as functional, for the initial appearance will not yet have been selected.  There was thus a disconnect in the claim to base function on selection if the function in fact preceded its selection.  For the properties approach, this is a non-issue because it does not base function attributions on origin.  The characterization of a part as specialized does *not* appeal whatsoever to its actual existential origin, either as a type or a token.  It is true that specialized parts are *typically* the result of either selection or design; but this is not what a properties approach is basing functional attribution on.  No matter the origin – which is to say no matter by what process the properties of the part were formed, or by what process the part became physically part of the container system – the nature of the system alone limits the kinds of parts (and their properties) which are or can be functional in it.  It is because of the demands of the system that its parts can have function, if they are suitably formed to provide it.

Indeed, on the contrary, probable origin is often deduced *from* the specialized properties.  Specialized properties provide prima facie evidence of the likelihood of specialized origin – the constraint of their properties by the end which they are so perfectly suited for.  We suspect such an origin only because the close match between their properties and this task/effect is so perfect that it seems unlikely that they are not

connected. Logically, however, they could be. Thus, with knowledge of the respective

origin processes typical of such specialized objects, we take it that the carburetor was

designed for the engine, and the heart incrementally selected for its pumping effect. But

we suspect or deduce the origin only from the highly specialized nature of their

properties. This is somewhat like finding a lock and a key that fits it – the natural

assumption is that the key was literally made for this lock. But, even if that somehow

happened not to be true – if the key were a natural formation, or the result of a quantum

fluctuation (and, crucially, even if we *knew* that) – we would still take it that the key is

*for* the lock. That conclusion as to its purpose is not a judgment about its origin, only its

ideal properties for this task. Any presumption of origin *follows* from this.

However, as this example indicates, there is a sense in which the container system

is a prerequisite for a part to become specialized: the container enjoys a logical, though

not necessarily physical or temporal, precedence. Undoubtedly, in the vast majority of

cases, this logical precedence is expressed via a temporal (etiological) precedence: thus

the gut-tube organism arises first, and then via random mutation and natural selection,

specialized parts such as eyes and legs arise within the organism and "under" the gut

tube, serving it. But a properties approach does not attribute function to these specialized

parts due to their having physically originated under the gut tube, but because of the

logical relationship between them. Due to the integral way in which organisms are

fashioned, the two situations are virtually identical. But examples with artifacts make the

situation somewhat clearer. With artifacts, typically a need is perceived, a design is

conceived, and an overall system with subsidiary functional parts is designed (e.g., car and engine, engine and carburetor). Thus, for example, a hammer is designed and produced due to the perceived need to drive nails.

But if a hammer were accidentally and randomly generated – one indistinguishable from a manufactured tool, complete with hardened steel head and graspable handle – we would still recognize it as a hammer. It would still be for driving nails, just as the random key is still for the lock it opens. In such cases, it is apparent that the container system (here, the user and his nail-driving needs) is not temporally or etiological prior to the specialized part – but it is still logically prior. The hammer could not be a hammer without them. Container systems are thus a logical prerequisite for their specialized parts. We might even go so far as to say that the container system is the "cause" of the specialized part *as* a specialized part, even if they arise physically independently of one another. It is by serving the container system that parts can become specialized.

Even in organisms, the difference between etiological or temporal precedence and logical precedence can be seen in cases wherein beneficial traits are known *not* to have been naturally selected due to these benefits. For example, the skull partitions which are essential for live birth were not selected for this reason, and we know this because the partitions predate the evolution of species that birth live young as opposed to laying eggs. But in terms of constraints and part specialization, things look different: while this feature was not the result of actual conditions imposed by a selection regime, the problem of live

birth does present constraints of a sort which skull partitions closely fit. A properties approach, therefore, has no trouble saying that skull partitions have a function, and indeed are for this function – even though their evolutionary origin is not connected to it. In this respect, a properties approach is more liberal than, and more intuitively appealing than, the etiological. However, the contributions approach does also come to very similar conclusions based on the benefit of the effect.

**Perfect Measure of Properties and Quasi-Specialization**

The main tenet of a properties approach is essentially that objects are *for* what they are particularly *good for* – they are for what their properties ideally suit them for. Indeed, "for" and "good for" become synonymous under a properties approach. Usability is purpose – the more suitable the use, the stronger the purpose. But some things are more usable for some purposes than other things. And some things can be used better for some purposes than other purposes. While the ideal of specialization is a part that is perfectly suited to its use, some parts can be used successfully even when they are not perfectly suited for their use – as indeed we have seen again and again in the previous chapters, from the nut adjusting the valve to the Bible stopping a bullet. A properties approach must be able to account for such cases. It can do so by conceiving an object's complement of properties on a sliding scale, from an ideal complement of properties for a use to a less ideal but still sufficient set.

To see how this might work, imagine a standard (flat tip) screwdriver used as a pry bar – a usage which occurs routinely, although it is not for this. A screwdriver is

ideal for driving screws due to the close fit of its properties to the task: from the way that the tip fits into the recesses in the head of a screw, to the way that the handle fits into the grip of a closed human hand, and the way that the shaft bears torsional loads. Aside from its overall shape, a screwdriver is suited to its purpose in less obvious ways. For example, its steel tip is hardened, so that it does not quickly wear under the repeated abrasion of metal-to-metal contact with the screw. But the tool's purpose is also marked by what it *lacks*. For example, the torsional stress which a human hand applies to the steel shaft (via the handle) does not require its hardening. All of these properties cohere to define the purpose of the screwdriver – they define what it is for by what it is good for: driving screws.

In contrast, a pry bar undergoes bending stress throughout its length, and might be braced against another object anywhere along its length. As a result, it is hardened along its entire length, as well as being larger in cross section and typically square in section rather than round. These traits allow more bending force to be safely applied to the bar, as well as for the bar to be braced stably against other objects. Moreover, a pry bar is typically curved to some degree so that it forms a fulcrum against any flat surface rather than requiring an available pivot point.

When a screwdriver is used as a pry bar, only a subset of the requisite properties are available, while other of its properties are unused. A screwdriver has a narrow metal tip, which allows it to fit in between objects to be pried apart. It also features a handle by which it can be comfortably gripped so that lateral force can be applied to it, and a

relatively strong metal shaft. But its shaft is entirely straight, meaning an external

fulcrum must be provided; moreover it is round (so that it can he held while turning),

making bracing against the fulcrum less stable. The fact that the shaft is not hardened

also means that much less force can be applied before the shaft will bend or break. For

all these reasons, serious mechanics have both screwdriver *and* pry bar sets in their

toolkit: "the right tool for the right job".

The screwdriver can be used as a pry bar only because it has a significant *subset*

of the relevant properties of a pry bar: a narrow end, a metal shaft, and a grip. If the

screwdriver were made of wood, or lacked a handle, or had any number of other crucial

differences from an actual pry bar, it could not substitute as one under almost any

conditions. But the lack of the *full* complement of properties of a pry bar (as well as the

presence of extraneous properties) means that the screwdriver is simply not "for" prying,

while a pry bar is. What a thing is for, then, seems identifiable on the basis of its

properties alone – neither origins (etiology via selection or design) nor actual

contributions to a system need enter directly into an evaluation of what a thing is *for*. An

item is 'for' its job when it has just the "Perfect Measure of Properties" (PMP) for the

task it needs to do. Lesser degrees of fit testify to what an object can be successfully

"used for" without being quite "for".

We can also see how this approach handles past cases. As noted in the last

chapter, Wright's accidental nut which adjusts a valve has the right thickness to achieve

this effect. But it also has six flat sides and a hole with internal threads, neither of which

are relevant to making this adjustment. Meanwhile, an adjustment screw has cylindrical

length and external threads which allow it to accomplish the additional function of future

readjustment. The screw thus has PMP for both jobs, while the nut has only an adequate

complement for present adjustment, and no applicable properties for future

readjustments. The screw is thus for its tasks, while the nut is not, although it is able to

perform one of them.

Wright also imagined a case where a soldier's pocket Bible saved his life by

stopping a bullet. As with the nut, due to its lack of design he held that the Bible did not

have the function of stopping bullets, however useful it happened to be in doing so. The

properties approach also finds that the Bible is not for stopping bullets – not because it

was not designed to, but because it does not have the ideal complement of properties to

reliably do so. Unlike modern body armor, the Bible does not cover much of the

soldier's torso; neither is it as hard and dense as armor, to help absorb as much of the

bullet's kinetic energy as possible. Moreover, the Bible, being a book, obviously has

manifold irrelevant properties for the task of stopping bullets – notably its division into

hundreds of individual paper pages, and the printing upon each one of them.

However, the Bible does have a relevant subset of properties which allow it to

stop some bullets in at least one area of the body. For the fortunate hypothetical soldier,

this was enough. The Bible was only able to do even this much by mimicking the

properties of actual body armor by being flat, thick and sufficiently dense. While its

form and properties do not comport perfectly with the demands of the task, as does that

of actual armor, the Bible can be used for this – under the right circumstances. Body armor is for stopping bullets not because it was designed to but because it has the Perfect Measure of Properties (PMP) for the job. Consider also that design, as we explored earlier, is no guarantee of success – designs can and do fail. It would be strange if design alone were enough to make an item for a task, even if it cannot accomplish the task. Properties, therefore, seem a more reliable basis than origin to say what the function of an object is.

In sum, an object with PMP for a given task is a *specialized part* with regard to that task. An object with merely adequate, but not ideal, properties for a task is a quasi-specialized part with regard to that task. But, importantly, such a quasi-specialized part is only such by imperfectly mimicking the properties of the specialized part that has a Perfect Measure of Properties. A PMP object has neither too few nor too many properties for the task, making it – and only it – truly "for" the task.

**Specialized Parts and Malfunction**

Any theory of function seems obligated to explain how a functional item can malfunction. The core notion of a properties approach is that a part with the Perfect Measure of Properties for a given role in a container system is thereby inherently "for" that task. The teleology of the part is bound up with its own properties. The account of malfunction in the properties approach would seem to follow from this principle, with the normativity of malfunction resulting from the teleology of the properties.

However, there is a serious problem in looking directly to properties for an account of malfunction. In any malfunction, a change in properties would seem to be implicated. The properties which served as the basis for the part having the function in the first place will have been materially compromised if the part is malfunctioning. Can the part still be for its function if the very basis for making that judgment has been significantly undermined?

One response to this problem would be to take a historical approach. Presumably, the part's perfect properties attested to its purpose at one point in time. We could argue that this fixes it purpose thereafter as well, with deviation from this baseline counting as malfunction. However, there are obvious disadvantages to such an approach, for it means that if the properties of a specialized part ever change, malfunction results. In an artifact, a part could thus never be repurposed or even improved. Worse yet, in organisms parts could not be allowed to evolve as a type, nor could they even change as tokens (e.g., baby teeth falling out are not malfunctions). The proper baseline for the functional properties of a part in organisms is also a difficult challenge.

Fortunately, a properties approach may be compatible with the expectation-based account developed for assigned contribution function, with an adjustment in the manner in which the function is "assigned". This is helpful because our expectations of function are more adaptable than the actual state of the properties themselves. In a properties approach, these expectations would result from a judgment that the item's properties represent – or *had* represented – a Perfect Measure for the task at hand. Observation of

properties thus take the place of knowledge of the design, selection or typology of the part, though these origins may in turn be implicated by a finding of PMP.  This means that PMP is somewhat analogous to an Assigned Function, while less than PMP is analogous to unassigned function.  However, there seems to be more of a spectrum built into PMP than there was with Assigned Contribution.

Of course, not all properties are equal when it comes to judging malfunction, for not all properties of a specialized part are equal.  Some may be indispensable, without which the object malfunctions; others may help but not be essential.  For example, a screwdriver with a missing grip is less of a problem than a fully worn or broken tip.  One can still use a screwdriver with a poor handle grip, but without a properly shaped tip the tool will not drive screws.  Malfunction is thus a relative term; the former is a modest malfunction while the latter is serious. Therefore, PMP itself is also a relative term: perfection is not a necessity, but an ideal.  There is a wide range of merely poor function between malfunction and full function.

**Summary of the Properties Approach**

In sum, in a properties approach the formal properties of an object account for what the part is for by showing what the part is *good* for, and indirectly indicating the systemic context in which it finds use.  Indeed, reference to a particular systemic context is seen as an inherent part of having a functional identity.  Specialized parts have the ideal complement of properties to be what they are for, and these properties themselves also testify to the context, or container system, in which they operate.  Container systems

provide the opportunity for parts to find a meaningful functional place in them.  These

container systems constrain the properties of the parts which can operate within them –

imposing what might be called "formal constraint" – but they do not necessarily explain

the actual origin of these specialized part.  Unlike in an etiological approach, in a

properties approach etiology is not necessary to determine the function of a part; their

own properties do this independently.  In fact, specialized properties provide the basis for

suspecting a specialized origin (i.e., design or selection) in the first place.

The properties approach described here, while still somewhat schematic, does

offer a number of compelling advantages over the previous approaches studied.  A

properties approach unifies teleology and function by looking to the properties of the

parts which can find a place in a container system.  It seems to provide a strong account

of property-based teleology that applies equally to natural systems and artifacts.  It offers

the principled simplicity of an etiological approach while avoiding the epistemic and

conceptual problems associated with it.  And its main benefit over a contribution

approach is how it integrates teleology directly into function attribution rather than

relying on an outside attribution of teleology that flows down into the contributing parts.

## Critique of the Properties Approach

### Specialized Parts as Bulk Matter

The core of a properties approach is found in the match between the specialized

part and the specific container system for which it is specialized.   It thus seems that

without such a match, the object would be regarded as undifferentiated, only becoming

recognized as specialized when its matching container system is known. If an unfamiliar part is encountered, we must therefore know what the corresponding container system is for us to recognize the part as specialized. But if they do not physically coincide (which the theory deliberately refrains from guaranteeing in principle) and the part is unfamiliar, how would we know if such a system exists, or what it is? Unlike the problems encountered with etiological theory, this is not merely an epistemic problem.

For example, consider a naturally occurring crystal that happens to have the perfect geometric properties required to serve some role in a laser. Suppose that it is superior to any that have been cut by hand or made in a lab. According to the simplest interpretation of a properties approach, the crystal would seem to simply be *for* the laser – even though its origin is unrelated to it.[3] Yet suppose the crystal arose naturally before the laser was even conceived. How can it be that an entirely natural object is for an artifact that may or may not even exist yet? In other words, is the crystal a specialized part or not? Does it become one only after the laser is developed? What if the builders of the laser are entirely unaware of the naturally existing crystal? As a naturally formed object predating the laser, the crystal seems to be merely an undifferentiated, "bulk matter" object – that is, one whose properties do not attest to a purpose. But in the context of the laser, it becomes a highly specialized part. Thus, it seems to be both at once.

---

[3] For comparison, according to Assigned Contributions theory, this crystal, once the match is recognized, can have this function assigned to it by selection.

Indeed, it would seem that *all* specialized parts are both specialized and bulk at once, as they are all (presumably) physical objects as well as functional parts. We see this even in our initial example: the heart in the butcher shop may still be a *recognizable* as a heart, but it is not *there* as one. It is there as meat – in other words, non-specialized biological matter. By contrast, were the heart in a cooler on the way to a transplant operating room, even though it is similarly disembodied it is still being treated as the specialized part that it (also) is. So it seems that the specialized identity is always in addition to, rather than instead of, its identity as a generic material object.

The question, then, that ultimately emerges from these examples is whether the identity of an object as a specialized part is merely a human interpretation of it when we recognize the kind of container system that the part would be useful in. Do we merely "read" the part as specialized because we recognize its usefulness, or is the specialization inherent in what it really is? If it is a matter of interpretation, what is it about container systems that prompt this reading?

**Container Systems as Teleological Cues**

Perhaps a container system is like a Rube Goldberg machine, in which parts that are preexisting and have other identities (functional and not) are brought together to form a new functional system, in the process giving all its parts new functional identities by becoming part of it. The Rube Goldberg machine arguably acts as a novel container system, making the parts within it specialized by giving them a role to play within it. For example, what was a mere rock now becomes a counterweight. The rock already existed

as a rock (i.e., bulk matter); but it did not yet exist, and could not yet exist, as a counterweight without something like the Rube Goldberg machine making it so. This would explain the conundrum of the laser crystal quite neatly – we simply interpret it as functional once we can imagine a function for it. It has, in effect, gained a new identity.

But if this is so, what seems decisive in the process is our interpretation of the overall container system as teleological. Its parts then attain their specialized identity by finding a role that contributes to the end of the system. This is not unlike a contributions approach, aside from the emphasis on part properties rather than the effects they produce. This would make the two apparently diverse approaches two sides of the same coin: we can look at the contributory effects, or we can look to the properties which make the effects possible (or which give the part the effect-dispositions it has). It would be essentially a choice of perspective. Of course, if this is so then no progress has been made, for we still require an account of what makes the overall systems teleological in the first place, especially for natural systems.

But this conclusion conflicts with the fact that in a properties approach the teleology – the for-ness of the part – resides, ostensibly, in the formal fit between the specialized part and the container system: not in the end purpose of the overall system. By hypothesis, it is supposed to be the fit between the specialized part and the container system which makes the part *for* what it does. Yet even cases where the teleology seems to reside in this kind of fit are not unaffected by interpretations of the purpose of the system. Return to the lock and key example above. It is hard to argue that the key, even

if randomly generated, is not in some sense "for" the lock that it opens.  Yet we know

that the container system context – the lock – is itself already somehow purposeful.  We

take for granted *what* a lock is, and that it is itself *for* something.  When the key opens it,

it is therefore natural to assume that the key participates in this teleology, making it for

the lock – even if we know it to have been randomly generated.

So let us consider a comparable biological example without such an assurance.

Consider a protein molecule which fits perfectly into another, causing it to change shape

and "open".  In this basic case, it makes no sense to say that one molecule is *for* the other,

even if the fit is precise.  They are just two large biomolecules that happen to fit together

in a precise way.  This example, it is true, does not posit a recognizable container system.

But this is part of the point: the container system in a properties approach seems to

implicitly include the presumption of a teleological context, which is how we can speak

of the specialized part "aiding" the system.  If we know that the two molecules and their

conjunction are vital to cell biology – say they are a hormone and a receptor on the

surface of a cell– we become instantly more willing to say that the two are "for" each

other in the same way that a lock and key are.  Perhaps this is because we already see the

overall cell as purposeful, and the protein molecules participating in that purposefulness.

The question of the teleological interpretation of the container system thus seems

crucial.  Is the teleology of specialized parts really found in the way that parts fit into and

are formally constrained by the container system (making it a fundamentally different

case than two molecules is in isolation fitting together)?  Or do we simply *interpret* the

overall container system (in this and other cases) as teleological, and thus see the special parts which fit them and aid the system as *for* that role. A properties approach claims to integrate teleology directly into its attribution of function, which it based on the fit between a specialized part and a corresponding container system. But does it ultimately rely on suggestive examples that smuggle in a particular teleological interpretation of that systemic context?

**Properties and Innate Functional Identity**

The answer is provided by a missing premise: A properties approach must hold that, in addition to being bulk matter objects, specialized parts have innate, objective identity as specialized parts – they are not merely bulk objects which appear functional from a given human or systemic perspective. Even if we could not recognize such a part, it would still have this identity. Without this presumption, the theory seems to become reliant on our interpretation of the nature of the container system.

Comparing the nature of specialized and unspecialized parts outside of any container system helps to clarify this premise. Even out of context, disconnected from their use, the heart and the screwdriver still carry the implications of their use and context with them. The screwdriver is a screwdriver because of its idealness for that use; likewise the heart. But consider the magnesium atom in chlorophyll. It too is ideally – indeed, uniquely – suited to its role in the structure of chlorophyll. And yet when out of context, it carries nothing of that context with it.

Why is this?  After all, even when outside of any particular context, all three could be characterized as matter with certain distinctive properties.  As we saw, even specialized parts are bulk matter objects as well.  What then makes the screwdriver a "screwdriver" even when it is not used as such, while the magnesium atom is always just magnesium even when it is being useful?  According to the properties approach, it is because the properties of the screwdriver are explained by that functional context, while those of the magnesium atom are not.  But this explanation, as was stressed above, is not a causal or etiological one in principle (though it may be in practice).  What it means is that these properties are required for the object to become part of the relevant system, *and to find identity as what it most truly is there*.

The chlorophyll molecule requires the precise properties of magnesium at least as much as the screw and the user's need to drive it require the precise properties of a screwdriver.  If anything, the properties required of magnesium by chlorophyll are of a far deeper and more fundamental nature than the properties of any tool are required by its user. Only magnesium can make chlorophyll.

But the screwdriver as an entity is defined not merely by its properties alone but by the *required-ness* of these properties *for* a container system.  The magnesium atom, by contrast, may be defined in large part by the properties that its presence in chlorophyll requires of it (e.g., its atomic number) – *but it is not at all defined by the fact that chlorophyll requires them*.

If etiological appeals are excluded on principle, this is an odd distinction. Neither the screwdriver nor the atom can owe its origin to the requirements of their respective container systems (otherwise, the properties approach would devolve or at least overlap with the failed etiological approach of chapter 2). What the screwdriver owes to the container system is the *need* of the container systems for these properties, for these are what *make* it a screwdriver (with or without actually being used as one). In contrast, the *need* of chlorophyll for magnesium does *not* make magnesium what it is (though, again, the properties of magnesium which chlorophyll needs may in fact be just those that make magnesium what it is).

This means that specialized parts have *an inherent identity as such* – one which we may discover merely through inspection of their properties (if we know of the kind of thing that they are). Specialized parts (at least those with a full measure of ideal properties) are not merely *used for* a purpose, or *usable* for a purpose – and they are not merely *interpreted* as functional due their suitability for a use. Even independent of actual use, according to a properties approach, specialized parts simply *are* functional. Thus a necessary premise in the properties approach is that specialized objects inherently are what they *truly are* because of their functional, specialized nature.

For a theory of function, this is a compelling and attractive attribute, integrating as it does not only function and teleology but the identity of the functional object itself. However, the metaphysical implications of such an assertion should not be overlooked. Can we truly say that any object has an inherent identity as particular type of thing – an

identity that goes beyond its immediate physical properties and appeals to what those properties might "mean" within another physical object? And an identity not only as a type of thing, but as a type of thing that is inherently *for* something else? At the very least, it is now clear that the power of a properties approach is rooted in some rather unusual metaphysical demands. Absent these, the properties approach devolves into a variation of a contributions account. It is not clear, at this juncture, whether these demands are enough to present a threat to the integrity of the theory. Their implications for a picture of teleology will be considered in the concluding chapter.

CHAPTER SIX

WHAT IS TELEOLOGY?

As we have seen over the past four chapters, the philosophical problem of

function is in large part a metaphysical problem of teleology. There are two aspects to

this issue. First, what is teleology? And secondly, can there be such a thing as teleology

in nature – that is, teleology without conscious intention?

To answer the first, we should start with the seemingly most secure, reliable,

intuitive case of teleology possible: deliberate human intention. When considered

schematically, what does this look like? When an action is done deliberately, we

sometimes note that it was "premeditated". By this we mean that the action was taken

with a clear mental picture of the consequence that would result from it. The

consequence was meditated on previous to taking the action which, it was believed,

would lead to it. This seems to be the very essence of teleology: an end which is the

cause of its own means. Teleology is thus a kind of loop: an end that is the cause of its

own means which in turn – in successful cases –lead to that end.

In cases of human intention, notably including the creation of artifacts through

engineering design, the ends are the cause of their own means because those ends are

consciously envisioned in the imagination of the agent. After envisioning this end, the

agent formulates a plan by which that end can be attained – whether that be designing and

fashioning a screwdriver from scratch or simply walking across the room to pick one up.

282

In a sense, the agent works backwards, reasoning from desired ends to appropriate means. Our own experience of this sense of intention, of having a purpose and considered ends behind our actions, is so familiar and compelling that it is difficult to doubt its reality.[1]

If this pattern of mutual interaction between means and ends is what teleology is essentially about, are there natural or non-intentional parallels?  What makes the question so challenging, clearly, is finding a substitute for the role of intentionality.  What else besides conscious imagination on some level can account for the "reverse" part of the teleological loop – the sense in which ends cause their own means?  In a natural efficient-cause setting, causality seems to work in one and only one direction: from means to ends – or, better, from causes to effects, since without both parts of the loop there can be no such things as proper means and ends.  Thus, in a natural, non-conscious setting, ends seem impossible.  But if the essence of teleology truly is its schematic pattern rather than its conscious instantiation, perhaps that pattern can be exemplified, and thus the criteria for teleology satisfied, in other ways.  In a sense, each of the four approaches to function studied above can be seen as a unique attempt to solve this central problem.

As has been stressed at various points throughout the work, particularly in chapter 2, a philosophical theory of function must attribute function in rough accordance with ordinary usage, helping to explain and articulate the intuitions which lie behind such common use.  It cannot simply stipulate what function is in some abstract theoretical sense independent of ordinary use.  This has been our yardstick in evaluating each of the

---

[1] Of course both philosophy of mind and contemporary neuroscience do call this simple experience into question.  Indeed, the question at hand of natural teleology is relevant here: Without an account of natural teleology, the source of human intentional teleology may become harder to explain.

four approaches to date. But we have also seen that the basic intuition of function seems to inherently involve teleology in some fashion. We saw in chapter 3 that theories of function which deliberately omit all teleological dimensions attribute a different kind of function, and to different kinds of objects and systems, than ordinary intuitive function attribution.

Thus, a sound philosophical theory of function seems bound to also involve a metaphysically coherent and plausible account of non-conscious teleology, implicitly if not explicitly. If the teleology were only of the familiar conscious variety, the associated theory of function would be unable to attribute function to non-intentional settings, including organisms and even artifacts if their precise design intention is not known or not directly relevant. Thus the plausibility of these implied views of teleology – their metaphysical plausibility and the degree to which they reflect the true basic principle of teleology – can also be used as criteria for evaluating theories of function. On this new dimension, which of the four very different approaches is to be preferred? Let us briefly consider each in turn.

**Teleology in Etiological Function Theory**

The idea of teleology as a feedback loop pattern between cause and effect was precisely the logic behind the development of etiological function theory, as we saw in chapter 2. Rather than look directly to intention, etiological function theory tries to find the pattern of teleology in the closely associated idea of causal etiology. In etiological theory, teleology and function coincide: a part has a function if and only if the origin of

the part is attributable to it having the function. While conscious design intent is one example of this pattern of origin (and for Wright the singular criterion for function in artifacts), the etiological pattern was generalized to include natural selection in organisms as well. An organic trait has a function if and only if its natural selection as a type was due to precisely the candidate effect.

Yet the etiological approach to function was seen to fail because the connection between origin and function could not be made tight enough to support the approach as a universal theory of function. Many things seem to have function without having the requisite kind of origin (such as first-generation traits), and vice-versa. Moreover, even knowing the actual origin of a thing is often problematic even when its functional role seemed secure. So the etiological approach failed as a theory of function. Yet its vision of teleology also suffered under close inspection. It claimed a strong purpose and normativity based on this pattern of origin that seemed unsupportable under a naturalistic metaphysics. What did it mean that selection, for example, imbued the trait with an unobservable, unenforceable "role" or "office" that the part is obligated to fulfill?

The etiological approach to function theory can be seen as the most direct and literal attempt to translate the pattern of intentional teleology into non-intentional scenarios. Unfortunately, it was also seen to be unable to live up to its ambitions.

**Teleology in Systemic Effects Theory**

For similar metaphysical reasons, systemic effects theory, considered in chapter 3, dismissed at the outset the very possibility of teleology outside of conscious intention.

As a consequence, the resulting theory of function had two unusual characteristics. First, it applied only to entirely non-conscious settings. Even artifacts as objects fell outside its scope, to say nothing of their actual intentional use. Because of their inherent relationship to intention as artificially made objects, artifacts are out of bounds for systemic effects theory, at least as currently developed. While both Cummins and Davies held out the prospect that the theory could potentially be adapted to artifacts and other intentional contexts, both declined to pursue the possibility. Indeed, there are good reasons for thinking that such a project is not promising, as was detailed in chapter 3.

Second, the function attribution that resulted often did not look much like function as it is ordinarily attributed. The explanatory value of systemic effects function related to what a thing does, but not what its purpose is or more generally to the meaning of its presence. Moreover, by having such a generic and wide-ranging concept of function, systemic effects theory was forced to attribute function to systems and objects which, under any other conception and in ordinary intuition, have no function whatsoever – such as the ionic bonds in a natural salt crystal. In sum, by omitting any consideration for teleology, systemic effects function was seen to fail as a theory of function.

**Teleology in Assigned Contributions Theory**

A contributions approach to function appeals directly to "goals". Without goals, there can be no contributions made towards them. But this simple appeal to goals masks the complexity of the relationship between a contributions approach and teleology. Notably, contributions alone do not imply the interaction of means and ends that seems

characteristic of teleology. In a contributions approach, systemic goals are stipulated, and any internal effects that aid the goal are considered functions. But there is no sense in which those effects are necessarily "for" the goals – they simply help the attainment of the goal. Thus, in a basic contributions approach, a thing can aid a goal without being "for" aiding the goal. This disconnect is precisely what necessitated the introduction of the distinction between function performance and function possession.

So while a contributions approach uses teleological language, and makes teleological allusions, in some sense it is not quite a true teleological theory. Rather, it piggybacks on some external teleological attribution of the system in which function occurs – a system which, by some unstated attribution mechanisms has "goals". Instead of a looping pattern, a contributions approach thus treats teleology in a "linear" fashion, with parts contributing to ends while not themselves being necessarily influenced by that end. Yet this is not necessarily a weakness of the approach – on the contrary, invoking but not incorporating teleology into function is what allows the contributions approach to avoid the pitfalls – especially the metaphysical ones – of the etiological approach, while staying essentially true to ordinary intuitions of function attribution.

When stronger teleology is called for – when it is necessary to explain why some things seem to be for the functions they perform while others do not – the contributions approach seems to resort to more conventional teleological concepts, such as etiology. This is where Assigned Contribution is introduced, with its three-fold modes of assignment: design intent, selection, and typology. Yet here too Assigned Contribution

sidesteps the metaphysical problems of appealing directly to etiology for teleology.

While the assignment of function can derive from knowledge about the origin of the part

(in two of the three modes), it is actually based on the human psychological *expectation*

of function that results from this knowledge – not the origin in and of itself.  This

principle is driven home by the third mode, typology, which does not appeal to any

etiological or teleological pattern at all, but simply highlights a very common source of

information that drives human expectations – the behavior of like systems with like parts.

Contributions approaches generally, and Assigned Contribution theory in

particular, thus represent a careful balance in the role of teleology in function.  While

they acknowledge the relevance of teleology for function, they do not incorporate

teleological principles directly into the theory.  When such principles seem relevant,

Assigned Contribution theory appeals to their role in human psychology rather than

asserting an objective, ontological existence for the teleological role.  In this way it is

able to both account for ordinary intuitions about function and make those attributions to

both natural and artifactual contexts – while still avoiding any un-naturalistic

metaphysical difficulties.  The theory effectively replaces the "backwards" ends-to-means

causality with an account of human expectations for function performance.

Yet as we saw in chapter 4, despite its cautious, arms-length approach to

teleology, contributions theory still incurs two significant complications from its

engagement with it.  The most notable of these is how to characterize the systems in

which function is to be attributed as goal-seeking in the first place.  The approach relies

on the independent identification of the system as teleological – as having an end that can be contributed towards. Indeed, the picture of teleology in contributions requires that some things ultimately be ends in themselves – for without this the chain of contributions cannot be terminated. In particular, there is the problem of attributing teleology to organisms as having ends. Boorse proposed to solve this with a cybernetic approach, but as we saw in chapter 4, this was not without its problems.

Yet this need not be a fatal flaw, for, in keeping with the centrality of human interpretations for the approach, all that is really required is the human taking of a system as teleological – as having goals. Even if this taking is not justifiable as objectively real independent of human interpretations, function can still be attributed to the systemic goal as stipulated. A contributions approach insists that function be relative to a goal, but it does not insist that the goal be somehow metaphysically real. Thus, whether or not non-conscious natural organisms truly have goals, or we merely treat them as if they do because they act as if they do (or if, as in a cybernetic analysis, the two premises are treated as equivalent) does not ultimately matter for the purposes of attributing function within them.

The other complication, however, is not so easily avoided. Function in the contributions approach goes all the way down, without apparent limit – it encompasses all parts at all levels, no matter how general, potentially including, for example, atoms, environmental conditions and even natural laws. As long as the candidate is part of a nominally teleological system, that part performs a function if it contributes to the goal.

It may even have an assigned function if that contribution is to be expected. Rather than highlight only unique entities that are "for" what they do, this dragnet implicates a wide range of entities with no obvious limit. This is a problem for the contributions approaches as theory of function, but its roots are in the oversimplified schematic of teleology under which it operates.

## Teleology in the Properties Approach

The properties approach, by contrast, takes a more sophisticated approach to teleology – one specifically designed to highlight only those specialized items that are for what they do. It looks for purpose not in the end goal of the system, but in the *interaction* between the part and the system in which the part has function. The systemic whole is an end for the part, and that whole also constrains the kinds of parts it will accommodate. The mutual pattern of teleology is thus built into the basic principle. The part has an end *in* the whole; the whole is an end *for* the part (not simply an end unto itself).

In essence, the properties approach argues that the feedback loop of teleology need not be instantiated as a causal loop: the loop can also be instantiated as an interaction of form. To be for an effect, the part finds a role within a logically higher, more general structure termed a container system. The part attains a new specialized identity within the system. If the match is precise enough, the part is for the system and the role it plays there. Whereas the teleology in the contributions account points outside the theory – indeed, outside the system, the teleology in the properties account is contained in the theory, and contained in the local system.

In this, notably, it resembles Kant's characterization of natural teleology, the first condition for which is that "the parts... [be] possible only through their relation to the whole" (Kant 2008, 5.373). Even more apt is the second condition for a natural end: that "the parts combine themselves into the unity of a whole by being reciprocally the cause and effect of one another's form" (ibid.).

The teleology of the properties account may present a more reliable and theoretically coherent picture of natural teleology than those of the competing accounts. However, as we saw, it does this at the price of a substantial metaphysical commitment to a thing having a singular innately teleological identity. Parts that have function are innately defined by how they are *for* their container systems. Parts that do not have function just are what they are, regardless of context or systemic contribution. If such innate teleological nature can be affirmed, then the approach may offer the most attractive picture of teleological function.

**Is Non-Intentional Teleology Real?**

The selection of a universal theory of function comes down to a comparison between the contributions and the properties accounts. The contributions account does less to resemble "true" teleology. But in its restraint, it articulates an account of function with minimal metaphysical complications. Even when it appeals to teleological-style origin, it does so to explain human expectations for function, rather than to attribute innate teleology to the part. Being "for" a function is seen as merely a human projection of obligation onto the part.

By contrast, the properties approach stays truer to the picture of pure teleology that posits ends as the cause of their own means. But in doing so, it is forced to posit an innate teleology in functional parts, one which does not depend on human interpretations or knowledge. If this is metaphysically plausible, it would therefore seem to be the superior account of function. But is it? The choice thus turns on whether there can be any such thing as genuine natural (that is, non-conscious) teleology – an objective teleology as opposed to a subjective human interpretation. This is a difficult question to even give precise meaning to, much less answer. What could it mean for a thing to be truly *for* a purpose, absent any intention that it be so?

Ultimately, whether a contributions or a properties approach is considered superior as a universal philosophical theory of function seems to depend largely on whether the basic concept of inherent, natural teleology is granted as metaphysically coherent and plausible. If it is suspect, a contributions approach that retains or explains most of the teleological content while avoiding all of its metaphysical complications seems preferable. If it is granted, a properties approach that is able to locate basic teleological patterns in the form of parts and systems, without relying either on unreliable etiology or actual conscious intention may be superior.

The question is whether it is coherent to posit this purpose as objective within a naturalistic metaphysics. The experience of similar claims in the etiological approach suggests reasons for caution. In chapter 2, we encountered the severe metaphysical problems that seem to result from such attributions, including the questionable

ontological status and explanatory value of the functional "role" which the part is

normatively obliged to fulfill.  Such objections remain unanswered.

A skeptical take on natural teleology might say that the very concept of teleology

derives from our experience of conscious intention, which is then inappropriately

generalized to non-intentional contexts.  In that case, any attempt to translate conscious

intention to non-conscious systems could only be heuristic or metaphorical.  If the idea of

non-conscious teleology is considered merely a projection into the world encouraged by

our own experience of conscious intention, the contributions approach to function would

seem to be the best available account, as it provides a systematic account of function that

comports with how it is ordinarily attributed and without incurring any additional

metaphysical commitments.  But if objective natural teleology is coherent and

metaphysically plausible, the properties approach to function would seem to be the best

available account of function, as it unifies function and teleology.

The status of non-intentional teleology is thus *the* central question for a

philosophy of function. Yet even in the absence of an answer to this fundamental

metaphysical question, with further development it may still possible to utilize the

properties approach by allowing the teleological nature of functional objects to be taken

as less than objectively innate – perhaps as a sort of "secondary property" which depends

not only on the object but on the observer as well.  As we saw in chapter 5, this would

bring the properties approach closer into line with the contributions approach.  Yet it

would still retain the conceptual simplicity of appealing to immediate properties, and

helping to identify presumed teleological interactions. The properties approach sketched out here is thus worthy of further study and development into a more comprehensive theory of function, even for those who view objective natural teleology skeptically. However, for an already fully worked-out approach to function which avoids all metaphysical complications while staying true to ordinary function attributions, the Assigned Contributions theory is recommended.

REFERENCES

Beckner, M. 1969. "Function and teleology." *Journal of the History of Biology* no. 2 (1):151-164.

Blachowicz, J. 2012. *Essential Difference: Toward a Metaphysics of Emergence*: State University of New York Press.

Boorse, Christopher. 1976. "Wright on functions." *The Philosophical Review* no. 85 (1):70-86.

Boorse, Christopher. 1977. "Health as a theoretical concept." *Philosophy of Science* no. 44 (4):542-573.

Boorse, Christopher. 2002. "A Rebuttal on Functions." In *Functions: New Essay in the Philosophy of Psychology and Biology*, edited by A. Ariew, Cummins, R. and Perlman, M. , 63.

Boorse, Christopher. 2011. "Concepts of Health and Disease." In *Philosophy of Medicine*, edited by Fred Gifford. Elsevier.

Borchert, D.M. 1996. *The encyclopedia of philosophy: Supplement*: Simon & Schuster Macmillan.

Buller, David J. 1999. *Function, Selection, and Design*: State University of New York Press.

Canfield, John. 1964. "Teleological Explanation in Biology." *The British journal for the philosophy of science* no. 14 (56):285-295.

Cummins, Robert. 1975. "Functional Analysis." *Journal of Philosophy* no. 72.

Cummins, Robert. 2002. "Neo-Teleology." In *Functions: New Essay in the Philosophy of Psychology and Biology*, edited by A. Ariew, Cummins, R. and Perlman, M. , 157.

Davidson, Donald. 1987. "Knowing one's own mind." *Proceedings and Addresses of the American Philosophical Association* no. 60 (3):441-458.

Davies, Paul Sheldon. 2001. *Norms of Nature: Naturalism and the Nature of Functions*: MIT Press. Original edition, 2001. Reprint, 2003.

Hempel, Carl G. 1994a. "The logic of functional analysis." In *Readings in the philosophy of social science*, 349-76. Original edition, 1959.

Hempel, Carl G. 1994b. "The logic of functional analysis." *Readings in the philosophy of social science*:349-76.

Kant, I. 2008. *Critique of Judgement*: Free Press.

Lewens, T. 2005. *Organisms and Artifacts: Design in Nature and Elsewhere*: MIT Press.

Melander, Peter. 1997. *Analyzing Functions: An Essay on a Fundamental Notion in Biology*: Almqvist & Wiksell International.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*: MIT Press.

Millikan, Ruth Garrett. 1989a. "An ambiguity in the notion "function"." *Biology and Philosophy* no. 4 (2):172-176.

Millikan, Ruth Garrett. 1989b. "In Defense of Proper Functions." *Philosophy of Science* no. 56.

Millikan, Ruth Garrett. 2002. "Biofunctions: Two Paradigms." In *Functions: New Essay in the Philosophy of Psychology and Biology*, edited by A. Ariew, Cummins, R. and Perlman, M. , 113. Oxford University Press.

Nagel, Ernest. 1961. "The structure of science." *American Journal of Physics* no. 29:716-716.

Nissen, Lowell A. 1997. *Teleological Language in the Life Sciences*: Rowman & Littlefield Publishers.

Schaffner, K.F. 1993. *Discovery and Explanation in Biology and Medicine*: University of Chicago Press.

Sober, Elliott. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*: University of Chicago Press.

Wright, Larry. 1973. "Functions." *Philosophical Review* no. 82.

Wright, Larry. 1976. *Teleological Explanations: An Etiological Analysis of Goals and Functions*: University of California Press.

VITA

Mark L. Bourgeois is an engineering instructor in the Segal Design Institute at Northwestern University, an ethics instructor in the Biomedical Engineering department at Northwestern, and the administrator of the Northwestern Center for Engineering Education Research (NCEER).  He was previously an engineer in the telecommunications industry for firms such as Lucent Technologies.  He holds undergraduate degrees in physics and philosophy from the University of Illinois at Urbana-Champaign and an M.A. in philosophy from Miami University, Ohio.  Beginning fall 2014 Dr. Bourgeois will be a postdoctoral fellow at the Reilly Center for Science, Technology and Values at Notre Dame University.