Normativity without Artifice: A New Foundation for Teleological Realism

Mark Bauer

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2007

Approved by:

William Lycan

Marc Lange

John Roberts

Jesse Prinz

Ram Neta

ABSTRACT

MARK BAUER: Normativity without Artifice: A New Foundation for Teleological Realism
(Under the direction of William Lycan)

Interest in teleological analysis has risen dramatically over the last several decades: teleo-functional accounts of biological systems, of gross mental state types, and of intentionality have all had their share of adherents. Such analysis has been attractive in part, because characterizing function as some item's job, office, or role allows that it might perform its work more or less well. (For example, in the context of intentionality, the possibility of malfunction is what is thought to secure intentional inexistence, e.g., misperception, false belief, etc.) For the application of teleological analysis to traditional philosophical problems, such as the "mind/body" problem or the problem of "intentional inexistence", to be successful, however, teleo-functional ascriptions to natural or nonartifactual systems must be construed literally. Yet, since a teleologically characterized item or behavior can succeed or fail at its function, teleo-functional ascriptions imply norms or standards of functional performance. A literal construal of nonartifactual teleological ascriptions presupposes, then, that there are literally norms within the natural world, which are independent of intentional and psychological agency. Any realist account of nonartifactual teleology must have at its core a realist account of nonartifactual normativity. In short, I develop just such an account of normativity and one that will serve as the foundation of nonartifactual teleological realism, thereby securing, I believe, a theoretical pillar requisite for naturalizing the mind.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

Chapter 1

TELEOLOGICAL REALISM AND THE PROBLEM OF NORMATIVITY

My aim in what follows is to show that teleo-functional ascription is often true beyond the realm of artifice. My motivation is not, however, innocent: If teleo-functional ascription can only be true of artifice, then I see no prospect for naturalizing the mind. With the equally distasteful options of dualism, fictionalism, or eliminativism if that naturalizing project fails, it is of some import whether teleo-functional ascription is true of the nonartifactual. My purpose is, in short, to shore up a theoretical pillar requisite for naturalizing the mind. That said, much of what follows will have little directly to do with the psychological and the intentional. Instead, the aim is to show that teleo-functional ascriptions are true of a wide range of natural phenomena, particularly within biology. With that in place, I will only make the most cursory moves connecting the analysis of biological systems generally to the problems associated with naturalizing the mind.

No one with much familiarity of the biological sciences would doubt, I suspect, the utility of teleo-functional ascription and analysis. The utility of such analysis rests, it seems, in understanding the complex operations of biological systems as analogous to the products of intentional design. Of course, that some form of analysis is useful in advancing human understanding does not entail that it is true. (It is a useful pedagogic heuristic to teach children the solar-system picture of the atom despite the falsity of that model.) So, are the nonartifactual systems of biology usefully analyzed as analogous to artifacts in virtue of their possession of function or merely in virtue of some aspect of our psychology? If the latter, functional analysis of the nonartifactual is strictly speaking false. Nonartifactual functions are only mere analogues of the genuine functions of the artifactual – that is, nonartifactual systems, though in fact lacking function, act as something possessing function might act. Teleological analysis of such systems is a useful heuristic, but nothing more. To reject such antirealism, then, is to think that the utility of understanding at least some nonartifactual systems as possessing function lies in the fact that they do possess it; so, that some nonartifactual system possesses a function is a fact not merely about our psychology but about the organization of the world.

There have been at least two consistent strands of thought that functional analysis applied to the nonartifactual can be no more than a heuristic device. The first is that talk of the purposive – talk utilizing 'in order to', 'for', 'so as to', etc. – just presupposes intentionality for its literal truth. This strand remains very much alive in contemporary philosophical literature. Dretske writes,

> Purposeful action – in contrast to mere behavior – requires thought, but thought alone is not enough. To qualify as purposeful, thought must *control* behavior. To be an agent it is not enough to be a thinker and a doer. The thinking must explain the doing. (Dretske 1999, 19)

Dretske's example of a "mere behavior" inappropriately teleologically characterized is a plant's changing color in order to attract pollinators:

> The botanists from which I take this example succumb to this temptation by explaining the plant's behavior in purposive terms: the plant changes color, they say, "in order to" attract pollinators. Clearly, though, we do not have anything like action, nothing like purpose, nothing that would justify "in order to". The plants would change color whether or not their behavior succeeded in attracting pollinators. (Ibid., 26)[1]

[1] In contrast to his statements here, Dretske, when representational function is the issue, endorses naturalized teleology. (See Dretske 1986 & 1995)

Even more recently, Davies, criticizing Kitcher's (1993) attempt to unify artifactual and nonartifactual theories of function via the concept of design, writes:

> But I take that genuine unification in this case requires a single, non-disjunctive concept of "design" that applies equally to artifacts and to the products of evolution due to natural selection. And there is no such concept. It is of course plausible to describe the processes that lead to the production of artifacts as involving design…. But it is not plausible to say the same about processes that lead to evolution via selection of natural objects, no matter how many marks of apparent design they bear…. For no one, not even Darwin, can show that there is design in the absence of a designer, since no one can reject the rather plausible conceptual claim that design requires a designer. (Davies 2001, 58-9).

And, since all such natural design is merely apparent in the absence of a designer, Davies concludes that any apparent purposiveness within the nonartifactual is just that apparent: "… the obvious point is that, while intentions fix the functions of artifacts, nothing fixes the functions of natural systems." (Ibid. 61) As a result, Davies suggests,

> … that certain of our psychological capacities and limitations incline us to conceptualize the capacities of stable, self-perpetuating systems as especially functional…. They strike as being for the sake of certain tasks that contribute to or result from the stability of the system… Exactly why we are so struck is not for me to say – *that is up to the psychologists*. My speculation is simply that we are so struck. (Ibid. 153, emphasis added)

In an environment where Kitcher just asserts, "one of Darwin's important discoveries is that we can think of design without a designer," (Kitcher 1993, 380) the intellectualism exhibited by Davies might seem to be parochial or reactionary. But, we can strengthen the intellectualist motivation by noticing the essentially normative character of teleo-functional ascriptions. Teleology looks to be an essentially normative affair, because the function of, say, some item need not be what it in fact does, is disposed to do, or even could do but is, rather, what it ought to do. The student raising his hand is to draw the attention of the instructor even when the raised hand goes unnoticed. Or, the water dance of the male American alligator is for courtship even when no female responds. Further, Langley's launching apparatus was to provide the initial lift for his Aerodrome, but, not only did it fail to do so, it could not have done so given the laws of physics. Or, a tokened structural gene has the function of coding for a certain protein, even when it not only fails to do so but cannot do so given the absence of a particular enzyme. That a possessed teleo-function is not to be equated to what an item or behavior does, is disposed to do, or could do but is rather what it ought to do reflects the fact that teleologically characterized events or behaviors are successes or failures. Success and failure presuppose some standard in virtue of which, say, a behavior is held to account as a success or a failure. Functioning properly, successful performance of function, is to meet some standard of performance; it is to do what one ought. Ascription of teleological functions implies standards or norms of performance.

As a result, a nonartifactual teleological realism requires, it would seem, a commitment to nonartifactual normativity. (In contrast, the antirealist only need suppose that the "ought-to-do"s and "supposed-to-do"s of teleological analysis reflect the epistemic and explanatory norms present in any scientific endeavor.) It is this commitment on the part of the teleological realist to nonartifactual norms that enhances the intellectualist's antirealism: norms, standards, and rules seem restricted to the province of the intentional and the psychological, and so literal teleological ascription is equally so restricted. Worse still for the realist, Davies (2001) suggests that any theory committed to nonartifactual normativity is wed to a metaphysics inconsistent with a contemporary naturalist worldview. (It will be central to my claim that nonartifactual teleo-functional ascription can be literally true that normativity is a widespread natural phenomenon beyond the province of the intentional and the psychological. See chapters 5 and 6.)

The second strand, not wholly disconnected from the first, is that treating nonartifactual teleo-functional ascriptions as literally true weds one to something like Aristotelian final causes, entelechies, or some form of backwards causation. At least sometimes when I act on purpose or for a purpose, I do so with an end in mind. That I have an end in mind is part of the explanation of my so acting; that end, as Dretske says, is in some sense controlling what I do. To think that the flower changes color in order to attract pollinators seems to assume that the end of attracting pollinators is similarly playing a causal role in the changing of color. And, if its causal role is not to be explained by some representation of that goal, then we would seem to be countenancing metaphysically strange forms of causation with no role in a contemporary naturalistic worldview.

The realist response to such concerns is to point to some grounding process, some naturalistic mechanism, that produces and/or accounts for some item's or behavior's possession of function. The relevant grounding process varies from etiology or selectionist history to the self-sustaining, self-generating, or otherwise homeostatic structure of biological systems. Where etiology or selectionist history serves as the grounding process, the teleo-function of some item or behavior is that effect that explains the presence and/or persistence of that item or behavior type within a population.[2] For the varied regulative or homeostatic views, the teleo-function of some item or behavior is

[2] See, for example, Wright (1973), Millikan (1984, 1989b, 1993, 1999a, & 2002), Neander (1991a & 1991b), Griffiths (1993), and Godfrey-Smith (1994).

that effect which it will persistently produce despite internal and/or external disturbances that should otherwise inhibit or eliminate the relevant effect.[3] In either case, the relevant grounding process is a naturalistic process that need not involve hypothesizing metaphysical strangeness; so, the second concern falls by the wayside. In addition, since we need invoke no intentional and psychological agency in the explanation of such grounding processes, we can forgo, the realist suggests, the presupposition of intentionality in teleological ascription.

While the provision of a naturalistic grounding process goes some way to dissolve the second sort of antirealist concern, it is far less efficacious in respect to the first. The antirealist can, as Davies explicitly does above, accept that the various offered naturalistic processes prompt us to see them as purposive. The antirealist, in denying the literal truth of nonartifactual teleo-functional ascription, does not deny that such functional analysis is useful nor that it might not be particularly useful in cases of selectionist history or homeostasis. That it is particularly useful in such cases, the antirealist suggests, is exactly because these processes or their products are in some sense analogous to how an intentional agent might produce or design its products. But, without such an agent, such grounding processes provide for something only merely analogous to genuine teleology.

The antirealist, I want to stress, is not being unreasonable in failing to see that this or that grounding process suffices for literal teleological ascription. Any process for which the initial properties determine the terminus of that process can be described with teleological language – that is, we can describe the initial properties as *for* the production

[3] The homeostatic/regulative view is initially suggested by Rosenblueth et. al. (1943). It is further developed by Sommerhoff (1950 & 1959), Braithwaite (1953), Beckner (1959), Hempel (1959), Nagel (1961, 1977a, & 1977b), and Boorse (1976 & 2002).

of the terminus. Yet, no one accepts that every process so described is literally teleological. The intellectualism of the antirealist provides a straightforward reason why not every process so described is literally teleological: a process is literally teleological when and only when a representation of a goal is part of the causal explanation of that process. The teleological realist's reasons for distinguishing some naturalistic processes as rightly teleological and others as not are less straightforward.

Borrowing a page from Wright (1973), functional analysis, when applied to the nonartifactual, is in the business of offering explanations. If such analysis is to explain the operation of a system or process, then we need to be able to distinguish between the functions of, say, some component and its accidental effects. For example, an explanation of the heart should distinguish its function as a pump for the circulatory system from its other accidental effects, e.g., the production of heart sounds, occupation of a region of the torso cavity, etc. The thought, then, is not every teleological description of some given process is rightly teleological, because some effects so characterized are only accidental in respect to the structure of the relevant system. The various offered grounding processes serve to distinguish which effects are rightly teleological and which are not by distinguishing which are not accidental and which are, respectively.

However, to distinguish between which effects are and are not relevant to the operation of some system, the antirealist will counsel, is not to distinguish between what is and is not rightly teleological. The former distinction concerning explanatory relevance is a general distinction that has nothing in particular to do with teleology. For example, the Moon's gravitational pull is explanatorily relevant to terrestrial tides while its magnetic field is not. That one is relevant and the other not is not a distinction in teleology: “the Moon’s gravitational pull serves the purpose of creating terrestrial tides” should be obviously metaphoric. The nonartifactual teleological realist needs to provide some reason to think that this or that grounding process not only distinguishes between which effects are explanatorily relevant and which are not but that distinction is relevant to literal teleological ascription.

The realist is not without resources here. One might point to, for example, selectionist history as grounding talk of “success” and “failure” in a way not relevant to other causal processes that might otherwise be teleologically characterized. The relevance of “success” and “failure” ties the grounding process to the normative character of the teleological and, thereby, provides some conceptual reason to think a grounding process such as selectionist history is relevant to literal teleological ascription.

However, that realist response only seems to have moved the conceptual question back a level. The antirealist can rightly demand to know why the utility of the language of “success” and “failure” with respect to selectionist histories suffices to think such terms are applied non-metaphorically. And, importantly, such terms in the context of selectionist history do look prima facie metaphoric. Some organisms act in ways that lead to the perpetuation of their lines and others do not. Are the organisms that survive “successes”? Well, they did not die out. But, it is unclear that there is anything in the universe like a standard governing or binding lineages such that organisms that do not die out have meet that standard. Prima facie, it looks more plausible that the utility of using the language of “success” and “failure” is a mere heuristic for us, human animals, to understand a natural process. James, for example, voices just this sentiment:

> We talk, it is true, when we are darwinizing, as if the mere body that owns the brain had interests; we speak about the utilities of its organs and how they help or

> hinder the body's survival; and we treat survival if it were an absolute end, existing as such in the physical world, a sort of actual should-be, presiding over the animal and judging his reactions, quite apart from the presence of any commenting intelligence outside. We forget that in the absence of some superadded commenting intelligence (whether it be that of the animal itself, or only ours or Mr. Darwin's), the reactions cannot be properly talked of as 'useful' or 'hurtful' at all. Considered merely physically, all that can be said of them is that they if they were to occur in a certain way survival will as a matter of fact prove to be their incident consequence. The organs themselves, and all the rest of the physical world, will, however, all the time be indifferent to this consequence, and would quite as cheerfully, the circumstances changed, compass the animal's destruction. In a word, survival can enter into a purely physiological discussion only as an hypothesis made by an onlooker about the future. (James 1890/1950, 140-141)

It is, at least, not obvious why the fact that some lineages die out and others do not should suffice to instantiate a norm or a standard governing biological lineages. Further, it is not at all obvious what that standard, if there were one, would do. Norms and standards, at least in the intentional and psychological context, are not epiphenomenal. They play causal and regulative roles in the governance of intentional and psychological behavior. But, in this selectionist case, what causal role does any such hypothesized norm play? It would seem to play none for the simple reason that it is causally irrelevant to whether lineages do or do not die out. If a grounding process such as selectionist history can at best provide metaphoric normativity, then it remains unclear why it suffices to provide anything more than metaphoric teleology.

Again, we see the theme around much of what follows will focus: if the offering of some naturalistic process is to ground literal teleological ascription, then it need make plausible that there are nonartifactual norms of performance.

The realist can, following a strategy consistently advocated by Millikan,[4] just refuse to engage the intellectualist tendencies of the antirealist: the project of providing the grounds for teleo-functional ascription, the realist suggests, is not to engage in conceptual analysis but to provide a theoretical definition of function that accords with theoretical practice and is naturalistically tractable. The worry that such teleo-functional ascription cannot be literally true rests on some conceptual claim about purposiveness and teleology, but the project, so the response goes, is not to divine the necessary and sufficient conditions of some concept but to propose one that will do the theoretical work required. Take the realist theory of your choice – selectionist history, homeostasis, etc. – the concerns from the antirealist are just irrelevant to its utility in science and so irrelevant to whether we ought to adopt that theory's stipulation for function. What we ought to ask in respect to the various offerings is which will serve as the theoretical backbone, say, for biological explanation.

It is, of course, the theorist's prerogative to introduce theoretical terms when needed. It is also quite understandable that one would introduce a homonym when the phenomenon was in some way analogous to that to which the original expression applied. But, we should notice that such a strategy employed here is just to concede the antirealist's point. Recognizing at least two senses of the teleological (one associated with the intellectualist intuitions and one with some preferred grounding process) is to recognize that, say, it is literally false under the intellectualist sense that a bat's wings are for flying. This is just the antirealist's point. Again, the antirealist concedes the utility of

[4] See Millikan (1984, 1989b, & 2002).

the teleological in respect to natural phenomena and can concede some grounding process or other will be what explains such utility.

Millikan (2004) rejects, however, that she is conceding two senses or kinds of purpose. A division between the purposes of intellect and the purposes of the natural order, Millikan suggests, is an artificial and arbitrary division, because no theoretically useful line can be drawn between the supposed two types of purpose:

> … the purposes of genes, of unlearned behaviors (smiling), of conscious intentional actions, of at least some cultural products (greeting rituals), and of artifacts are all purposes in exactly the same sense of "purpose". In all cases the thing's purpose is, in one way or another, what it was selected for doing. (Ibid, 13)

That a thing's purpose is what it was selected to do is not the result of conceptual analysis, Millikan suggests, but reflects a "common underlying pattern beneath the surface features that we recognize as the marks of purposiveness across a variety of domains" (Ibid.) And, since the same marks are present in the intellectual and nonintellectual cases, the intellectualism of the antirealist is without theoretical value. Now, it is essential for the antirealist to be such that he agree that there is a univocal sense of purpose; after all, it is because there is a univocal sense of purpose that the antirealist seeks to distinguish between literal and metaphoric applications of that selfsame concept. Further, the antirealist can accept Millikan's conceptual anthropological thesis about what marks prompt us to speak of purpose. But, the antirealist need not be compelled to accept Millikan's conclusion that no useful theoretical line can be drawn between intellectual and nonintellectual purposes. If the only norms or standards are those of artifice, then there is a substantive metaphysical distinction: the purposes of the intellect can literally be achieved and the purposes of nonartifactual cannot. Pointing to "selection" does little to help matters, because there too rests the same question with respect to metaphoric application and there too the prima facie answer is that the "selection" of "selectionist history" is metaphoric.

In short, if we are to cast aside the antirealist's intellectualism and understand how literal nonartifactual teleological ascription is possible, then we need to understand how nonartifactual normativity is possible. Selectional or regulative views that fail to account for such normativity will always fail to offer a compelling case that the relevant grounding process is right or relevant to teleology. What I hope to show is that normativity is not a phenomenon restricted to the province of the psychological and the intentional but is, rather, a widely spread natural phenomenon. Once that is established, we will be in position to see how it is that teleological ascription can be literally true of the nonartifactual.

Chapter 2

MINIMAL TELEOLOGY

A number of philosophers, from Kant (1790/2000) through contemporaries such as Cummins or Davies,[5] have consistently held that, though teleological analysis is indispensable to biological explanation, it is literally false with respect to the nonartifactual. To establish the literal truth of nonartifactual teleological ascription, I suggested in the previous chapter that the teleological realist must show how nonartifactual normativity is possible, and I proposed to demonstrate how that is the case. I want, however, to set aside temporarily the question of whether nonartifactual teleo-functional ascriptions are literally true. It would be helpful, before attending to that question, to make it clear just what is up for dispute – that is, what is it exactly that some want to accept and others want to deny as genuinely teleological? Toward that end, I will make some brief and general comments in this chapter about what teleological analysis minimally involves as a methodological heuristic. In the subsequent chapter, I will identify the particular functions up for dispute in the biological case. I will concentrate on that case, because that is where the teleological realist will hold that some teleo-

[5] Cummins and Davies do not style themselves in the business of teleological analysis at all and use instead the phrase "functional analysis." The distinction is verbal. Cummins' and Davies' intent is to avoid tying their form of functional analysis to etiology. Yet, teleology is no more synonymous with etiology than it is with feedback, entelechies or any of the other various proposals concerning how to understand teleological expressions like "in order to," "so as", etc. I place their proposals within the category of teleological proposals, because their sort of functional ascriptions, namely contributions to systematic effect, are of a traditional teleological sort. (See Cummins 1975 & 2002 and Davies 2001)

functional ascriptions are literally true. Additionally, if we can find that the biological case supports literal teleological ascription, then this will be a direct benefit later in returning to questions about naturalizing the mind.

As both Waddington (1968) and Mayr (1974) point out, whenever the beginning or initial properties of some process determine some end-state, those beginning or initial properties can be teleologically characterized. That is, we can characterize the initial or beginning properties with the typical teleological language of "for", "in order to", or "so as," etc. Teleological analysis is, however, only paradigmatically useful with respect to complex systems.[6] Following Godfrey-Smith (1996 & 2002), the distinction between complexity and simplicity is best thought of as the distinction between the heterogeneous and the homogeneous, respectively. From an explanatory standpoint, a system is complex not in the first instance due to the heterogeneity of its pieces and parts but due to the heterogeneous states of affairs under which it can fail to produce some effect. Following normal causal inference, the states of affairs under which the system will fail to produce the effect of interest exposes the causal factors needing to be in place for the production of that effect. A complex system is just one for which numerous causal factors play a role in carrying off some systematic effect. Characterizing the contribution of these distinct causal factors to the overall systematic effect makes for functional ascription. Identifying that something needs to be in place (say, oxygen) for the system to operate does not on its own expose what it does (say, oxygen is used in respiration); it is the latter that is the aim

[6] That functional analysis seems especially useful in biology might reflect nothing more than the complexity of biological systems. As Gates, a physicist, wrote,

> To begin with, one point should be made emphatically clear. Biology is by far a more difficult subject than physics. It is much more complex, and it has more variables. It is more difficult to do a controlled experiment and to understand the basic laws of nature which apply in biology than it is in physics. (Gates 1962, 2)

of the functional analyst. The utility of teleological analysis rests in characterizing the contribution of differing causal factors to some overall systematic effect, and it is of particular value the more complex the system or the more numerous the distinct causal factors involved.

Starting with some system and systematic effect, the teleological analyst proceeds by decomposing the system into individually less-talented components or homunculi, where the cooperative causal output of these components explains how the systematic effect of interest is produced.[7] (That the functionally characterized units are *individually less-talented* than the whole system prevents the hypothesis of dormitive virtues.) The functional analyst is a 'reverse engineer', and it is the methodology of reverse engineering that supplies the minimal analogue between the analysis of the artifactual and the nonartifactual. The analyst begins knowing that VCRs playback video or wings enable the flight of sparrows, and the task is to describe how the various components of the system enable such a systematic capacity. The sense of function invoked in such analysis is that of contributory effect: to describe the function of some component is to characterize its contribution to the production of the systematic effect. Functional ascription, then, is to specify the work, job, office, or role of some item or behavior in respect to the overall systematic effect. Such functions are teleological in at least the following respect: an item or behavior performs so-and-so function *in order to* or *for* the system to produce the systematic effect. In respect to the utility of such analysis, the "in order to"s, "for"s, "so as to"s, etc. need carry no further conceptual (or, for that matter,

[7] Dennett's (1978 & 1987) "design stance" is a classic example of such functional thinking. The characterization of function as contributory effect is, however, widespread in the literature; see, for example, Nagel (1957, 1961 & 1977b), Hempel (1959 & 1965), Cummins (1975), Boorse (1976 & 2002), Adams (1979), Lycan (1981), or Davies (2001).

metaphysical) weight than the specification of some component's job or office in respect to the operation of the whole system. If this minimal respect in which functions are teleological is insufficient for such functions to be genuinely teleological – that is, if the minimal sense of contributory effect is insufficient to render such functions more than mere analogues to genuine functions, it is at least the respect in which the utility of teleological analysis rests, namely in identifying the working parts of a system and describing the work those parts perform.

The individuation of the varied contributing homunculi can be more or less gross. By organizing systematic failures into general failure types, one generates a gross functional organization of the system. Each failure type exposes some distinctive contribution required for the production of the systematic effect. Such a gross homuncular characterization of the system is further subject to refinement by treating each gross homunculus as a system in its own right subject to the same decompositional analysis. And again, one proceeds by asking under what states of affairs does that gross homunculus fail to do its work in order to expose its functional architecture.

Often, the analyst does not proceed by such a top-down approach (i.e., gross functional architecture subject to further decomposition). It is common enough that we begin from the bottom and move upward. So, we might individuate some physical structure of the system (e.g., an ear, a leg, or a bump on the back of the head) and ask whether it plays a contributory role in respect to the systematic effect. The methodology is little changed from this direction. If the individuated component, structure, or behavior plays some role in the relevant system, then it will be the case that the suppression or removal of that component should produce some system failure. Characterizing the function of some component from the bottom-up is to characterize how the relevant suppressed effect contributes to overall system operation when everything proceeds as it ought.[8]

The simplest model of homuncular decompositional functional analysis would produce a gross functional flow chart of cooperating homunculi, each of which was then decomposed into further subhomunculi, and so on. Graphically, the functional flow chart would look like an inverted tree-like structure. However, homuncular decompositional analysis will often produce far more complex functional architecture, when, for example, we find ourselves with homunculi with conjunctive or alternative functions, with serial or circular homuncular architecture, or with redundant homunculi.

A homunculus with conjunctive functions is a homunculus which simultaneously performs more than one function. Coming from the bottom-up, it is easy to see how we would find ourselves with such functions. If we begin by delineating, say, some structural component of the system, that component might well have multiple effects feeding into differing higher-level functions or homunculi. For example, if we start by individuating the gross morphological feature of the fennec fox's ear, we would find that morphological feature simultaneously plays roles both in the fox's auditory system as well as its thermo-regulatory system.[9]

From the top-down, it is harder to see how we might have homunculi with conjunctive functions: if the homunculus is proposed to fill a functional gap, then it ought

[8] Wright's skepticism about system individuation leads both to his etiological proposal but also interestingly leads to his ignoring completely the top-down approach of functional analysis. (See Wright 1973) The subsequent chapter will show that Wright's skepticism about system individuation and specification of overall systematic effect is unfounded in the biological case.

[9] There is widespread agreement that the overly large ears of the fennec fox are a desert adaptation to keep it cool. (See Lariviere 2002 & Sheldon 1992.)

to be the case that we have at most one function for each homunculus. There are, nonetheless, a couple sorts of cases in which conjunctive functions arise in top-down analysis. For example, we have two homunculi the decomposition of which generates a subhomunculus for each, and each of those subhomunculi make the same sort of causal contribution to their respective homunculus by the same causal means. In such a case, a reasonable hypothesis is that one subcomponent of our system does work for two different higher level systems or homunculi. For example, with a multicellular organism, there will be a system to distribute nutritive resources to its interior and a waste elimination system from its interior. Both systems require a similar subhomuncular unit, and it would not unreasonable to think that there was a single subhomunculus, namely the circulatory system, servicing both. Alternatively, we might have two distinct hypothesized subhomunculi for some higher-level function, but we discover that the two functions prompting the individuation can be serviced by a component that had one single causal effect. E.g., for the chemical engine of an individual cell to carry out its functions, molecular inputs to that engine need to remain available and, so, constrained so as not to drift away. So, we might postulate some functional homunculus spatially constraining the appropriate molecular species. Additionally, the chemical engine also needs that supply of molecular inputs replenished over time. So, we might postulate a homunculus in the business of gathering or acquiring the relevant molecular species. Both functions might be carried out by a single structure if that structure could both be a container and an attractor, and we have such a structure in the cell wall. The cell wall can simultaneously both serve as a spatial constraint and via osmosis can serve to replenish depleted molecular inputs.

A homunculus possesses alternative functions when it has more than one function but the performance of these functions is not simultaneous. For example, the throat alternatively serves the functions of intake for oxygen and for foodstuffs. Again, from bottom-up analysis, it is easy to see how this situation might arise, say, if we start by individuating the gross morphological structure of the throat. And, we would find homunculi with alternative functions in top-down analysis for reasons similar to those prompting conjunctive functions. For example, we might characterize an energy acquisition system and a toxin avoidance system within E.Coli. Both systems independently require a subsystem to move the bacterium. It is a reasonable hypothesis that one subsystem, namely the flagellum, can service both of the higher level functions of energy acquisition and toxin avoidance, though it does not service them both simultaneously.

Functional circularity is a further complication. Functional circularity is not a type of function possessed by some homunculus but is rather a description of how the homuncular parts and pieces are related to each other. For example, the functional organization of metabolism within organisms is cyclic. The metabolic function or homunculi within organisms is accounted for by a sequence of causal contributions that form a closed or semi-closed loop of functional work. (For a review of the varied metabolic cycles, see Perry et. al. 2002 or Prescott et. al. 2002.)

Lastly, the development of functional architecture can include redundant homunculi, namely more than one homunculus doing the same work. Again, from the bottom up, it is easy to see how we might produce such architecture. For example, in the mammalian lung, either cyclooxygenase (COX-1 or COX 2) can serve in the signaling pathway in prostaglandin production. (See Bauer et. al. 2000) From the top down, I am unsure that one would have reason to postulate redundant homunculi, because the systematic failure type would involve each redundant homunculus failing to do its work. This is not, I think, anything particularly problematic, because actual practice will always involve a combination of both directional approaches.

I do not intend the above to be a full accounting of the ways in which we might produce more elaborate or complicated functional architecture.[10] Rather, I only intend the above to stress that it is not somehow a theoretical problem that functional analysis does not always generate clean and tidy inverted tree-like structures.

In providing a homuncular analysis and developing the functional architecture that explains some systematic effect, one is in the business of providing and developing a mechanistic hypothesis. As Brandon (1984, 346) suggests, to provide a mechanistic hypothesis is just to offer a model of the process underlying some effect of interest. The provision of a homuncular flow chart underlying some systematic effect – that is, a model of how varied contributing interactions of homunculi produce the relevant effect – is exactly to provide a mechanistic hypothesis. That hypothesis finds support when the analyst can identify effects that would supply the relevant causal contributions. That the homunculi are themselves functionally characterized and subject to further functional decomposition makes such a flow chart no less a mechanism; it is, after all, a model of the causal factors underwriting a systematic effect and which is arrived at by standard causal inference. When we decompose the homunculi composing a mechanism, we are then are offering further mechanistic hypotheses to explain how the relevant homunculi

[10] For a far more detailed analysis of functional architecture, see Wimsatt (2002).

do their work. We should avoid the simple-minded thought that nature carves up neatly into two levels, the mechanistic and the functional. As Lycan writes,

> One and the same space-time slice may be occupied by a collection of molecules, a piece of very hard stuff, a metal strip with an articulated flange, a mover of tumblers, a key, an unlocker of doors, an allower of entry to hotel rooms, a facilitator of adulterous liaisons, a destroyer of souls. Thus, we cannot split our theory of nature neatly into a well-behaved, purely mechanistic part and a dubious, messy vitalistic part better ignored or done away with. (1981, 33)

One last note in respect to functional analysis. The analyst, in decomposing the systematic effect, need not be limited in the assignment of function to individuated components of the system itself. Cummins (1975) and Hempel (1959), for example, when writing on functional analysis, seem to take it that decomposition is limited to the system itself, and so they restrict functional analysis to components of the *containing* system. However, no system operates in a vacuum. The aim of decompositional functional analysis is to account for how the systematic effect is produced, and that requires the specification of individual contributions from both features internal and external to a system. The specification of functions to the environment proceeds along the just same lines as it does in respect to internal features, namely by way of specifying contributory effect to some systematic effect. Inclusion of external features in functional analysis is often important, because, in order to understand the work performed by internal features, we have to understand how they cooperate or coordinate with external features. For example, Vogel, in discussing water filtration by sponges, writes:

> If the flagellar were inoperative, would water pass through a sponge anyway? Or, to put the matter in more realistic terms, does flagellar action account for *all* the water passing through a sponge, or can ambient water currents make a contribution to filtration? It appears now (Vogel 1977, 1978) that not only do

> ambient currents help, but that structure of sponges is most exquisitely adapted to take advantage of such currents… (Vogel 1981, 190)[11]

As we see with Vogel, often when accounting for systematic effect, a functional analysis of the components alone will fail to account for that effect. In such cases, some of the functional workload is being carried by the environment, and it is not unreasonable to characterize external conditions functionally as contributing to overall systematic effect.[12] In fact, that part of systematic operation can depend on external features carrying some of the functional burden generates a sort of methodological constraint. Whether with Vogel's principle ("Do not develop explanations requiring expenditure of metabolic energy until simple physical effects are ruled out.") or in Clark's cognitively-oriented version ("know only as much as you need to know to get the job done.")[13], the general methodological advice amounts to the same: avoid hypothesizing internal architecture when the functional workload can be carried by the environment.

---

[11] Vogel (1981) is packed with a number of nice biological examples of cases where, in order to understand some internal feature's contribution, we need to understand how it works in cooperation with features external to the *containing* system.

[12] Brandon also provides a nice example where decompositional analysis drives one not into smaller units of the system but outside to that in which the system is embedded:

> Consider the movement of the intake value in a normal piston engine. How do we explain that? Certainly not in terms of the molecular and atomic parts of the valve. At minimum the mechanistic explanation must make reference to the movement of the camshaft(s). I would think that a complete explanation would also involve the movement of the pistons, the intake of the fuel-air mixture, the firing of the spark plugs, the explosion of the mixture and the exhaust of the spent gas. That is, we explain the movement of the intake valve by embedding it in a larger mechanistic system. (Brandon 1984, 348)

[13] Vogel (1981, 182) & Clark (1989, 64). The full statement of what Clark dubs his "007 principle" is the following:

> In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. That is, know only as much as you need to know to get the job done.

Such an inclusion of environmental conditions as contributory to systematic effect would be odd or just confused if one tied, as has been frequently the case in the literature,[14] the utility of functional ascription to accounting for the presence or existence of some item or behavior within a system. On such a view, to ascribe a teleological function is more than the specification of a contributory effect. It is, in addition, to account for the presence of some item *within* or behavior *of* that system. However, accounting for the presence of some item or behavior is not part of a minimal conception of teleological function (nor, I will later suggest, is it part of any thicker conception of teleological function). Even in the artifactual case, that some item or behavior has some job or office to serve is often independent of the reason for its presence. That is, an artifact need not be the result of explicit intentional construction. All that is required is that one does use it for some purpose, not that one creates it for that purpose. The stick on the forest floor becomes a club by my appropriation of it for that purpose. Such an appropriation need not even require that I, in fact, wield it as a club but only that I am prepared to do so when danger presents itself.[15] Nor even when some item is the result of explicit manufacture does that suffice in the artifactual case for it to possess a function. Whatever function Stonehenge served its creators, it clearly no longer serves that function for the present population. Or, alternatively, I have a key to some now unknown lock that I use as a bookmark. The explicit manufacture of that key does not provide its function; its function is to be a bookmark and nothing more. The intertwining of

[14] See, for example, Ayala (1970), Braithwaite (1955), Hempel (1959), Lehman (1965), Nagel (1961), Scheffler (1959), or Wright (1973).

[15] Sorabji (1964) nicely presses that artifacts need not be produced or made by anyone and that appropriation for a use more than suffices. (See McLaughlin (2001, chapter 3) for an extended discussion of this point.)

contributory effect and an explanation of the presence of some item is likely to be explained by the fact that providing the reason for some artifact's presence often is just to provide the function of that item.[16] But, providing the reason for something's existence or presence is neither necessary nor sufficient for its possession of function.

Teleological ascription minimally (that is, all that is required for it to serve as part of a methodological heuristic) is just the description of component contribution to overall systematic effect. Its value rests in providing a way to decompose complex systems into working parts and describing the work of those parts. With these general comments in hand, I want to turn now to the particular functions generated in the biological case.

[16] Cummins (1975, 746) provides a similar diagnosis for the tendency to treat functional explanation and "why is it there" explanations as one in the same.

Chapter 3

TWO SOURCES OF FUNCTIONAL ASCRIPTION IN BIOLOGY

The intent of the previous chapter was to locate what at least should be common ground between the realist and the antirealist in respect to functional ascription. Both should be able to agree that human theorists find some value in understanding complex systems by characterizing the mechanisms of those systems via contributory effect. The realist position should be just, contra the anti-realist, that at least some of those functional characterizations are literally true – that is, at least some of the nonartifactual does possess genuine teleological functions. I want now to extend that common ground from the previously excessively abstract to a somewhat less abstract application to the biological. Once that is in place, we should be in a position to ask which, if any, of the functional ascriptions in the explanation of biological systems are literally true. Though the intent is to maintain a common ground, those teleological realists, who have wed the truth of nonartifactual functional ascription to etiology or selectionist history, will find in what is suggested below the role of selectionist history and etiology radically downplayed. Such realists should, nonetheless, be able to accept the common ground as such – an arena in which functional analysis is of some utility – and see themselves as picking some piece of that territory as where functional ascription is literally true in addition to being useful.

To apply the previous chapter in a single sweep to the biological is to suppose an abstract system type of which each biological entity is an instance. But, to suggest some abstract system type of which all biological entities are instances would seem to suggest some definition of what constitutes "life". While I do intend to offer an abstract system type and systematic effect types to serve as the source of biological functional explanation and ascription, I do not intend by that to be offering anything like a definition of "life". I do not intend, then, to provide some identificatory and distinguishing marks that set the biological apart from everything else. I intend, instead, only to point out what is particularly noteworthy and in need of explanation in respect to those varied physical-chemical entities lumped together in biological textbooks. As Kauffman suggests, it is an open question whether the abstract system type of which biological entities are an instance extends to cover other processes and structures in the universe such as "lifeless galaxies, stars, the molecular clouds in galaxies, or lifeless planets". (2000, 106)[17] In fact, so much the better if it does for that is just more reason to resist latent vitalist tendencies.

[17] An interesting speculation in this vein is Smolin's (1997) suggestion that whole universes reproduce daughter universes from the creation of black holes, that these daughter universes inherit with variation the physical constants of their parent universe, and that, in turn, differing physical constants make some universes more or less "fit" in respect to which predominates the multiverse. Smolin's speculation presents no problem for this chapter's thesis. It is just what we should expect in specifying an abstract system type: there might be many unforeseen ways to realize it. The only reasonable mark that the offered abstract system and effect types are appropriately distinctive, as opposed to being too inclusive, overly vague, or just plain vacuous, is that their specification would lead one to the sort of functional ascriptions that biologists in the field make. I find Smolin's speculation intriguing for another reason as well. I will later suggest that selectional histories are one way to generate the nonartifactual normativity required for teleology. Smolin's speculation might well have the surprising result that our physical laws (at least in respect to the physical constants) turn out to be laws in the normative sense of that term.

I. Autocatalysis as a biological systematic effect

a. *Autocatalysis and work cycles*[18]

For all the variance in biological types, they are just physical-chemical systems. Each biological type is, nonetheless, capable of a trick that much of the rest of the physical-chemical universe is incapable: members of biological types produce more members of that same type by sustaining themselves over some duration of time through the repeated extraction of thermodynamic work. The explanation of how biological types accomplish this trick is, I will suggest, one of the key sources of functional ascription in biology. The simplest example of this trick in respect to replication[19] is an autocatalytic reaction. I suggest that we generalize that simple example to create a model of a system type with the systematic effect of replication, and I will then suggest that generalized model applies to all instances of biological typing.

Autocatalysis is just the catalytic task of producing more of the same. So, take some molecular species A, A is autocatalytic when one catalytic task it performs is the production of a further A. For example, some molecular species A is capable of ligating two other molecular species (A' and A'') into A. That A is autocatalytic does not require that the only catalytic task it can perform is the production of further A's. A's might

---

[18] This section is heavily indebted to Kauffman (2000). Much of the key ideas contained herein are drawn directly from that work. Much of the emphasis below on autocatalytic systems or collectives as the model for living systems is also reflected in works of Ganti (2003) and Bechtel (in press).

[19] Following Morowitz et. al., "replication is defined as any energy-requiring growth process in which an organized assembly of molecules produces similar assemblies over time. We do not require sequence-mediated information transfer, nor a precise doubling of the assemblies." (1988, 281)

catalyze any number of other molecular species; it is just that it is capable of producing more A's that makes it autocatalytic.[20]

Let's complicate the simple autocatalytic example slightly. Instead of molecular species A catalyzing A's, assume that A catalyzes some molecular species B (say, by ligating B' and B'') and that B's catalyze A's (say, by ligating A' and A''). A's are no longer autocatalytic. Rather, we might think of A's and B's as forming an autocatalytic set or collective. Though both A's and B's might have any number of other catalytic tasks besides the respective production of B's and A's, A's and B's achieve a sort of catalytic closure: the catalytic tasks of producing A's or B's each require their own product to do the relevant task – that is, in order for A's to catalyze B's, A's need to be catalyzed by B's, and vice versa. The set of catalytic tasks achieving this sort of catalytic closure is in effect autocatalytic, because it creates a closed repeating loop of catalytic tasks. We could further still complicate matters by putting any number of intermediate and tangled pathways leading from A to A. No matter how complicated that whole set of catalytic tasks, the set remains autocatalytic insofar as catalytic closure is maintained.

Let's return to the simple autocatalytic reaction (A catalyzes A): the molecular species A is a system with an autocatalytic effect. The A-B autocatalytic collective also forms a system with the systematic effect of autocatalysis. More generally still, any autocatalytic collective – a catalytically closed task set – will form a system with the

[20] Prions are a nice example of such simple autocatalysis. Prions are supposed to be the cause behind a number of infectious, genetic, and spontaneous disorders, e.g., scrapie, Creutzfeldt-Jakob disease, Kuru, etc. Unlike a virus, a prion has no genetic material. Instead, the prion is a single protein with the same amino acid sequence as proteins normally produced within the cell. The only difference is conformational – that is, the three-dimensional shape differs between the normal type and the prion type of the same amino acid sequence. The interesting bit is that, when prions come into contact with normal type proteins, they will induce the normal type to alter its conformation to that of the prion type. We have then a bit of autocatalysis and replication that requires no genetic code and no intermediate steps. (See Prusiner (1995) for a nice review.)

systematic effect of autocatalysis. We can distinguish then the genus of autocatalytic systems into species in virtue of the varied catalytic task descriptions that comprise catalytic closure.

That the A-B autocatalytic collective has the systematic effect of autocatalysis introduces functional hierarchies: A production and B production are the lower level tasks required to perform the higher level task of autocatalysis. There is no reason why such hierarchies cannot be indefinitely more detailed in decomposition. For example, the task of B production by A's is accomplished by A cleaving some molecular species C into C' and C'' that in turn jointly catalyzes B from, say, some molecular species D. We have further decomposed the A-B system by decomposing A's task into smaller task units. Systematic typing can be performed at the various levels of task hierarchy, since each provides some catalytic task description that ultimately adds up to the systematic effect of autocatalysis.

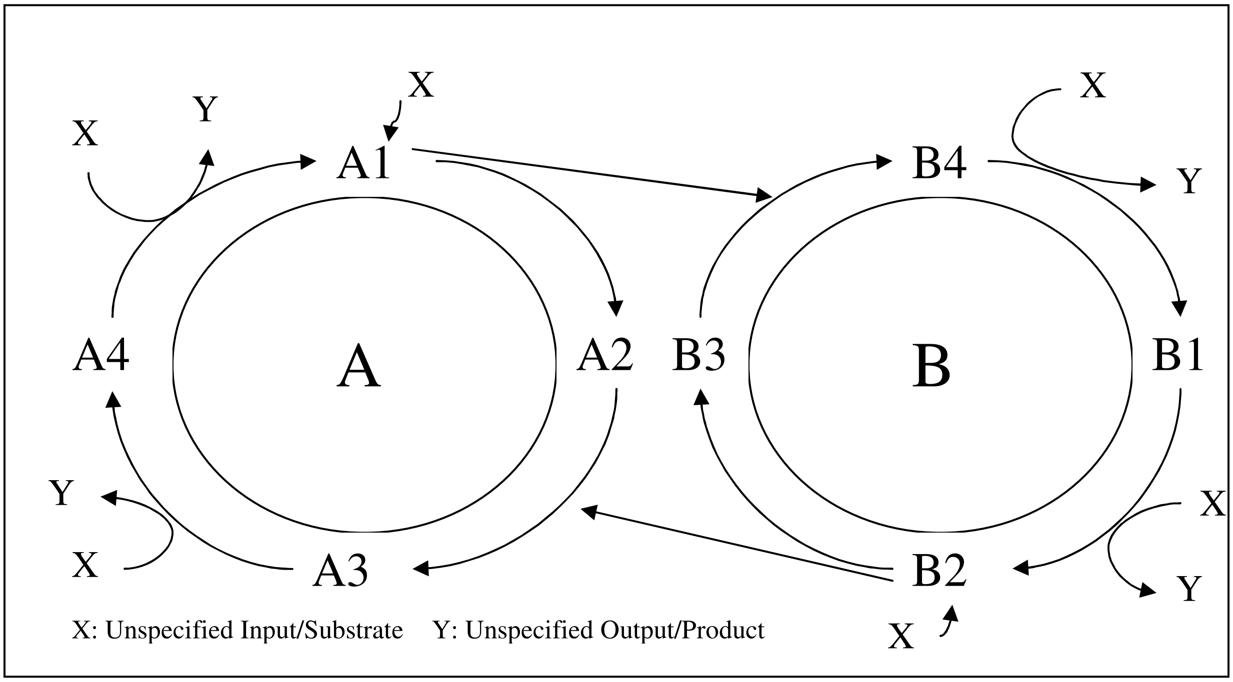
That within the genus of autocatalytic systems is any number of varied task hierarchies achieving the same systematic effect provides two morals: 1) lower level hierarchical variation can be irrelevant to system typing; and, 2) the specification of a higher level task need not specify some particular structural product. The first should be obvious given that the genus of autocatalytic systems above contains two species (the A-A collective and the A-B collective), so let me say something about the second. While the task specification of the A's and B's of the A-B autocatalytic collective specifics particular structural products (B's and A's, respectively), the specification of the system effect type – autocatalysis – does not. The catalytic closure that is autocatalysis is achieved just by having it be the case that the performance of some task presupposes its product, and its product only need be something capable of performing the same task, i.e., producing something presupposed for its own catalysis.

I want to suggest that these two morals do not apply solely to the highest task specification, namely autocatalysis, but can be recapitulated at any hierarchical level. In respect to the first, systematic typing at some lower hierarchical level via task description can ignore variances in some still lower level task specification. In respect to the second, lower level task descriptions need not be of particular structural products but can rather be descriptions of that product's task specification. For example, for the A-B collective, rather than saying that A's task is the production of molecular species B, we can say that A's task is the production of something capable of producing A's; that A's produce something capable of producing A's provides, after all, for catalytic closure. We can do the same with A's themselves. Now, we have a way to type all two-step autocatalytic collectives together. Of course, these two steps might sit on top of any number of varied hierarchies, and we can recapitulate these morals further down. Systematic type individuation is itself then hierarchical organized. Which level of systematic typing ought to be chosen will very much depend on the detail required for the explanatory project at hand.

I want to stress that generalized or non-structural task descriptions are not a bit of philosophical slight of hand. Such non-structural task descriptions are an essential part of contemporary molecular biology. For example, the binding of antibodies to receptors works as one might think of a key to a lock. The antibody binds to a receptor with a complementary shape. But, there is some wiggle room in which receptors an antibody will bind – that is, the antibody's capacity to recognize complementary shape is finite.

So, for any antibody there is a group of receptors in which it will fit, and vice versa, for any receptor there are a number of antibodies that will fit within it. There are intersections in structural molecular shape space between that of antibodies and receptors that carry out in effect the same binding task, and we can describe the binding task without needing to specify a particular structural product or substrate. That this is the case is important in the very least to how much of present drug research is conducted. For example, our capacity to build estrogen mimics depends on the fact that a variety of structures can do the same work. (Kaufman 2000, 11-13)

Just to make it explicit, let's notice an architectural type that will be of some import later. Suppose two autocatalytic collectives, say A and B. Each autocatalytic collective is some 4 step cycle: e.g., step 1 catalyzes step 2, step 2 catalyzes step 3, step 3 catalyzes 4, and 4 catalyzes 1. For A, suppose that the initial task is the cleaving of a molecular species into two. The first of these products is capable of the second step in the autocatalytic process. The second of these products, however, binds with the fourth step in the B autocatalytic collective to produce the initial step in the B cycle. We have then an autocatalytic collective A that produces a byproduct that in turn feeds the autocatalytic collective B. Further, let's suppose that the second step of the B cycle cleaves itself into two products: the first will be the third step in the B cycle, and the second will bind with the third step of the A cycle to produce the fourth step in the A cycle. (See figure below). Our two autocatalytic collectives are now linked to one another such that a byproduct of each is required for the other. Our two autocatalytic collectives form then a single autocatalytic collective. I note this sort of case, because it will be important later to think of living systems as autocatalytic collectives comprised of further autocatalytic collectives and that comprise further higher level autocatalytic collectives.



If the autocatalytic collective is to serve as a model of biological systems, two further features need to be added, namely directionality to autocatalysis and the capacity to produce thermodynamic work cycles. We will need the first, because, stealing from Kauffman (2000, 69), organisms do not revert back into their foodstuffs – that is, there is a direction to development, growth, sustenance, and ultimately organismic reproduction. In respect to the second, biological entities are not, as is often remarked, passive. They are active agents in their environments initiating operations for development, persistence, and reproduction. Each of these operations require some way of constraining free energy to extract thermodynamic work repeatedly, and so we will need to include some characterization of how a system can build constraints to extract work repeatedly.

Let's start with directionality. In a chemical reaction, some atoms or molecules (substrates) undergo transformations leaving some product(s). However, the direction of the flow of matter from substrate to product depends upon displacement from equilibrium. For example, in a closed system with a single substrate-product reaction (A-B), the concentration of A's to B's will tend toward equilibrium. So, at equilibrium, A's will convert into B's and vice versa with no net production of A or B. When displaced from equilibrium with more A's than B's, A's will have the net effect of producing B's; thus, the whole system will tend toward equilibrium. If the concentration of B's is higher than that of A's, then inverse scenario will take place. These facts are just what follows from the second law of thermodynamics. If we open the system and couple a free energy source to the reaction, we can drive the reaction beyond equilibrium. Further, if this free energy source is coupled to only one of the reactants, say A, then we can provide a directionality to the chemical reaction and build up a concentration of, say, molecular species B. Let us restrict our attention to only those autocatalytic collectives that exhibit such a directional flow.

Let's turn to thermodynamic work. Take a closed cylinder filled with a noble gas and assume that the gas molecules are largely concentrated in one end of the cylinder. By the second law, over time the molecules will tend to become distributed throughout the cylinder. From the first law (i.e., energy is conserved), the positive change in entropy results in a negative change in free energy. (Or, more intuitively, as the molecules disperse through the cylinder, there is decreasing net dispersal available.) The disbursement of the molecules is spontaneous – that is, it does not require the input of some further energy source. Such a process or reaction is called "exergonic". An exergonic process involves the release of free energy (i.e., free energy is diminished), but on its own such a process does no thermodynamic work. As Kauffman writes,

> What is work? Physicists have an answer – work is force acting through distance – given by a number, or scalar, representing the sum of forces through the distance. But it will turn out that in any specific case of work, the specific process is organized in some specific way. Work is more than force acting through distance; it is, in fact, the constrained release of energy, the release of energy into a small number of degrees of freedom. (Kauffman 2000, 83)

In our cylinder, the dispersal of the molecules, though a release of free energy, is not a constrained release of energy within some degrees of freedom. To extract work from our cylinder, we need a way of constraining the release of energy generated by the exergonic process. So, for example, put a partition with a small opening in the middle of our cylinder and in that opening place a fly wheel. The partition and fly wheel both place constraints on the release of free energy; entropy can only increase by the molecules passing through the opening and pushing on the fly wheel. By moving the fly wheel, the release of free energy is constrained in such a way as to extract, in this case, mechanical work.

Let's notice a few things about the above. First, and perhaps obvious, in an empty cylinder, our fly wheel does not move. The fly wheel requires the input of free energy to move it. The moving of the fly wheel is an endergonic process, namely one in which an input of free energy is required to drive it to completion. Second, a wheel with a single blade and capable of only a quarter rotation can only extract work on one occasion, namely the first time a molecule strikes it. But, a fly wheel constructed such that each rotation brings the blade back to the starting position, say at the top, will extract work repeatedly as entropy increases. A system that extracts thermodynamic work and in so doing resets itself, as our better flywheel does, is capable of work cycles. Third, the construction of constraints, the partition and the fly wheel, takes work. Thus, it takes work to extract work. With the better fly wheel, we have a simple case of this: the extraction of thermodynamic work is put to the work of resetting the flywheel so that flywheel can extract further thermodynamic work.

Now, an autocatalytic collective that links exergonic and endergonic processes will be one that both 1) has directionality and 2) will be capable of one or more thermodynamic work cycles. (In respect to the second, autocatalysis driven by the linking of exergonic and endergonic processes is like the flywheel being reset by its own extraction of thermodynamic work.) Let's restrict the class of autocatalytic collectives to just these, namely those that do so link exergonic and endergonic processes to extract thermodynamic work and can do so in cycles.

b. *The autocatalytic collective: a model of living systems*

This brings us to Kauffman's suggestion that such a class of autocatalytic collectives will do as the model for all living systems. He calls such a system or model an "autonomous agent", namely "an autocatalytic system able to reproduce and to perform one or more thermo-dynamic work cycles." (2000, 49). Why think that the autonomous agent will serve as a model for all living systems? Well, on first blush, the autonomous agent looks to have just those distinctive features pointed to by cell theory. As Bechtel notes,

> One of major claims of cell theory, as it developed in the 19th century through the endeavors of Schleiden, Schwann, Virchow, and others, is that the cell is the fundamental living unit. [Bechtel 1984; 2006]. Two things are salient to this claim. First, the cell is a unit – it is an entity whose identity is maintained over time despite exchanges in matter and energy with its environment. Second, as a living entity, a cell is an active agent. Unlike a rock or crystal, for example, it initiates operations that affect both itself and the environment. (Bechtel, in press, 4)[21]

[21] Bechtel (in press) adds a further claim that cells must be capable of repair. His amendment is, I think, better left out, because it is a contingent feature of cellular operation. As Morowitz et al. (1988) argue, what

Both of these points are nicely realized by conceiving of the individual cell as an autocatalytic collective comprised of autocatalytic collectives that link various endergonic and exergonic processes.[22] First, an autocatalytic collective is something that by its nature maintains its stability across time. Second, if we link that stability to the construction of constraints for the extraction of thermodynamic work, then we capture the way in the which cells are active. And, if the latter is to work, then we need the cell building constraints by the doing of thermodynamic work, leading us back again to the autocatalytic collective which is just the sort of cyclic organization required. So, the following claim by Kauffman is not all that surprising:

> All free-living cells are, by this definition, autonomous agents. To take a simple example, our bacteria with its flagellar motor rotating and swimming upstream for dinner is, in point of fact, a self-reproducing system that is carrying out one or more work cycles. So is the paramecium chasing the bacteria, hoping for its own dinner. So is the dinoflagellate hunting for paramecium sneaking up on the bacterium. So are the flower and flatworm. So are you and I. (Kauffman 2000, 8)

But, before jumping to Kauffman's conclusion that you and I and all living systems are autonomous agents (autocatalytic collectives able to reproduce and do at least one thermodynamic work cycle), let's go somewhat more slowly and attend to the facts on the ground.

The basic energy cycle of every terrestrial living cell forms an autocatalytic collective linking exergonic and endergonic processes. The free energy input varies from aerobic respiration, anaerobic respiration, fermentation, or photosynthesis. But, despite

---

is necessary for a cell is, first, phase separation in order to create a closed internal environment in which directed chemical reactions can occur ("the cell is a unit") and, second, a chemical reactor ("a cell is an active agent").

[22] Of course, viruses and prions are not cellular. Microbiologists (see, for example, Prescott et. all. (2002) or Perry et. al. (2002)) do not classify viruses and prions as alive due to the absence of metabolism. This fits with the suggested model above, because absent metabolism neither is capable of thermodynamic work cycles.

the variance, that input is used to bind ADP (adenosine diphosphate) and Pi (orthophosphate) to form ATP (adenosine 5'-triphosate). So, we have an exergonic process (e.g., photons falling on the planet) linked to an endergonic process (e.g., the binding of ADP and Pi). ATP is a high energy molecule; it is so not in sense that there is a great deal energy stored in some bond but that its breaking down is highly exergonic. ATP exergonically breaks down into ADP and Pi almost completely and does so because it readily transfers its phosphate to water. ATP's breakdown is further coupled with endergonic processes to extract chemical, mechanical, and transport work. (Prescott et. al. 2002, chapter 8, & Perry et. al., 2002, chapter 8)

That the basic energy structure of terrestrial living cells is organized as an autocatalytic collective linking exergonic and endergonic processes is not likely just a contingent fact about terrestrial life. Given the considerations from thermodynamics, it should be of no surprise that a system, such as a living cell, will sustain itself over time through thermodynamic work cycles via autocatalytic organization; it is difficult to imagine what other organizational structure would do.

These comments only take us as far as metabolism, but from them it is at least the case that living cells are comprised of autocatalytic collectives. We can, however, push the case further still and suggest that living cells as a whole are autocatalytic collectives. A living cell persists by the repeated extraction of thermodynamic work, but again the extraction of thermodynamic work requires the building of constraints and these in turn require thermodynamic work. The autocatalytic collective looks to be the right sort of functional organization, just as it is with metabolism, to explain how the whole cell is capable of this feat.

Ganti (2003) suggests, for example, that minimally a model of extant cells requires at least three subsystems:

> Three subsystems – the cytoplasm, the membrane, and the genetic substance – are present in every cell. There are many other subsystems with specific functions which are present in different types of cell, but these are not present in every cell. Therefore we can assume that if we establish the functions of these three subsystems, construct abstract models of them, and finally organize them into a single system, we will obtain an abstract model of the cell. (Ganti 2003, 83)
>
> … in seeking for the abstract minimal system of the cell, we have to construct abstract models of the following three subsystems:
>
> 1. [For the cytoplasm,] A soft (chemical engine) fulfilling the task of a chemical motor, i.e. of performing chemical work. This chemical motor must have a functionally stable inner organization, must be provided by chemical regulation, and must be capable of synthesizing chemical substances for itself as well as for other systems.[23]
> 2. [For the membrane,] A soft (chemical) system which is capable of spatial separation, of being selectively permeable to chemical substances, and of growth in the presence of raw materials.[24]
> 3. [For the genetic substance,] chemical system which is capable of storing and copying information, i.e. capable of self-reproduction in the presence of raw materials.[25] (2003, 84)

Ganti (as well Kauffman (2000, 101-3) and Morowitz et. al. (1988)) have offered models of these subsystems; each of which is modeled as an individual autocatalytic collective. Further, the three are interconnected in such a way that they form a whole autocatalytic collective. Below is just a summation of Ganti's "chemoton" model of the individual cell:[26]

---

[23] That is, we need a basic system of metabolism or something capable of thermodynamic work cycles.

[24] The membrane, as Morowitz et. al. stress, is required to generate "an entity thermodynamically separated from the environment," thereby "some of the reactions occurring therein [within the membrane] can be thermodyanically improbable in an equilibrium system." (1988, 281 & 283)

[25] On the basis of extant cells, Ganti assumes that some information-storing system (e.g., DNA) is required for replication, but that assumption is questionable when applied to the protocells from which extant cells have evolved. Morowitz et. al. (1988), Wächtershäuser (1988), and Bechtel (in press) have argued that only the first two subsystems are required for a minimal protocell, because the information driving replication can be built into the structural features of these subsystems.

[26] For the detailed version of the chemostat model, see Ganti (2003 chapter 3).

A chemoton consists of three different autocatalytic (i.e. reproductive) fluid automata, which are connected with each other stoichiometrically. The first is the metabolic subsystem, which is a reactive network (optionally complicated) of chemical compounds with mostly low molecular weight. This must be able to produce not only all the compounds needed to reproduce itself, but also the compounds to reproduce the other two subsystems. The second subsystem is a two dimensional fluid membrane, which has the capacity for autocatalytic growth using the compounds produced by the first subsystem. The third system is a reaction system which is able to produce macromolecules by template polycondensation using the compounds synthesized by the metabolic subsystem. The byproducts of polycondensation are also needed for the formation of the membrane. In this way, the third is able to control the working of the other two solely by stoichiometrical coupling. As they work, the three fluid automata become a unified chemical supersystem through the forced stoichiometrical connections. This means that they are unable to function without each other, but the supersystem formed by their cooperation can function. (2003, 4)

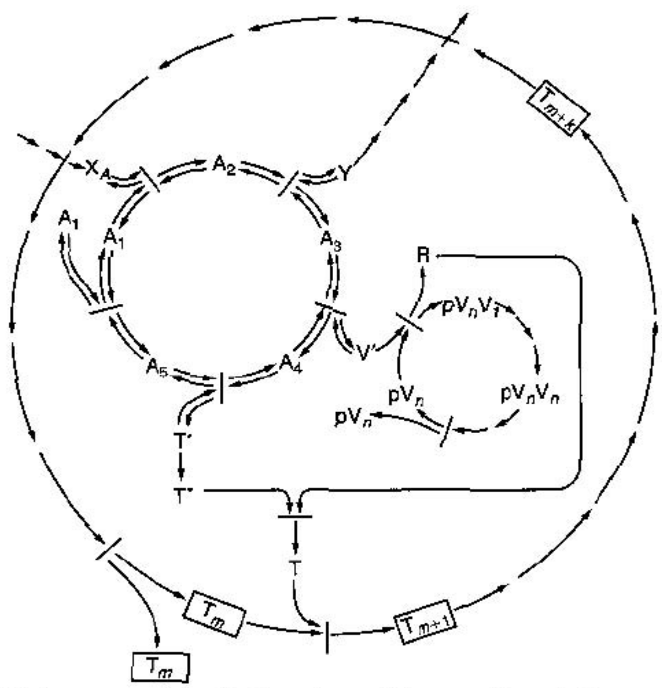


Fig. 1.1 Minimum model of chemotons. *Three* self-producing systems are coupled together stoichiometrically: cycle $A \rightarrow 2A$, *template polycondensation* $pV_n \rightarrow 2pV_n$, and membrane formation $T_m \rightarrow 2T_m$. This coupling results in a *proliferating program-controlled* fluid automaton, known as the chemoton.

Figure from Ganti (2003, 4)

Ganti has provided above a model for how not only to understand the essential features of an individual cell as comprised of autocatalytic collectives but also how to understand the whole operation of the cell, as a supersystem, as an autocatalytic collective persisting through the cyclic extraction of thermodynamic work. Again, given the considerations from thermodynamics, it is difficult to imagine what other sort of functional organization – that is, other than the autocatalytic collective – would be capable of the relevant work.

But, the other important fact about individual cells over time is that they succeed in the production of more individual cells. Cellular persistence and stability are contributory effects toward the production of the autocatalytic event leading to a new individual cell itself capable of continuing the autocatalytic process. Individual cells are then comprised of autocatalytic collectives (think here of metabolic cycles like the Krebs' cycle), are autocatalytic collectives (the supersystem of the chemoton), and (by the action of replication) comprise autocatalytic collectives.

This gets us as far as including individual cells, especially free-living individual cells, within the model of autonomous agency, but we would still seem to be some distance from the multicellular organism. Can we treat the multicellular organism as comprised of autocatalytic collectives, as itself an autocatalytic collective, and as comprising an autocatalytic collective which realizes the building of more multicellular organisms? I think that we can on much the same considerations as with the individual cell.

A general feature separating the multicellular organism from, say, a cooperative aggregate of individual cells is the separation and specialization of germ and somatic cell lines. Following Michod et. al.:

> We use the terms "germ" and "soma" in the sense of there being two kinds of cells in a multicellular group, cells that are specialized in contributing to the next generation of individuals and cells that are specialized in the vegetative functions and do not directly reproduce the next generation of individuals. Even organisms often regarded as not having a germ-line, such as plants, have cells specialized in reproductive and vegetative functions, and so meet our criteria of reproductive specialization. Specialization of cells in reproductive and vegetative functions is an almost universal feature of multicellular life. (Michod et. al. 2003, 96)

The soma-germ differentiation[27] results when cellular immortality becomes decoupled from totipotency. (Your skin cells, say, might be immortal – that is, they can continue to divide over and over again,[28] but they are not totipotent – that is, they will not, if grown in a Petri dish, form a new person.) Such decoupling occurs just when intracellular conflict becomes more costly than group level fitness benefits within a cooperative aggregate of individual cells. As a result, the group develops (on an evolutionary time-scale) a set of mechanisms (e.g., apoptosis, mutation rate, reproductive mode, etc.) to limit such lower level fitness conflicts; the effect of which is to limit totipotency to a certain set of individual cells, e.g. germ line. Intracellular competition in respect to

[27] Volvox, a noncladistic group of green algae, presents a nice example of a recent multicellular organism for which the only cell type differentiation is that of germ and soma. See Kirk (2005, 2003, & 1998) & Miller (2002).

[28] Your skin cells are not in fact immortal in this sense during normal operation in the body. After a certain number of divisions, a cell line will become senescent (i.e., it will stop replicating). This senescent mechanism can be suspended, however, and cell lines can be maintained indefinitely. Immortalizing a cell line, i.e., suspending the senescent mechanism, is a standard research procedure for in vitro experimentation.

relative fitness is, thereby, diminished, if not removed. (Michod et. al. 2003 & Michod 1996, 1997, 1999)[29] As Michod et. al. write:

> Indeed, the essence of an evolutionary transition in individuality [from free-living cells to multicellular organisms] is that the lower-level individuals must "relinquish" their "claim" to fitness, that is to flourish and multiply, in favor of the new higher level unit. This transfer of fitness from lower to higher-levels occurs through the evolution of cooperation and mediators of conflict that restrict the opportunity for within-group change and enhance the opportunity for between-group change. Until, eventually, the group becomes a new evolutionary individual in the sense of being evolvable – possessing heritable variation in fitness (at the new level of organization) and being protected from the ravages of within-group change adaptations…. (Michod et. al. 2003, 96)

With the above in mind, we should be able to say that the multicellular organism is composed of autocatalytic collectives, namely the individual cells. It is an autocatalytic collective via the cooperative interaction of its cells for sustenance and persistence. And, importantly, the multicellular organism is part of an autocatalytic collective by being totipotent.

Groups of varied species of organisms can also form autocatalytic collectives, and one way to think of stable ecosystems is along such a model. Let's look to a very minimal example of an ecosystem, namely one with two or at least few actors interdependent on one another. Biologists distinguish a variety of cross-species interaction types under the heading "symbiosis". I want to concentrate on just one of these, namely obligatory mutualism,[30] as the minimal model of an ecosystem. Mutualism is an reciprocally beneficial interaction between two types of organisms, and that mutualism is considered

[29] Some organisms are neither clearly unicellular nor multicellular. Terrestrial slime molds, for example, spend their early life as free-living cells. Latter, they will latter aggregate to form a stalk for reproduction, and importantly only some the amoebae within the stalk will get the opportunity to form fruiting bodies. See Gross (1994).

[30] Sometimes "mutualism" is just defined as an obligatory reciprocally beneficial interaction, and nonobligatory mutualism goes by the name "protocooperation". See, for example, Prescott et. al. (2002, 598).

obligatory when both actors are metabolically dependent on one another. In other words, when two species are obligatory mutualists, it is not possible for them to live without one another. A classic example of such obligatory mutualism is lichen.[31] [32]

Lichen is not a single organism but is a biological entity "composed of a fungal partner, the mycobiont, and one or more photosynthetic partners, that may be either a green alga or a cyanobacterium." (Nash 1996, 1) Before Schwendener's (1869) discovery of the dual nature of lichen, it is not surprising that lichens were taxonomized into genus and species. But, even following that discovery, there remained a very strong tendency to think of lichens as individual organisms rightly classified into genus and species. In fact, Prescott et. al. continue to suggest just this:

> The remarkable aspect of this mutualistic association is that the morphology and metabolic relationships are so constant that lichens are assigned generic and species names. (2002, 599).

But, among lichenologists, such talk is treated with disdain. For example, Tehler writes,

> Lichens are not organisms. Lichens are small ecosystems, associates with two or more components: an algal producer and a fungal consumer. Consequently, lichens as such cannot be used in phylogenetic classifications because they have no phylogeny. (1996, 217)

Lichen classification, instead, is based on the fungal component of the lichen. But, what I want to draw attention to is Tehler's comment: "Lichens are small ecosystems…".[33] Such

[31] Though biology textbooks tend to treat lichens as all classic cases of obligatory mutualism, Nash points out that the symbiotic relationships in the varied lichens are themselves varied and only some of these are cases of obligatory mutualism. (Nash 1996, 2-3) Adjamdjian (1993) has gone further and argued that all lichens are better understood as cases of controlled parasitism.

[32] Another classic example are the protozoan-termite relations The termite, such as the Formosan subterranean termite, lives on cellulose, but the termite is incapable of efficiently digesting cellulose due to its incapacity to synthesize cellulases (enzymes required for the hydrolysis of cellulose to glucose). The cellulases are provided by a protozoan, Trichonympha, inhabiting the termite's gut. Trichonympha, in turn, is dependent on the steady digestion of cellulose by the termite for its own diet of cellulose derived carbohydrates. (Cleveland 1923 & Prescott et al. 2002, 599) Some species of termite, such as *Coptotermes formosanus,* use a "double digestion" system, where some cellulase is endogenous and some supplied by protozoans in the hindgut. (Nakashima et. al. 2002 & Zhu 2005 et. al.)

obligate mutualism, as seen with lichen, provides a model of a stable minimal ecosystem. Further, given what is required for an autocatalytic collective, such obligate mutualism is a case of an autocatalytic collective: in the case of a lichen, the mycobiont by enhancing water uptake due to its low water potential and reducing light intensity allows the photobiont to exist in habitats that would otherwise be inhabitable, and the photobiont by providing carbon nutrition (and, in the case of cynolichens, a nitrogen source) allows the mycobiont to exist in habitats where it would be out-competed by other fungi. (Nash 1996, 2-3)

Extending the model of an autocatalytic collective to cover stable ecosystems does not imply that we must consider whole ecosystems as living systems. Ecosystems lack an important feature of life, namely replication. This fact should not be viewed as some sort of problem for extending the model of the autocatalytic collective to the ecosystem, however. An individual organism forms an autocatalytic collective through its continued persistence in the linking of exergonic and endergonic processes to extract thermodynamic work. The activity of organismic replication forms another higher level autocatalytic collective. That the ecosystem does not replicate just entails that there is no further higher level autocatalytic collective of which it is a part. That the hierarchy of autocatalytic collectives tops out at some point does not present some sort of theoretical problem with the model; rather, it is clearly preferable to some infinite upward chain of hierarchies.

---

[33] Farrar (1976) makes the same point that a lichen is best understood not as an organism but as an ecosystem.

c. *Biological functions rooted in autocatalysis*

This is where things stand. Biological agents are just physical-chemical systems but are physical-chemical systems that exhibit the distinctive, though perhaps not unique, features of replication and persistence through the repeated extraction of thermodynamic work. The model of an autocatalytic collective linking endergonic and exergonic processes is just such a system. The fundamental structure of metabolism of every living thing looks to be a straightforward example of such a model. Further, the model applied to whole organisms (free-living individual cells or multicellular organisms) makes sense of their capacity to sustain and replicate themselves. So, it should not be unreasonable to adopt Kauffman's suggestion that living systems are just what he calls "autonomous agents". If so, then we have identified the gross system type of which every living system is an instance and a gross systematic effect, namely autocatalysis through the repeated extraction of thermodynamic work, of such a system.

From the previous chapter, the utility of functional analysis and ascription rests in the characterization of individually less-talented units' contribution to some overall systematic effect. With a system type and systematic effect type of every living system in hand, we have a source of functional analysis in the biological case, namely explaining how living systems accomplish autocatalysis through thermodynamic work.

What then are the functional ascriptions the literal truth of which is up for dispute? We can divide these into the two gross functional types of persistence and replication reflecting the fact that an organism is both an autocatalytic collective and comprises an autocatalytic collective, respectively. In the first gross type, we find all those familiar functions dealing with energy acquisition, processing, and internal and external stabilizing functions. The latter includes everything from immunology to the structural integrity of cell structure on the principle that the autocatalytic collective will fail unless conditions appropriate to the thermodynamically improbable environment required to complete work cycles are maintained. Replicative functions include the contributory effects of anatomical structure and behavior (e.g., mating patterns) to replication and ontogenic developmental. The functional explanation of development fits within replicative functions, because catalytic closure is not achieved until one mature organism produces a mature adult organism in a position to engage in the next replicative cycle. All the steps along the way, namely those of development, fall within the catalytic cycle from one adult to the subsequent adult offspring.

d. *The role of etiology given the autocatalytic model*

The intent has been to clarify the class of functional ascriptions up for dispute in order to be in the position to ask whether they are appropriately understood as literal or metaphoric. The above proposal, by grounding a source of functional ascription and explanation in autocatalysis, is not wholly neutral, however. In particular, it does violence to a traditional motivation for the etiological theory of function.

The aim of functional analysis, I have said, is to decompose some systematic effect into individually less-talented contributing units. However, if there is a difference between proper[34] systematic effect and merely incidental effects, not any individuation of systematic effects (nor of systems themselves, for that matter) will do. For example,

[34] I am following Millikan in the usage of the term ‘proper’:

> I intended (as suggested at LTOBC [Millikan 1984] 2) Webster’s first meaning of ‘proper’, which coincides with that of Latin *proprius* meaning *one’s own.* (2002, 116)

contraction by the heart would seem to be a proper systematic effect of the heart, whereas its production of sound a merely incidental effect. Cummins, for example, has been criticized by Millikan (1999a & 2002) for failing to see that, before we can get to explaining how some system does its work, we need to know what the work is which is in need of explanation. We can explain the production of sound by the heart and do so by decompositional analysis. But, from a biological perspective, sound production is not what is in need of explanation; rather, it is the contraction of the heart that is need of biological explanation. Etiology or selectionist history is offered in part as providing a way to identify and individuate those proper systematic effects from the merely incidental. Since it is the contraction of the heart, the etiologist might suggest, that explains the persistence of hearts within a biological lineage, it is that effect which is in need of explanation from the biological perspective. The production of sound by the heart is not part of the explanation of the persistence of hearts in some biological lineage and is, from the biological perspective, an accident or coincidental effect of the system. That etiology might so provide a guide to the identification of proper systematic effects is utilized as a motivation to take some particular etiological account of function as superior to non-etiological alternatives such as Cummins (1975.).[35]

The autocatalytic proposal above would seem to undermine this motivation for an etiological account of function, however. We can identity a gross systematic effect that would lead to the functional ascriptions we find in biological textbooks. Importantly, that systematic effect provides a way to distinguish proper versus merely incidental effects of the system. Autocatalysis through thermodynamic work cycles is a distinct, if not unique, effect of biological systems. It is that effect which is in need of explanation from a

[35] See, for example, Wright (1973) or Millikan (1999a & 2002).

biological perspective insofar as it carves off the biological from much of the rest of the universe. Further, in the more detailed case (such as that heart), the gross systematic effect of autocatalysis provides the measure of whether, say, the production of sound is the proper systematic effect of the heart. (It is not, because the production of sound by the heart does not have any apparent contributory effect to the autocatalytic collective that is the organism in which it is housed.) There is, then, just no need to mention or rely on selectionist histories or etiology either to identify proper systematic effect or for subsequent decomposition.[36]

However, the committed etiological theorist might see the fact that the above proposal has so radically downplayed, if not displaced, the importance of selectionist history as a sign that it has gone fundamentally wrong.[37] Yet, we should remember that we need stable living systems to provide something over which natural selection can operate, and the above proposal attempts to provide just that. Natural selection has a role to play in explaining the long term perseverance of life on this planet, its diversity, as well as the stability of traits within lineages, but the properties distinctive of living systems that sets biology apart from physics and chemistry is not to be found by looking to natural selection. As Bechtel writes,

> Evolution via natural selection is a process that over time can develop systems with greater autonomy. Although not denying the traditional accounts of evolution (e.g., that evolution requires mechanisms of variation and selective retention), the focus on autonomous systems provides a rather different perspective. First, it places the organism in the central role and emphasizes that an

[36] Millikan (2002) recognizes that, once the systematic effect has been identified, selectionist history is not required in the functional decomposition of that effect. It is for this reason that she sees her own proposal (originally developed in detail in Millikan (1984)) and that of Cummins (1975) as complementary, not competitive.

[37] I will momentarily suggest that evolutionary considerations do provide a further source of functional ascription and explanation, though I doubt the committed etiological theorist will find consolation in that proposal.

> organism needs to be able to maintain itself as an autonomous system. Otherwise, there is nothing to evolve. This does not mean that individual organisms must be totally self-sufficient. Organisms can evolve to rely on features of the environment that are regularly present to them. But they need to create and maintain all the mechanisms upon which they rely in order to use these resources. Second, each addition to the basic system involves a cost in that the system must generate and repair these mechanisms itself. Evolution is not just a matter of introducing and selecting new genes, but requires a system that builds and maintains new traits (i.e., new mechanisms). (Bechtel, in press, 22)[38]

That said, I do not see the above autocatalytic proposal as having eliminated the etiological account of function from the running. What it has done is eliminate the claim that the etiological account has some proprietary or special insight into the nature of biological function that other accounts fail to have. The etiological proposal remains, as it should remain, a proposal for when attributed functions can be understood as literally true. (In chapter 6, I will suggest that the etiological view has got it right in respect to phylogenic functions.)

II. Controlled heritable variance as a biological systematic effect

I want to suggest a further source of functional ascription that reflects a secondary systematic effect of the living systems encountered in biology textbooks. That secondary systematic effect is the effect of controlled heritable variance.

So far, the specifications for Kauffman's autonomous agent do not include anything like heritable variation. Without heritable variation, the autonomous agent's lineage is awfully fragile and is unlikely to be durable over time, however. It is fragile, because the sequence of autocatalytic events will persist only so long as conditions remain stable. Think of the simple A-B autocatalytic collective above with the following

[38] Cummins (2002) makes similar remarks.

addition: photons striking molecular species A are required for its ligation of the two B fragments. (Now, the example is one that does some thermodynamic work.) Our little collective will keep running as long as B fragments are in supply or at least as long we don't turn out the lights. If the required conditions for the operation of the collective evaporate, so too does our collective.

Each extant organism, however, is the result of an uninterrupted chain of autocatalytic events extending back over at least 3.5 to 3.8 billion years[39] of often catastrophic changes in terrestrial conditions.[40] That chain of autocatalytic events has persisted through a series of transformations that have allowed descendent collectives to operate in novel conditions. The last 3.5 to 3.8 billion years have been witness to an explosion in the biosphere both in respect to its diversity and shear range. (In respect to the latter, microbial life alone appears to have entirely permeated the terrestrial surface, oceans, and atmosphere.) Kaufmann's autonomous agent requires the further property of heritable variability to accomplish this further trick of biological systems, namely persisting in novel conditions, if it is to serve as the model of what is found in biology textbooks.

"Heritable variability" encodes the two distinct notions of genetic[41] memory and genetic novelty. Genetic memory is the retention of some ancestral traits in descendents as a consequence of the replicative duties of genetic operators. Genetic novelty, in

---

[39] These dates reflect the oldest fossilized remains of prokaryotic cells. (Prescott et. al. 2002, 422)

[40] The six mass extinction events (Precambrian, Cambrian, Ordovician, Devonian, Permian, and Cretaceous) over that period alone are clear evidence of the dramatic and catastrophic shifts in terrestrial conditions. (See Stanley 1986 & 1987)

[41] I intend 'genetic' somewhat loosely including not only obvious genomic structures, such as DNA and RNA, but also "epigenetic" factors, such as chromatin structure, or other information bearing structures affecting heredity. (See Gibbs (2003) and Jablanka & Lamb (1989 & 1998) for some of the work on epigenetic factors.)

contrast, is the generation of novel variations on ancestral traits in descendents by genetic operators. The evolvability of a system – the capacity of a system to create adaptation – reflects the extent of genetic memory and genetic novelty within that system. As Bedau and Packard explain,

> A system's evolvability depends on its ability to produce adaptive phenotypic variation, and this hinges on both the extent to which the system's phenotype space contains adaptive variation and the ability of the evolutionary search to locate it while avoiding maladaptive traps. Two main factors control the effectiveness of the evolutionary search process: the way in which genetic operators transverse genotype space, and the way in which that genotypes are phenotypically expressed (the genotype-phenotype mapping). For evolutionary search to explore a suitable variety of viable evolutionary pathways, genetic operators must generate sufficient amounts of the right kind of genetic *novelty*. At the same time, since evolutionary adaptations are built through successive improvements, genetic *memory* is required for the evolutionary process to retain incremental improvements … so, evolvability requires genetic operators to balance these competing demands [the demands of genetic memory and novelty] successfully. (Bedau and Packard 2003, 144)

That evolvability requires the balancing of the competing demands for genetic novelty and memory leads Kimura to suggest:

> These considerations inevitably suggest that there must be an optimal mutation rate[42] for the survival of a species under a given rate of environmental change. If the mutation rate is too high the species will be crushed under a heavy mutational load; if it is too low the species will not cope with adverse environmental changes. (Kimura 1960, 21)[43]

The evolvability of a system is importantly, given Kimura's comments, a systematic property that can be more or less optimal given the balance of genetic novelty and genetic

[42] The mutation rate of a system simultaneously reflects a system's genetic memory and genetic novelty. (Bedau and Packard 2003)

[43] The idea that there is an optimal positive mutation rate conflicts with Williams' (1966) view that mutations rates will be as low as physically possible. Williams' view relied on the claim that mutations are generally harmful. If conditions were frozen and given that mutations are generally harmful, then it would make sense that we would see a downward move in mutation rate. And, this is captured in Kimura's sense of optimality, because in such a case genetic memory would clearly be the favorite over novelty. But, if conditions change over time, excessive genetic memory is unlikely to lead to the long term persistence of the lineage and it is the overly long term persistence of terrestrial life that is in need of explanation.

memory in respect to a rate of environmental change.[44] Focusing on a given rate of environmental change, as Kimura does, generates a diachronic measure of a system's evolvability. But, in order to include speciation events that do not involve the eradication/replacement of the ancestral species nor geographic dislocation or isolation (i.e., in order to include sympatric speciation)[45], we could also provide a synchronic measure of a system's evolvability by looking not to the rate of environmental change but to the structural features of an environment not presently utilized by a species. Such structural features provide novel operating conditions without presupposing a shift in the *in situ* operating conditions of the species. The capacity of a species to search these novel present conditions will also be reflected in its capacity to balance the demands of genetic memory and novelty; hence, we can provide a synchronic measure of a system's evolvability in addition to the diachronic measurement suggested by Kimura. We might take the notion of evolvability even further by creating a measure of supra-optimality as that maximal point for lineage sustenance balancing both the synchronic and diachronic measures of optimality.

However, let's limit the discussion of systematic evolvability and optimal mutation rates to the strictly diachronic sort envisioned by Kimura, because most of the theoretical work has been limited to that. Both in respect to infinite populations with

---

[44] As a property of the system given the structural balance of genetic memory and novelty as well as the rate of environmental change, evolvability does not depend, as Bedau and Packard (2003, 155) stress, on the presence of natural selection.

[45] Sympatric speciation is the divergence of species within a non-geographically isolated population of breeding individuals. Unlike sympatric speciation, allopatric, peripatric, and parapatric speciation all involve some degree of geographic or niche isolation. The existence of sympatric speciation does remain contentious. Various mechanisms have been suggested, however, that could give rise to sympatric speciation events, see Burger & Schnieder (2006) and van Doorn et. al. (1998) for the relevance of assortive sexual selection, Parker and Partridge (1998) for the relevance of sexual conflict, and Kawecki (1997) for the development of habitat specialization within populations. Verzijden et.al. (2005) and Skarstein et.al. (2005) have applied sympatric speciation models to charr, and Barluenga et.al. (2006) has done similar work in respect to cichlids.

fixed fitness functions and, more recently, in respect to finite populations with dynamic endogenous fitness functions, mathematical and computational models appear to show both that systematic evolvability evolves over time and that it does so in the direction of optimality.[46] That evolvability should evolve in the direction of optimality should not be all that surprising if evolvability evolves at all. If evolvability evolves, evolvability, as a systematic property of a system, is a heritable systematic property. Since evolvability is the measure of a system's capacity to create adaptations, it will be through that heritable capacity that a lineage will or will not persist over the long run. So, we should expect that lineages persisting over the long run have a mutation rate closer to optimality than those that do not. Importantly, it should neither be all that surprising that evolvability evolves. All that is required for that to be the case is that the mutation rate be encoded/be the result of some heritable structural feature itself subject to variation.

I set out this section with the claim that a systematic effect of biological systems is *controlled* heritable variation. We should now be in a position to assess that claim. The contrasting claim is that heritable variation is not controlled by biological systems. What would it mean to say that heritable variation is uncontrolled? Well, it would seem to mean something like, say, the mutation rate (a simple example of evolvability) is random. 'Random' might be read in two different ways in respect to mutation rate: 1) no fixed mutation rate – that is, any level of mutation in each reproductive event is equally probable; or, 2) an arbitrary fixed mutation rate – that is, there is some fixed value but which value it is is random. To say then that heritable variation is controlled by a system

[46] In respect to the work on infinite populations with fixed fitness functions, see, for example, Gillespie (1991), Holsinger and Fledman (1991), and Liberman and Feldman (1986). In respect to the work on finite populations with dynamic endogenous fitness functions, see Bedau and Packard (2003).

is to say that the particular value of the mutation rate is a non-random effect of that particular system.

Of the three above possibilities, only the last (controlled heritable variation) is plausible if there is an optimal mutation rate for a lineage. Possessing heritable and variable structural features controlling mutation rate is far more likely to lead to long term persistence of a lineage than its absence, because it allows the lineage in effect to discover and maintain the optimal mutation rate. If the mutation rate is random in the first sense, then it is very likely over the long term that the lineage will be mutating at suboptimal rates. If the mutation rate is random in the second sense, then it is possible that the lineage happens to be mutating at the optimal rate at some given point in time. The optimal rate, however, will change over time as ecological conditions change. Adopting a fixed strategy (random in the second sense) is not likely to work out over the long run, because over the long run the optimal rate will itself be subject to change. These quick conceptual points are born out in the mathematical and computational work cited above in respect to the evolution of evolvability. In addition to these abstract theoretical demonstrations for controlled heritable variance, a number of features of living organisms appear to provide empirical examples of just such control.

One such potential example would be the repair and proofreading mechanisms involved in DNA replication. During DNA replication, errors can occur when an incorrect base is added to one of the original chains of nucleotide bases. DNA polymerases in effect proofread the adding of bases for mismatched base pairs as well as replace mismatched bases with the appropriate match. It is the proofreading and repair activity of DNA polymerases that accounts for the relatively low error rate in replication.

The effectiveness of such proofreading and repair affects the mutation rate of an organism. Since the effectiveness of proofreading and repair is itself a phenotypic trait subject to evolution, this would seem to be an example of controlled heritable variance as a systematic effect.

An additional somewhat striking example of controlled heritable variance is the "SOS response" in bacteria in which bacteria utilize mutation as a self-defense mechanism. When under stress from, say, antibiotics like Cipro, bacteria such as E. Coli will switch on genes increasing the mutation rate during replication by 10,000 times. This hypermutation can consequently result in the rapid development of resistance to antibiotics. (Stix 2006). This seems to be a fairly straightforward example of controlled heritable variance, because the bacterium is capable of altering mutation rate in favor of genetic novelty in order to search the possibility space for a better variant of itself.

Less striking than the direct control over mutation rate with the "SOS response" is female mating preferences tied to phenotypic traits in males. Female mate selection is often tied to morphological phenotypic traits (e.g., the male peacock tail, the rooster's comb, or the size of male nailtailed wallaby) or phenotypic mating display behavior (e.g., the frill display of the frill-necked lizard, neck sac inflation by the greater prairie chicken, or the water dance of male alligator). Such phenotypic mating cues are typically costly, whether metabolically or more indirectly by increasing, say, conspicuousness to predators. Just thinking of metabolic costs, the general health and ontogenic fitness of a male can be reflected, say, in the quality of the peacock's tail. What is not reflected is the particular underlying factors behind that general health and ontogenic fitness, e.g., a variation in the thermo-regulatory system, in the food acquisition strategy, etc. As a result, some such female mating preference should have the result of targeting offspring in the direction of more successful variation and turning it away from less successful variation. This is a form of, I think, controlled heritable variation. The female sex preference provides a mechanism to enhance the rate at which genetic novelty is introduced into a population when a male has hit on a better strategy or heightening genetic memory when a male has hit on a worse novel strategy.

Given the long term persistence of life on this planet coupled with both the computational demonstrations and empirical examples of controlled heritable variance, it is not unreasonable to suppose that a systematic effect of which living systems are capable is controlled heritable variance. Such a systematic effect provides for a distinct and further source of functional ascription within biology.

Biological entities are and are part of autocatalytic collectives with the systematic effects of autocatalysis and controlled heritable variance. It is these two systematic effects of biological organisms that inform functional explanation and ascription in biology. With the particular functions of the biological case in hand, I will turn in the next chapter to the criteria that might decide whether such functions are genuinely teleological.

Chapter 4

LITERAL TELEOLOGICAL ASCRIPTION

In the two previous chapters, I suggested 1) that the utility of teleological ascription and analysis rests in the characterization of component contribution to overall systematic effect as well as 2) that teleological ascription in the biological case is to account for the systematic effects of autocatalysis and controlled heritable variance. Does a characterization of component contribution suffice for literal teleological ascription? Both the realist and the antirealist, though for different reasons, agree that it does not. For the antirealist, such teleological ascriptions can only be merely analogous to genuine function due to the absence of intellect. A nonartifactual realist, on the other hand, requires some naturalistic grounding process, such as selectionist history or homeostasis, for such ascriptions to be literally true. I agree with both the realist and the antirealist that a characterization of component contribution is insufficient for literal teleological ascription. However, I want to suggest that there is a univocal reason – one to which both the realist and antirealist should agree – to deny its sufficiency.

What is missing – that further condition for literal teleological ascription – is the following: the relevant item must be sensitive to the success and failure of its teleologically characterized activity. I will provide some conceptual auto-anthropological[47] evidence for this suggestion by noting that, across the realist/antirealist divide and across disciplines, a failure to be so sensitive is routinely cited as grounds to reject teleological ascription as literal. However, I want additionally to suggest the stronger claim that such sensitivity suffices for literal teleological ascription. As conceptual evidence for the stronger claim, I will show that the adoption of it as the criterion for literal teleological ascription explains the plausibility of both the realist and antirealist positions, and, as a result, both positions can be understood as attempts to provide the ground for such sensitivity. The disagreement between the realist and antirealist transforms then into a disagreement over what is required for something to be sensitive to its success and failure.

I. The sensitivity condition on literal teleological ascription

A little argument to set the stage. If the characterization of some effect as contributory suffices for literal teleological ascription, then every process or event is rightly teleological. If every process or event is teleological, then it is false that some process is not purposive. There is, then, just no (extensional) distinction to be drawn between purposive and nonpurposive activity. There is, however, an (extensional) distinction between purposive and nonpurposive activity. So, the characterization of some effect as contributory cannot suffice for literal teleological ascription.

To buy that little argument, we need reason to accept that first conditional premise as well as that there is an (extensional) distinction between purposive and nonpurposive

[47] I take the phrase "conceptual auto-anthropology" from Dennett (2005). The basic idea is that, in explaining and elucidating some concept, we are engaged in an anthropological exercise. (The anthropological exercise is *auto*, because it is in respect to our own concepts.) Explanation and elucidation of a concept does not proceed by way of a priori analysis but rather proceeds by the formation of an anthropological hypothesis.

activity. The latter is a common enough assumption, and so, if we can provide some further condition(s) on literal teleological ascription, I take it that the assumption is reasonable.

To the conditional premise. The individuation of any activity or process allows the individuation of the terminus of that process or activity. In any case in which the beginning properties (external and internal) to that process determine the terminus, we can characterize the contribution of each beginning property to the terminus. In short, any process or activity can be explained by citation of contributory effect. Every process or activity is at least "quasi-finalistic" or "teleomatic" to use the language of Waddington (1968, 55-6) and Mayr (1961 & 1974, 97-8), respectively. [48] So, we have the first premise: "if the characterization of some effect as contributory suffices for the truth of literal teleological ascription, then every process or event is rightly teleological."

Assuming an (extensional) distinction between purposive and nonpurposive activity, something more than the characterization of contributory effect is required for literal teleological ascription. That something more, I want to suggest, is the following: the relevant item must be sensitive to the success and failure of its teleologically characterized activity.

Let's start with what should be a straightforward enough claim: nothing about a physical behavior or event type as such marks it as genuinely teleological. Breathing, walking, stretching, door opening, moving a chess piece, uttering a sentence, etc. are on some occasions rightly teleological and others not. As Ryle stresses,

[48] Butts (1990) suggests that we should understand Kant (1790/2000) as having similarly stressed this point.

> … there is no particular overt or inner performance which could not have been accidentally or ‘mechanically’ executed by an idiot, a man in panic, absence of mind or delirium or even, sometimes by a parrot. (Ryle 1949, 45)

Since what makes some behavior or event genuinely teleological is not to be found by looking to the behavior or event itself, one must, Ryle suggests, look beyond the tokened event or behavior to determine whether that token is rightly teleological:

> When we describe someone as doing something by pure or blind habit, we mean that he does it automatically and without having a mind to what he is doing. He does not exercise care or vigilance, or criticism. After the toddling-age we walk on pavements without minding our own steps. But a mountaineer walking over ice-covered rocks in a high wind in the dark does not move his limbs by blind habit; he thinks about what he is doing, he is ready for emergencies; in short he walks with some degree of skill or judgment. If he makes a mistake, he is inclined not to repeat it, and if he finds a new trick effective he is inclined to continue to use it and to improve upon it. He is concomitantly walking and teaching himself to walk in conditions of this sort. *It is of the essence of merely habitual practices that one performance is a replica of its predecessors. It is the essence of intelligent practices that one performance is modified by its predecessors.* (Ryle 1949, 42, emphasis added)

Ryle has, I think, essentially got it right here.[49] What we are looking for, in looking beyond the performance of some tokened event or behavior, is whether that event or behavior was and is subject to modification in respect to the goal or teleos of that event or behavior. Mere replica performances are not rightly teleological, because the respective agent, in performing such behavior, is insensitive to the success or failure of the putative goal of that behavior. In contrast, the agent, who demonstrates some sensitivity to the success or failure of its behavior, is rightly construed as acting purposively and its

[49] Intelligent practice or behavior, for Ryle, is purposive behavior, and its contrast is the merely automatic. (See, for example, Ryle 1949, 20 & 25) There are no representational overtones in Ryle’s use of “intelligent” or “purposive”.

behavior rightly characterized as teleological. If Ryle is right here, then sensitivity to success and failure is necessary and sufficient for literal teleological ascription.[50]

Whether Ryle is right is a question of whether Ryle has rightly characterized the applicability of a concept. It is an anthropological question and is not a strictly philosophical one. It is not a question of whether we ought to apply the concept or category of the teleological, how is it possible for us to have the teleological concept, and so on. Rather, given that we do apply the concept, categorizing some of the world as teleological and some of the world as not, the question concerns the characterization of that practice. The evidence that I will offer in favor of Ryle's suggestion is, then, of an anthropological sort; I will offer a sampling of what is said about teleology as well as the positions taken in respect to its applicability and show that Ryle's suggestion make sense of both.

To see that sensitivity to success and failure is at least necessary for literal teleological ascription, we ought to look to those cases where authors deny that some event or behavior is rightly characterized as teleological. For example, though von Uexküll (a father of modern cognitive ethology) wants to maintain that animals act in accord with "nature's plan" (a plan teleologically characterized), he denies that any animal, besides ourselves, can act in a goal-directed fashion. They are, von Uexküll suggests, just insensitive to the success or failure of their behavior and, thereby, only engage in mere replica performances:

> Actions directed toward a goal do not occur in animals at all…. According to information I have received concerning sound perception of night moths, it makes no difference whether the sound to which animals are adjusted be the sound manifestation of a bat or one produced by a rubbing a glass stopper – the effect is

[50] James (1890/1950, 6-10) makes strikingly similar comments.

> always the same [namely, fleeing upon perceiving a high tone]…. (von Uexküll 1934/1957, 42)

Shettleworth provides a more recent example from cognitive ethology. In considering whether parids (e.g., chickadees and titmice) act purposively in storing food, Shettleworth writes:

> Food storing has been cited (Griffen 1984) as behavior suggesting that animals have conscious foresight. In storing food that will be used later, chickadees certainly look as if they are behaving in a consciously planful manner…. However, parids store food even under circumstances where they seem unlikely to anticipate retrieving it. For instance, in our laboratory, some birds persist indefinitely in storing peanuts in places where they drop out of reach. These observations … do make it clear that the cognitive and brain mechanisms … can be studied while remaining agnostic about the nature of animal's possible awareness. (Shettleworth 2002, 126-7)

Whatever we might want to make of "consciousness" and "awareness" here, it is clear that Shettleworth wants to deny that parids act purposively in food storing. They do not seem to have a goal "in mind" or guiding their behavior, because they are just insensitive to the success and failure of the activity characterized as food storing: "some birds persist indefinitely in storing peanuts in places where they drop out of reach".

Or, in the psychological tradition, consider Wooldridge's famous sphex wasp:[51]

> When the time comes for egg laying the wasp *Sphex* builds a burrow for the purpose and seeks out a cricket which she stings in such a way as to paralyze but not kill it. She drags the cricket into her burrow, lays her eggs alongside, closes the burrow, then flies away, never to return. In due course, the eggs hatch and the wasp grubs feed off the paralyzed cricket, which has not decayed, having been kept in the wasp equivalent of deep freeze. To the human mind, such an elaborately organized and seemingly purposeful routine conveys a convincing flavor of logic and thoughtfulness – until more details are examined. For example, the wasp's routine is to bring the paralyzed cricket to the burrow, leave it on the threshold, go inside to see that all is well, emerge, and then drag the cricket in. If, while the wasp is inside making her preliminary inspection the cricket is moved a few inches away, the wasp, on emerging from the burrow, will bring the cricket

[51] It is famous at least within philosophy of mind given Dennett's (1978) extensive use of it, where Dennett repeatedly uses this case to distinguish the merely tropistic or automatic from the purposive.

> back to the threshold, but not inside, and will then repeat the preparatory procedure of entering the burrow to see that everything is all right. If again, the cricket is removed a few inches while the wasp is inside, once again the wasp will move the cricket up to the threshold and re-enter the burrow for a final check. The wasp never thinks of pulling the cricket straight in. On one occasion, this procedure was repeated forty times, always with the same result. (Wooldridge 1963, 82)

The failure of the wasp to be sensitive to its own success (it has already checked that all is well in its burrow) suffices, Wooldridge suggests, to deny that its behavior is rightly purposive.

Or, consider a more classic example in the psychological tradition from James:

> If some iron filings be sprinkled on a table and a magnet brought near them, they will fly through the air for a certain distance and stick to its surface. A savage seeing the phenomenon explains it as the result of an attraction or love between the magnet and the filings. But let a card cover the poles of the magnet, and the filings will press forever against its surface without its ever occurring to them to pass around its sides and thus come into direct contact with the object of their love.
>
> …. Loves and desires are to-day no longer imputed to particles of iron or of air…. The end, on the contrary, is deemed a passive result, pushed into being *a tergo*, having had so to speak, no voice in its own production. Alter the preexisting conditions, and with inorganic matter you bring forth each time a different apparent end. (James 1890/1950, 6-8)

The iron filings do not act purposively, because they are insensitive to the success or failure of the putative goal, namely to reach their love, the magnet. Alter the conditions, and the filings show no resolve to alter their behavior to reach that putative goal.

Here is a recent example from the philosophical tradition:

> The botanists from which I take this example succumb to this temptation by explaining the plant's behavior in purposive terms: the plant changes color, they say, "in order to" attract pollinators. Clearly, though, we do not have anything like action, nothing like purpose, nothing that would justify "in order to". The plants would change color whether or not their behavior succeeded in attracting pollinators. (Dretske 1999, 26)

That plants are insensitive to their success and failure in respect to attracting pollinators, Dretske thinks, suffices to deny their behavior is rightly teleological; such plants only engage in mere replica performances.

Lastly, an example in the philosophical tradition with an Aristotelian flavor, Sorabji writes,

> … a necessary condition for the correct application of the word 'function'[52] [is] that some efforts are or would if necessary be made to obtain the effect …. For example, suppose that by some convenience herbs in medieval times had sprung up spontaneously in places where there were crowded assemblies, though not elsewhere. Suppose they had this same effect of keeping off the smell. But suppose no efforts were made and no efforts would have been made if necessary to obtain this effect from herbs. I do not think that under these circumstances it would be true to say that herbs had the *function* of keeping off the smell of crowded assemblies. And I think it is the absence of any effort and of any likelihood of effort to obtain this effect from herbs that would prevent it from being true. (Sorabji 1964, 290)

If the citizens make no or would make no efforts to secure the effect of keeping off the smell, then this is a reason, Sorabji suggests, to deny that the herbs have that effect as a function; alternatively, that the citizens are insensitive to the success or failure of the herbs keeping off the smell provides cause to deny function. This example is doubly interesting, because it is not an example of acting purposively, as were the others, but is an example of the possession of purpose. The herbs, even if they had the function of keeping off the smell, would not be in so doing acting purposively. The herbs are not sensitive to the success or failure of keeping off the smell, even if the citizens are. It is the citizens that could be sensitive to that effect and could take, when appropriate, actions to ensure the success of that effect. And, if they were to do so, then plausibly the herbs would possess the function of keeping off the smell. This distinction between acting purposively and possessing a function does not despite appearances introduce anything

[52] Sorabji is here explicitly considering teleo-functions only.

new into the equation. In the previous examples, there was already a distinction between acting purposively and the possession of function. Individual acts can possess function but cannot themselves act purposively; the latter is reserved for the acting agent. The individual acts possess function in virtue of the agent being sensitive to the success or failure of those acts. The herbs, when they possess function, do so in the same way that an individual act does so, namely the relevant actors hold them to a standard of success and failure in respect to that function.

In each case above, the reason to deny that the behavior is rightly teleological reflects the insensitivity of the relevant entity to the success or failure of its own behavior (or some further effect in the case of the herbs). Absent such sensitivity, such behavior is blind habit, mere reflex, merely automatic, etc., because each occasion of such behavior, as Ryle nicely puts it, is just a mere replica performance. The behavior never alters, is incapable of altering, in respect to the putative goal, or alternatively the behavior is not subject to modification in virtue of the putative goal. Thus, the relevant entity just looks to be insensitive to the success or failure of its behavior, and consequently that behavior is not rightly characterized as purposive.

Now, Ryle suggests that such sensitivity is not merely a necessary condition but that it suffices for literal teleological ascription. Ryle's mountaineer, even absent Ryle's behavioristic overtones, appears to be acting purposively, because he alters his behavior over time in respect to the goal of getting across the mountain. His successes or failures are consequential in respect to the walking behavior he exhibits, and the behavior he exhibits would be explained in part by reference to his sensitivity to the success or failure of getting across the mountain. Consider von Uexküll's night moths. If this or that night moth, after having been exposed to some number of squeaky stoppers without ill effect, no longer flees from stoppers but continues to flee bats, then it would seem that night moths are sensitive to the success and failure of their predator avoidance behavior. The success or failure of the relevant behavior would be consequential to the night moth's activity in just the same way that mountaineer's successes and failures are consequential to his behavior. We could, if we follow Ryle here, think of the night moths as acting purposively, because the explanation of their adapting their behavior makes reference to the putative goal of that behavior.

Authors other than Ryle also cite such sensitivity as sufficient for literal teleological ascription. For example, from cognitive ethology, Byers (2002) and Brittan (1999) both suggest that pronghorn does' varied behavioral responses to changing and varied predation threats to fawns suffices for "conscious planning" (Byers) or to "assure that the behavior is purposive" (Brittan, 61). (Both Byers and Brittan also explicitly reject behavior as purposive when it lacks such sensitivity – Byers' example is the contrastingly insensitive predator response of killdeer, and Brittan's is the automatic behavior of Wooldridge's sphex wasp.) Additionally, the sufficiency claim is explicitly endorsed within the regulative and homeostatic views of teleology. For example, Rosenblueth et. al. (1943) suggest that feedback controlled behavior not only suffices for teleology but is synonymous with it. The non-teleological is just that behavior for which no feedback maintains the teleologically characterized effect in the face of internal and external disturbances.[53] For these regulative or homeostatic views, behavior is purposive when the

[53] See Sommerhoff (1950 & 1959), Braithwaite (1953), Beckner (1959), Hempel (1959), Nagel (1961, 1977a, & 1977b), and Boorse (1976 & 2002) for similar views.

relevant entity is capable of compensating for disturbance so as to realize the relevant goal – that is, when it is capable of adjusting so as to ensure behavioral success.

That said, both the intellectual antirealist and the etiological realist seem to reject that such sensitivity suffices for literal teleological ascription. Without the presence of a representation, the intellectual antirealist is just unwilling to concede that a teleological ascription is literal.[54] And, without the presence of a supporting etiology, the etiological realist is unwilling to treat teleological ascriptions as literal. I want to suggest, however, that, despite appearances, both positions do in fact implicitly endorse the sufficiency claim, because the plausibility of both views depends on the sufficiency of the sensitivity criterion. The thought is that, if sensitivity to success and failure explains the plausibility of both positions, we have some further reason to think that such sensitivity is what rests at the core of the purposive/nonpurposive distinction.

II. The intellectualist's (implicit) endorsement of the sensitivity condition

Let's begin with the intellectualist picture of the teleological, where literal teleological ascription requires that the respective entity represent the goal or purpose of an act or behavior. Even on the intellectualist picture, it is not sufficient, however, for a behavior to be purposive that the relevant entity represents some goal in respect to that behavior. As Dretske suggests,

> Purposeful action – in contrast to mere behavior – requires thought, but thought alone is not enough. To qualify as purposeful, thought must *control* behavior. To be an agent it is not enough to be a thinker and a doer. The thinking must explain the doing. (1999, 19)

[54] For example, though Kant would agree with the homeostatic and regulative views that the self-preserving structure of biological organization is what makes teleological explanation particular apt in biology, he rejects that such teleological explanation is literal absent representation.

We might say, then, that it suffices for an act to be purposive on the intellectualist picture if the agent represents some goal in respect to that behavior and if that representation somehow controls that behavior.

Despite the widespread appeal of such a picture, it is not at all obvious why the intellectual or representational condition needs to be in place for a bit of behavior to be purposive. Most cases of skill-based activity by intentional and psychological agents lack any obvious or explicit representations of goals but are, nonetheless, clearly purposive. For example, when I am skiing down a hill, I will perform any number of actions to make my way down the hill: I will avoid trees as they present themselves, alter pressure on the downhill ski in response to changing incline, or alter my position over the skis in response to micro-changes in snow conditions. The heterogeneous behavioral alterations that take place in getting down a hill are nowhere captured in consciousness or in mind. In fact, the attempt to consider such things consciously tends to lead to disaster as events overtake the speed of explicit thought. And yet, this heterogeneous set of behavioral events in getting down the hill are not mere habit nor blind reflex. The present tokened responses, as skilled responses, vary appropriately to the changing conditions in light of the overall goal of getting down the hill. That I lack intentions for each of these is further seen in the general difficulty in voicing or instructing another in the appropriate skill set required to get down the hill. The same type of phenomenon is seen in driving a car, riding a bicycle, or any other skill requiring numerous behavioral modifications to acquire the overall goal given varying circumstances. While I might be able to attend to some behavioral modification required in respect to the goal, it seems mistaken to think that only when I do so has the behavior been rendered purposive.[55] Or, consider a case like the speaking of a sentence. Any competent English speaker will use 'the' on some occasions and 'a' or 'an' on other occasions. The linguistic behavior is subject to modification given the circumstances and is subject to modification in cases of failure. And yet few competent English speakers, as attested by interviewing any number of freshmen, could tell you what the purpose of using the definite article is. Even among those that can, a bit of introspective research should reveal that there is rarely anything like an explicit intention preceding and resulting in yours and their use of the definite article. Such skill-based behavior in the absent of explicit intention is treated as purposive, I suspect, just because the relevant entity is sensitive to its successes/failures in performing such behavior. But, the point to note for now is just that, given our willingness to treat such behavior as purposive in the absence of explicit intention or representations of a goal, it is not prima facie clear why the intellectualist condition needs to be in place for literal teleological ascription.

The intellectualist has a well-worn theoretical move to explain why behavior is purposive in the absence of an explicit representation of the goal: though no explicit or conscious intention is before mind, there is some subterranean (subpersonal, unconscious, implicit, tacit, etc.) intention or representation of the relevant goal in virtue of which such behavior has been, is, or comes to be modified. Let's notice, however, that the intellectualist is now engaged in a bit of theorizing. These subterranean representations are put forth to explain how it is that recognizably purposive behavior could be such in the absence of the explicit representation of a goal. Uncharitably, the subterranean representation, however, is just a way to salvage the intellectualist condition in the face of

[55] See Millikan (2004 chapter 1) for a similar discussion.

clear counterexamples. But, if we adopt the suggestion that sensitivity to the success or failure of the act is required for literal teleological ascription, we can explain the plausibility of the intellectualist condition and the plausibility of postulating subterranean representation in the explanation of purposive skill-based activity.

An intentional and psychological agent, who is capable of representing some goal as well as the conditions in which it finds itself and is capable of practical reasoning, looks to be just the sort of the thing that could be sensitive to the success and failure of its acts. The intentional and psychological agent looks to be then a plausible answer to the question "how is it possible that something could be sensitive to the success and failures of its acts?". That it might be *an* answer to that question does not do justice to the intellectualist view, because that view involves the stronger claim that only through intellect might some agent possess the appropriate sensitivity. It is because of this stronger claim that the intellectualist rejects that nonartifactual teleological ascriptions can be anything more than metaphor. Further, it is on the basis of the plausibility of the stronger claim that the postulation of subterranean representations is not merely an ad hoc theory salvaging move but is, rather, what must be in place if we were right in the first place that such skill-based behavior is rightly purposive. So, if sensitivity to success and failure is at the core the purposive-nonpurposive distinction, then that criterion should help explain the plausibility of the stronger claim that purposiveness is only possible through intentional agency.

An explanation of the intellectualist's strong claim (such sensitivity is only possible via representation), I think, lies in the Cartesian mechanistic view that dominates the natural sciences at least from the 17^{th} through the 19^{th} centuries. That view that the world can be wholly explained in terms of movements and forces suggests a deterministic world incapable of the sensitivity required for purposive activity. If the world is nothing more than billiard balls and the forces that act on them, the mechanisms in the world, once initial and boundary conditions are set, just run their course. There is no space for the behavioral plasticity required for the rightly purposive. It is not surprising then that Descartes looking to the biological saw nothing there but automata. That apparently rigid deterministic character of mechanism – that world subject to Newtonian calculus (set boundary and initial conditions and calculate the result) – explains why it is so routinely assumed that "the mechanistic displaces the purposive."[56] To offer a mechanistic explanation of some phenomenon is to expose it as rigid and deterministic and, therefore, lacking the plasticity required of the purposive. If the world is to be exhaustively explained by such deterministic mechanisms, then there is no place in the physical world for the purposive. In fact, we can see Bergson's (1907) *élan vital* and Driesch's (1909) *entelechie* as conceding just this point: to make room for purposiveness in biology, one must add some new metaphysical bit to the world on top of the mechanisms explained by physics and chemistry.

But, more plausible than Bergson or Driesch is Kant's position in the *Critique of Judgment*. As Nagel writes,

> Kant was heir to the thought of Descartes and Newton, and subscribed to the principle that all material processes of nature must be explained by "merely mechanical laws." However, the apparently purposive character of the organization and behavior of living things seemed incapable of being understood in terms of "the mere mechanical laws of motion"; and they had to be viewed *as if* they had been produced by design. (1977b, 289)[57]

[56] This is Dennett's phrase but not a principle that he endorses (1978, 234).

[57] Underhill (1904), Mayr (1974), Butts (1990), and Ariew (2002) all stress this point as well.

And, it is this position from Kant, recognizing the utility of teleological explanation as indispensable in biology but, because of the endorsement of Cartesian or Newtonian mechanics, treating teleological explanation as merely a heuristic, that continues through contemporary philosophy. (For repeated variations on this theme, see Broad (1925), Hospers (1958), MacIntyre (1957), or Malcolm (1968), Cummins (1975), Dretske (1999), or Davies (2002).[58]) Since the natural world as rigid deterministic mechanism cannot be literally teleological, the teleological is reserved for the intentional and psychological agent who, it seems, is clearly capable of the requisite plasticity. (This sets up an obvious tension between a mechanistic view of the natural world and mentality, and that tension will, in turn, generate the philosophical diversion of the mind-body problem.)

Dennett (1978) has argued that the intellectualist advocates of the principle "the mechanistic displaces the purposive" have, in their advocacy, repeatedly confused *explaining* with *explaining away*, where one assumes that to provide any non-question-begging mechanistic explanation is to *explain away* and not *explain* the purposiveness of

[58] With philosophers such as Hospers, MacIntyre, or Malcom, the view that the mechanistic, i.e. that subject to mechanistic explanation, cannot be rightly teleological was entangled with irrelevant moral concerns. While that I acted purposively is clearly relevant to decisions governing responsibility and such decisions have ethical import, it is not true that the conceptual grounds for judging an action purposive are essentially ethical. I can judge that you spoke deliberately and with purpose without invoking any ethical considerations. As Dennett nicely puts it, what we need to take on within the intellectualist picture to judge an action purposive is the intentional stance but the moral or ethical is a still further stance, namely the personal stance:

> …. One adopts the intentional stance toward any system one assumes to be (roughly) rational, where the complexities of its operation preclude maintaining the design stance effectively. The second choice, to adopt a truly moral stance toward the system (viewing it as a person), might often turn out to be psychologically irresistible given the first choice, but it is logically distinct. Consider in this context the hunter trying to stalk a tiger thinking what *he* would do if he were being hunted down. He has adopted the intentional stance toward the tiger, and perhaps very effectively, but though the psychological tug is surely there to disapprove of the hunting of any creature wily enough to deserve the intentional treatment, it would be hard to sustain a charge of either immorality or logical inconsistency against the hunter. We might, then, distinguish a fourth stance, above the intentional stance, called the personal stance. The personal stance presupposes the intentional … and seems, to a cursory view at least, to be just the annexation of moral commitment to the intentional…. (Dennett 1978, 240)

some behavior. To *explain away* is to explain some phenomenon but in so doing render the phenomenon merely apparent. So, for example, if we start wanting to explain why it is that celestial bodies move across the night sky and end up explaining that phenomenon by reference to the motion of the Earth, we have explained away the initial phenomenon as merely apparent; i.e., the celestial bodies do not really move across the night sky but only appear to do so given our position on the Earth's surface. In contrast, a Newtonian mechanistic explanation of the Earth's movement does not *explain away* that movement (i.e. show that the Earth does not really move) but explains that movement.

An advocate of the principle "the mechanistic displaces the purposive" need not be suffering from any such confusion, however. If mechanistic explanations expose some behavior as rigid or automatic, then mechanistic explanation does explain away purposiveness. Such explanations expose the relevant behavior to be lacking what is requisite of the purposive. The mistake in such advocacy lies not in thinking mechanistic explanation can explain away the purposiveness of some behavior (it can if it exposes the lack of some requisite feature) but lies, instead, in assuming that the mechanistic is identical to the automatic. (If the mechanistic was identical to the automatic, then the intellectualist would be quite correct that all mechanistic explanation displaces the purposive.)

The irony is that the assumption – the mechanistic is identical to the automatic – must be false if the intellectualist picture of the purposive is to be plausible. Remember that the capacity of an intentional and psychological agent to represent some purpose or goal in respect to some behavior or activity was insufficient for that behavior or activity to be purposive. My representing the goal of winning the lottery does not suffice for my winning the lottery to be purposive unless I cheat. I can purposively purchase a ticket, but I cannot purposively win the lottery by having that as a goal. What made the intellectualist position plausible was that the representation of the goal was causally implicated in the sense of controlling the relevant purposive behavior. My representing a goal, my capacity to represent the conditions of the world, and adjust my behavior so as to better realize that goal through those representations are what comprise the picture offered by the intellectualist. That story then explains how it is possible for an entity to be appropriately sensitive to the success and failure of its acts: it is so sensitive by its capacity to adjust its behavior in light of representing its goals and circumstances and move to act via practical reasoning. However, what makes the intellectualist picture a plausible explanation of how there could be purposive behavior is exactly that it provides a causal model of the process required to be appropriately sensitive to the success and failure of one's acts. The plausibility of the intellectualist picture rests, in short, on its provision of a mechanistic explanation for purposive behavior. Unless the intellectualist picture is to use Ryle's phrase "broken-backed" – that is, unless the intellectualist picture explains away the very purposiveness that it purports to explain, the intellectualist cannot subscribe to the principle "the mechanistic displaces the purposive". Thus, offering a mechanistic explanation, lest there be no purposive activity, cannot suffice to expose a bit of behavior as automatic and, thereby, nonpurposive. Without the principle "the mechanistic displaces the purposive", the reason to exclude the possibility of genuine nonartifactual teleology has fallen by the wayside.

An auxiliary motivation to exclude the nonartifactual from the genuinely teleological is waiting in the wings, however. The advocate for a nonartifactual teleological realism needs to supply some alternative explanation for how the appropriate sensitivity to success and failure is possible. But, sensitivity to success and failure is sensitivity to some norm or standard of performance in virtue of which some act is a success or failure. A norm or standard looks to be something inherently representational or intentional. Without an explanation of how there can be norms or standards that are not products of intentional and psychological agency, nonartifactual realism degenerates into a mysterious position, and the intellectual antirealist wins, it seems, by default. It is exactly because the nonartifactual realist cannot, without invoking mysteries, explain such normativity that Davies (2002) argues we must accept the antirealist position. I will show in the subsequent chapters, however, that the category of the normative extends beyond the intentional and psychological and that the possibility of such nonartifactual normativity can ground the possibility of genuine nonartifactual teleological functions.

The point in this section, however, was just to notice the following: adopting "sensitivity to success and failure" as the core of the distinction between the purposive and nonpurposive explains the motivation for and plausibility of the intellectualist picture; and, if it explains that picture, there is reason to think the conceptual auto-anthropological evidence for "sensitivity to success and failure" is correct. The intellectualist's representational requirement on literal teleological ascription looked initially to be poorly motivated given the commonplace practice of teleological ascription in the case of skill-based activities for which no representation is in any obvious way implicated. However, that requirement became more plausible if we understood it as an answer to "how could it be possible that anything is appropriately sensitive to its successes and failures?" Further, if we added in the long-standing assumption that the mechanistic is identical to the automatic, we could understand both why an intentional and psychological agent looked to be the only possible answer to the how-is-it possible question as well as why, if we are correct in ascribing purpose in the case of skill-based activity, there must be some sort of subterranean representation at play. Unhappily for the intellectualist, the assumption that the mechanistic is identical to the automatic is untenable. That said, holding the sensitivity condition at the core of the purposive/nonpurposive distinction both supplies and explains an auxiliary intellectualist motivation, namely that normativity is rightly restricted to the intentional and psychological. Normativity is an issue for literal teleological ascription exactly because the relevant entity must be sensitive to some standard in virtue of which its actions are successes and failures. Again, the sensitivity condition appears to be at the heart of things. It is not unreasonable to conclude that, despite initial appearances, the intellectual antirealist endorses the sufficiency of sensitivity to success and failure, since only by so doing does the intellectualist's position enjoy some measure of plausibility.

III. The etiological realist's (implicit) endorsement of the sensitivity condition

Let's turn to a commonplace realist view of nonartifactual teleology, namely the varied etiological views. (The homeostatic or regulative views explicitly endorse sensitivity to success and failure as sufficient to identify purposive behavior, so we need not bother with these.) Though a number of variants are present in the literature, the following captures, I think, at least the general spirit of etiological views: The function of a heritable trait T within a lineage is effect E, because the ancestral production of E by T tokens in part explains the proliferation of T within the lineage. A present token T possesses the function of E, even when incapable of E, because present tokens possess the relevant function in virtue of the selective efficacy of ancestral productions of E. If these tokens fail to E as their ancestors did, they have failed to perform their function; consequently, they are malfunctional and not afunctional (i.e. lacking in function).[59] However, like the intellectualist picture above, it is not all that clear what etiology is supposed to supply in order to render the relevant functions genuinely teleological. That is, what is it about having the sort of etiology proposed that makes plausible that the relevant function is more than merely analogous to the genuinely teleological?

It is frequently suggested, most prominently perhaps by Wright (1973), that etiology allows us to distinguish between the function of an item and what it merely does by accident. Using Wright's example (Ibid., 140), there are many things that hearts do besides the circulation of the blood, and, while the production of a thumping noise might contribute to diagnostics, the function of the heart is not to produce a thumping noise but is to circulate the blood. Some effects of an item, though they might be described as contributory, seem to be rightly accidental and are not then genuinely teleological. Etiology serves as the ground for distinguishing accidental and non-accidental contributions of some systematic component. Hearts have been maintained in mammals because of their effect of circulating the blood and not because of their contribution to diagnostic procedures by the production of sound. Or, borrowing another of Wright's examples (Ibid., 147), large belt buckles might contribute to the wearer's health by stopping a bullet, but the stopping of bullets is not what explains why it is that belt

[59] See, for example, Wright (1973), Millikan (1984, 1989b, 1993, 1999a, & 2002), Neander (1991a & 1991b), Griffiths (1993), and Godfrey-Smith (1994).

buckles have been and continue to be produced. So, some contributory effects are only contributory by accident and are not rightly included as the function of some item. Etiology, so the thought goes, provides a way to carve off those effects that contribute accidentally from those that do not and by so doing distinguish ersatz and genuine functions, respectively.

Even supposing that etiology provides *a* way to carve off accidental effects from non-accidental contributory effects, it cannot be by so carving up the accidental from the non-accidental that etiology grounds genuine teleological function. The distinction between accidental effects and non-accidental contributory effects need make no appeal to etiology. In the context of explaining the operation of some system type, the distinction between non-accidental and accidental effects is just the distinction between what is and is not explanatorily relevant, respectively. Roughly, one can distinguish two sorts of accidental effects: type-incidental effects and token contributory effects. In respect to type-incidental effects, any physical operation or process will have any number of auxiliary and downstream effects. The question in explaining a system type is which of the varied effects of some component item or process are relevant to the operation of that system. Those incidental effects, e.g., the thumping of the heart, are incidental, because they do not contribute to the operation of the system type. That an effect is incidental is in the first instance its failure to play a contributory role to the operation of the relevant system type. The distinction between incidental and non-incidental effects is not one that is in someway particular to etiology. In respect to token contributory effects, the explanation of a system token's operation might involve the contributions of random, unique, or token specific effects. Using Wright's example, a screw that happens to fall inside an engine might have contributory effects to the operation of that engine. And yet, it is an accident of sorts that that screw is having this contributory effect. It is an accident not because of some mysterious absence of etiology or explicit intention on the part of the engine designer but because the effect of the screw is irrelevant (just as incidental effects are) to the explanation of how engines of that type operate. Its contribution is an accident, because engines of that type do not utilize such screws in their operation. In this tokened case, the explanation of this particular engine would involve reference to that screw, but that is an explanation of a token and not a type. Again, the distinction between what is accident and what is not does not rest in the first instance on etiology but rather on what is and is not relevant to the explanation of a system type.

Since the distinction between accidental effects and non-accidental contributory effects is not in any way particular to etiology, we have two options: 1) we can continue to accept that the distinction between accidental effects and non-accidental contributory effects suffices to distinguish ersatz and genuine functions; or, 2) we can continue to accept that it is something peculiar to etiology that makes for genuine function but that something is distinct from the accidental/non-accidental distinction. The first option is clearly a nonstarter for the etiological realist. But, it looks problematic on independent grounds. If the accidental/non-accidental distinction suffices to distinguish genuine from ersatz function, then each explanatorily relevant characterization of contributory effect suffices to be genuinely teleological; this puts us back at the beginning with every process counting as teleological and, thereby, losing the (extensional) distinction between the purposive and the nonpurposive. The etiological realist needs to choose the second of these options and needs to provide some further reason why etiology makes for the genuinely teleological.

An alternative reason to think etiology is peculiarly appropriate for literal teleological ascription is that etiology can explain the presence of some component item or process within a system. Since frequently we answer a query for the purpose of an artifactual item by providing the reason for its presence, etiology, particularly selectionist history, would seem appropriate to distinguish the genuinely teleological from the merely analogous. However, as already pointed out in chapter 2, that which is rightly teleological, even in the artifactual case, does not require that it came into being for its *telos* nor does it suffice that something came into being for some *telos* for it to have that *telos.* So, that an etiological explanation might account for the presence of some item cannot be what makes such etiologies the plausible ground of genuine teleological function.

Let me suggest, then, that what makes etiology so attractive as the ground for distinguishing genuine and ersatz nonartifactual teleology is that such etiologies appear to provide evidence for exactly the property pointed to previously, namely a sensitivity to success and failure.

Consider the case of a stable phenotypic trait. For example, efficient digestion within non-ruminant and non-frugivorous mammalian herbivores requires some way to sift smaller easier to digest particles from larger more difficult to digest particles in order to excrete the larger particles and, by so doing, concentrate digestive work on what is easiest to digest. (Björnhag, 1989) Clauss has suggested that three-toed sloths (*Bradypus*) accomplish this by adopting an upright resting posture. (Clauss 2004) While the three-toed sloth, when active, moves through the trees hanging upside down, it spends very little time active and most of its time is spent resting and squatting in some tree fork. (Goffart 1971) The exorbitant time the sloth spends resting upright, Clauss (2004) suggests, allows the sloth to take advantage of gravity to sift larger less-dense particles from smaller denser particles.[60]

Following Clauss, the sleeping posture contributes to the end of efficient digestion by allowing the sloth to take advantage of the gravity gradient and, thereby, sift ingested particles by size. For the sake of simplicity, assume that sleeping posture is a monogenic heritable trait. Over a small number of generations, we should expect some number of mutant sloths to be born that do not adopt the upright sleeping posture. These mutants, upside-down sleepers, will suffer inefficient digestion, because they will not sift particles well; as a result, we should not expect that many will succeed in passing on their mutation to offspring. Consequently, the upright sleeping posture is a stable phenotypic trait in the genus *Bradypus* despite repeated mutation events over its history. The explanation of the stability of the upright sleeping posture is that of the various sleeping postures the sloth might take (and does take at time given mutations) only the upright sleeping posture serves the end of efficient digestion. The lineage "acts" in such a way to ensure that when it gets sleeping posture right in respect to digestion that strategy is repeated and that when it gets it wrong it does not repeat the mistake. The lineage as a whole looks to be sensitive to its successes and failures in respect to efficient digestion.[61]

---

[60] Though the three-toed sloth is strictly foliovorous, its cousin, the two-toed sloth (*Choloepus*), is frugivorous and is, as a result, less dependent on particle separation. Understandably, the two-toed sloth adopts multiple sleeping postures. (Merritt 1985, Chiarello 1998, & Clauss 2004)

[61] Notice that the focus in the above example is on the diachronic stability of a trait and not on its origination or initiation. As Millikan has repeatedly stressed, etiological explanation via selection is to explain the stability of function and not its origination. (See, for example, Millikan (1984 chapters 1 & 2)

The etiological proposal is plausible exactly because it provides a naturalistic mechanism for how systems (that is, whole lineages) might exhibit sensitivity to success and failure. However, if that proposal is acceptable, then the appropriate etiologies cannot merely present the appearance of such sensitivity but must be in fact a case of sensitivity to *success* and *failure*. This, however, requires that we make sense of how the history of a lineage is evidence for a norm of performance. And, it is this last where the etiological proposal seems to give out, because it is not clear how such histories provide evidence of norms of performance.

While it is certainly common enough to speak of adaptive "success" or "failure", it is not at all clear why we should think such language is literal. As stressed in the first chapter, norms and standards, at least in the intentional and psychological context, are not epiphenomenal. They play causal and regulative roles in the governance of intentional and psychological behavior. But, in these selectional cases, it is not clear what some norm of performance might do, if there were one. Some organisms act in ways that lead to the perpetuation of their lines and others do not, and we call the former "adaptive successes" and the latter "adaptive failures". That some organism is an adaptive success relative to another is just that it perpetuates its lineage and the other does not. Where is the regulative action of the putative norm of performance? There just seems to be no need for it, no need to postulate some further regulative process. Some things live and some die out; that's all there is to it.

---

for an extensive discussion of this point.) Critics, such as Nagel (1977b) and Cummins (2002), have missed this point. Both Nagel and Cummins, for example, criticize the etiological theorist as having misunderstood the nature of selection by trying to use selectional explanation to explain something it cannot, namely origination. But, this is not the aim of the etiological theorist. Rather, the aim is to explain the diachronic stability of function, and, as in the example above, the continuing emergence of mutation makes such stability something in need of explanation.

In respect to literal teleo-functional ascription, importantly, it seems that only by being held to some standard of performance is function assigned or imposed on some item or behavior. The possession of function, after all, cannot be a fact about the physical constitution of some item, because that item can possess a function even when physically incapable of performing it. If a selectional history was a naturalistic analogue of our holding an item to account or liable for its behavior in respect to some standard, we might have reason to think that selectionist histories do instantiate some norm of performance and, thereby, assign or impose literal teleological function. It is unclear, however, how a selectionist history provides for that. A selectionist history is just that, history; it is not present to provide some naturalistic mechanism to hold some item to account. Nor is it clear that an appeal to the present conditions of some item will help matters. An item or behavior can fail to perform its function not as a result of some internal damage or malformation but because the present conditions prevent it from performing its function. When a female is infertile, an individual sperm will fail to perform its function not because of malformation but because present conditions prevent it. Or, when a bird's wings fail to achieve lift in a heavy windstorm, the bird's wings have not failed to perform their function because they are kaput but because present conditions prevent it. Just as an appeal to present constitution will not do for function assignment because present constitution might render an item incapable of performing its function, so too an appeal to present conditions will not do because an item can possess a function exactly when present conditions prevent its successful performance. So, neither present nor historical conditions suffice to impose function or hold an item to a standard of performance. It is just unclear then why one should think of selectionist histories as instantiating or imposing *literally* some norm of performance. As Davies demands of the nonartifactual teleological realist,

> As naturalists we must ask, What in the process of natural selection makes it true that descendent tokens are "for" the performance of a given task? What in the process of past selective success determines, shapes, or imposes such functional roles? What natural features of the causal mechanical processes that constitute a selective history have the power to determine that descendent tokens are for some task?
>
> …advocates 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. So what in the process of natural selection is responsible for emergence of such norms? What causal-mechanical properties of our history have the power to produce norms that attach to descendent tokens and remain attached even when tokens do not possess the physical capacities required to fulfill the norms? (Davies 2001, 139-140)[62]

Further, the normative modesty of the teleological antirealist seems more than sufficient to make sense of the talk of "success" and "failure" in adaptive or selectionist explanation. A selectionist history might well explain why some present token T exists over variants within the lineage. Also, that the present token is a more or less perfect copy of ancestral tokens that performed E explains why we might *expect* that it will E. And, if present conditions are like historical conditions, the selectionist history will also lead us to *expect* that the present token T, if it performs E, will enjoy the same reproductive advantages as its ancestors. However, all of this is consistent with the normative modesty of teleological antirealism. We can say that "T ought to E" on the basis of some selectionist history, because we have formed certain expectations in respect to T in virtue of that history. *We* are, the teleological antirealist counsels, what holds

[62] Davies sees no prospect in providing a naturalistic answer to the above. As a result, he suggests that any view committed to nonartifactual normativity must be opposed to naturalism.

some present item or behavior to a standard of success or failure, when the item or behavior meets or fails to meet our expectations, respectively. Since it is us and not the world holding items or behavior to standards of success and failure, the resulting assignment or imposition of function is, by the antirealist's lights, a fact about our psychology and not some other fact about the world.

That said, the purpose of this section was to explain the motivation behind the etiological requirement. Neither the appeal to the accidental/non-accidental distinction nor the appeal to explaining the presence of some item succeed in providing a motivation to think an appropriate etiology is a plausible requirement for genuine nonartifactual function. Instead, it is that whole lineages appear sensitive to the success and failure of a teleological characterized activity that makes etiology the plausible ground to distinguish genuine and ersatz nonartifactual function. Despite the criticisms above in respect to normativity, the language of "adaptive success" and "adaptive failures" further seems to tie etiology to the sensitivity condition. As a result, it should not be unreasonable, despite initial appearances, to see the etiological theorist as endorsing the sufficiency of sensitivity to success and failure for literal teleological ascription.

The purpose here was to show that sensitivity to success and failure lies at the core of the distinction between the purposive and the nonpurposive. Such sensitivity, we have seen, seems to be uncontroversially an essential part of the teleological. Both the intellectualist antirealist and the etiological realist, however, appeared to deny that it was sufficient. Yet, in both cases, the plausibility of those positions' particular requirements on teleological ascription – that is, what explained the plausibility and relevance of the representational requirement and the etiological requirement – was that such requirements were an instantiation of sensitivity to success and failure to the teleological characterized activity. It is not unreasonable, then, to follow Ryle in thinking that such sensitivity is at the core of the purposive/nonpurposive distinction. But, accepting Ryle in this regard only seems to make the intellectual antirealist's case, because normativity seems rightly restricted to the intentional and psychological. Over the next two chapters, I will try and show that nonartifactual normativity is a feature of the world, notwithstanding the intellectualist's misgivings, and, consequently, that nonartifactual teleological ascription can be appropriately understood as literal.

Chapter 5

A FIELD GUIDE TO THE NORMATIVE

With contributory effect as insufficient for literal teleological ascription, I suggested in the previous chapter that sensitivity to success and failure did so suffice. As success and failure are only such in respect to some standard or norm, teleological nonartifactual realism presupposes nonartifactual norms or standards to which entities could be appropriately sensitive. Whether literal teleological ascription rightly extends beyond the intentional and psychological depends, therefore, on the existence of nonartifactual norms. The purpose of this chapter and the subsequent chapter then is to find the nonartifactual norm within the physical world and show that non-intentional and non-psychological entities can be appropriately sensitive to such norms. To have confidence that what is picked out as a norm within the natural world is rightly such, it would do to have a field guide such that we might recognize a norm when presented with one. In this chapter, I will develop such a conceptual field guide. Given that field guide, I will, in the subsequent chapter, show that there are nonartifactual norms governing biological systems, and, as a result, nonartifactual teleological ascription can be literally true in biology.

Crudely, the norm is a type of regulative principle, and our field guide should suffice to identify such a regulative principle from other causal principles in the field. What follows is not then an attempt to produce a reductive conceptual analysis of “normativity” and its various constituent and related concepts. Rather, the aim is to extract sufficient distinguishing features such that we can identify a norm when presented with one. What is intended is very much a field guide, just as a book of North American birds is a field guide. Such books do not provide definitions or identity conditions for, say, robins, but they aim rather to provide sufficient identificatory marks such that one might distinguish a robin from a dove, a bluebird, etc. I will begin with a few negative comments about what will not do to identify a norm before turning to a positive proposal.

I. Ersatz normative indicators

First, the terms ‘norm’, ‘normal’, and ‘standard’ are each ambiguous between roughly two distinct senses, and only one of these concerns us here. Such terms can stand for, be near synonyms for, the typical, the frequent, the average, etc.: e.g., “The norm for the American household is 2.5 children”, “It is normal that Larry’s dog barks at the passing mailman,” “The standard weight of books produced in the 18th century was 2 lbs.”, etc. Such usages, serving to indicate the average, the typical, the frequent, etc., are what I want to set aside here. What was of concern in the previous chapters was not that certain behavior or traits are frequent or typical for biological types; no one, I suspect, disputes that. Rather, what was of concern was that sense of “norm” in which an act, event, or entity is correct or incorrect. Deviation from a norm in the first sense is just the quantitative distance from the mean; such deviance is not acting incorrectly or a failure. Driving over the speed limit is to violate a norm in the second sense – that is, it is to act incorrectly, even when driving over the speed limit is the norm in the first sense.

Second, it is a common enough refrain when discussing the normative to talk about the "prescriptive" versus the "descriptive" as if this captured the difference between the normative and the non-normative, respectively. The prescriptive is, however, a species of the normative, namely what Sellars' (1968 [chapter 7] & 1969) described as the "ought-to-do"s. The genus of the normative includes norms, however, that are not rules or prescriptions to act. "Fire exits ought to be unblocked" or "A hammer head's face ought to be 1 ¼ inches diameter" are standards that are not prescriptions. Each might imply[63] some prescription or another such that one brings about, say, that fire exits are not blocked, but the standard itself does not prescribe any course of action. We should not think, then, that the difference between the normative and the non-normative is somehow captured in distinguishing the prescriptive from the descriptive.

Lastly, the presence or absence of 'ought' in a statement is no sign that the statement is or is not one of a norm, respectively. The statement of a norm does not require the presence of an 'ought': "The Manchester train arrives at 6 o'clock" on a printed train schedule is not a prediction but is a statement of a standard. Nor does the presence of an 'ought' suffice to render a statement into that of a norm. It is correct to

---

[63] Sellars would not have agreed to the "might"; he thought that "ought-to-be"s *do* imply "ought-to-do"s:

> … though ought-to-be's are carefully to be distinguished from ought-to-do's they have an essential connection with them. The connection is, roughly, that ought-to-be's imply ought-to-do's. Thus the ought-to-be about a clock chimes ["Clock chimes ought to strike on the quarter hour."] implies, roughly,
>
> > (Other things being equal and where possible) one ought to bring it about that clock chimes strike on the quarter.
>
> *This* rule belongs in our previous category [ought-to-do], and is a rule of action. As such it requires that the item to which *it* applies (persons rather than chimes) have the appropriate concepts or recognitional capacities. (1969, 508)

Sellars' claim here ("standards imply rules of action") rests on the assumption that the existence of a standard presupposes an intentional and psychological agent seeing to it or ensuring conformity to that standard. Part of the purpose of this chapter is to undermine the plausibility of that assumption.

say "The train ought to arrive at 6 o'clock" either after reading the train schedule (presumably a list of standards of train arrival) or after having some factual evidence that the train usually arrives at 6 o'clock. This is not because there are two senses of 'ought' in English.[64] As White writes,

> … The current half-truth that 'the word "ought" is used for prescribing' is no more indicative of the meaning of 'ought' than the opposite half-truth that word 'ought' is used for predicting. In both the subjunctive-governing use and the indicative-governing use 'ought' has exactly the same sense…. What ought to be is, as its etymology in several languages shows, what among the alternatives is *owing* in these circumstances, and under this aspect in order that the requirement be met. It is as if the situation were a pattern with one missing piece, namely what ought to be.
>
> In the indicative use 'ought to V' the only requirement is conformity to the facts and the only aspect is factual. Since, however, the relation expressed by 'ought', unlike that expressed by 'must', is not one of necessity, but of what is owing, what ought to be is that which follows non-deductively from given or presupposed circumstances. For example, if the train left London at 10 o'clock at its usual speed, at what time ought it to arrive at Hull? If he usually works late at the office, then he ought to be there now. If the square root of 900 is 30, then the square root of 837 ought to be about 29. (White 1975, 140-1)

'Ought' serves exactly the same function in the differing contexts of prediction and prescription, namely to indicate what is *owing*. 'Ought' is no different in this respect than any other modal auxiliary, e.g. 'must', 'can', 'may', etc. Not only do each of these find their way into normative contexts (moral, legal, etiquette, conduct, etc.) but they also find their way into a variety of non-normative contexts (economic, physical, biological, romantic, etc.). Notwithstanding the context, 'must', 'can', and 'ought' perform the univocal functions of indicating what is necessary, possible, or owing, respectively. That is, what must be, can be, or ought to be is what is necessary, possible, or owing,

[64] Both Frankena (1950) or Gauthier (1963, 10-12), for example, suggest that there is a prescriptive and predicative sense of 'ought'.

respectively, given the legal, moral, physical, economic, etc. facts.[65] Consequently, I will forgo Sellars' language of "ought-to-be"s and "ought-to-do"s for the more straightforward language of standards and rules, respectively. The moral here is simply that one should not look to a linguistic analysis of 'ought' to distinguish the normative from the non-normative.

II. Constructing the field guide

Enough of the negative. Let the minimal schematic form of a norm be "When φ, ψ". The consequent specifies what is sanctioned or obligated; the antecedent specifies those conditions under which it is so sanctioned or obligated. (The conditional form is suggested just because it makes perspicuous that that which is sanctioned or obligated is often done so only under certain conditions.) E.g.,

> "When soccer is being played, a ball in play is not touched with the hands except by the goalie."
>
> "When x is a house in Forest Hills, x will not be more than 2 stories tall."

The minimal schematic form is intended to cover norms generally – that is, both standards and prescriptions; we attend, thereby, to the Sellarsian lesson that the normative is not identical with the prescriptive. Further, we should not assume that the schema is applicable only to hypothetical norms given its logical form as a conditional. In a Kantian voice, the difference between the hypothetical and categorical norm is the difference between those norms dependent on one's contingent and relative ends and those that are not, respectively. A norm can count then as categorical when applicable only in certain

[65] For detailed considerations along these lines, see White (1975, chapter 10), Wertheimer (1972, chapter 3), & Lycan (1994, chapter 8).

conditions as long as those conditions are not the contingent and relative ends of its subject. The schema is thus perfectly generic, because it rightly does not distinguish between the categorical and the hypothetical nor between standards and prescriptions.

That said, the schema is hardly sufficient, because it is far too generic; it is, after all, equally well a schema for claims other than the normative. This is, I think, a benefit of the schema and not a disadvantage in order to get at the question of what roughly divides the normative from the non-normative. "When soccer is being played, a ball in play is not touched…." can equally well be the specification of a rule of soccer or a factual claim about soccer, albeit a false one. The difference between a statement of fact and of a norm is not to be found in the schema of such statements. Instead, what makes a statement factual or normative is what is subject in the first instance to revision when the world fails to correspond to the statement made. Someone, who makes a factual claim such as "when soccer is being played, a ball in play is not touched…." and then observes a player touch the ball during play, is under an obligation to revise his claim, not the world.[66] The statement as a statement of a rule of soccer, however, is not subject to revision in the first instance by the fact that a player touches the ball; the world ought to be revised, not the statement. While both factual and normative claims are evaluated in

[66] We often make general factual claims, especially in the biological case, for which we reasonably expect exceptions and for which the observed exception is no reason to revise the general claim. For example, "Angelfish [*Pterophyllum scalare*] breed only when water quality is high" is generally true and for which there are observable exceptions. The odd breeding in poorer water quality is not a reason to revise the general claim, because it is intended to be generally true, not universally so. So, it is not surprising that the presence of an exception fails to put us under an obligation to revise the claim. There remains, nonetheless, a threshold of non-correspondence, no matter how vague, at which point mounting exceptions do present an obligation to retract or revise the general claim.

respect to the world, it is the practical consequences of such evaluation that makes for the difference between a statement of fact and a statement of a norm.[67] [68]

The above, though rough, does suffice to provide the sense in which the world is correct or incorrect. Just as a statement of fact is subject to modification so as to correspond to the world, the world is subject to modification so as to correspond to the statement of a norm. Whether the statement of fact or the world ought to be modified is all and only what is conceptually required for correctness and incorrectness: a statement of fact or the world (in the case of a norm) is correct when and only when it ought not be

---

[67] Clearly, the "direction of fit" of the normative is not the sole province of the normative. Propositional attitudes, such as desires, with satisfaction conditions rather than truth conditions will share the same "direction of fit" as the normative. "When soccer is being played, a ball in play is not touched …" can equally well be the expression of a desire or the statement of a rule of soccer. The logical grammar of normative statements and ascriptions of propositional attitudes is sufficiently distinct, however, that we need not be concerned with running the two together. A properly made ascription of a propositional attitude requires an agent to which to predicate the propositional attitude, while a properly made normative statement does not.

[68] I have recently discovered that Wertheimer independently makes much these same points. I reproduce his comments below:

> I call a System a System of Actuality (hereafter, SA) if, when it contains some law '(x) (Fx $\supset$ Vx)' and a case is discovered in which 'Fn. -Vn' is true, then the law is judged to be wrong. A scientific theory is a paradigm of an SA. I call a System a System of Ideality (hereafter, SI) if, when it contains some law '(x) (Fx $\supset$ Vx)' and a case is discovered in which 'Fn. -Vn' is true, then –Vn (the state of affairs, not the proposition) is judged to be wrong. A moral code is a paradigm of an SI. (Wertheimer 1972, 89)
>
> The definitions of an SA and an SI are designed to capture a long-recognized distinction between two sorts of things we call laws. The distinction is expressed in terms of two ways of treating a law. The difference is not a linguistic one; the same grammatical form and the same vocabulary, even the same unambiguous sentence, can (not must) be used to state a law of an SA and a law of an SI. The difference lies, not in the expression of the law, but in what is done with the law. (Ibid, 91)
>
> … it is not a semantic question whether a System is an SA or an SI. Whether a *specific* System is one or the other is a question of fact about how the people that use the System treat its laws…. If in general a counterinstance to a law is treated as a reason for criticizing the law, then the System is an SA. And if in general a counterinstance to a law is treated as a reason for criticizing the violation, then the System is an SI. Roughly, a law of an SA is a description of a regularity, and a law of an SI is a norm (which is not to say that it must lack truth-value). This is a difference in illocutionary force, and thus in the use of sentences, not necessarily in their meaning or grammatical form. (Ibid., 140-1).

modified; and, a statement of fact or the world (in the case of a norm) is incorrect when and only when it ought to be modified.[69]

Stating a norm, however, need imply no obligation on the part of the speaker to revise the world. Reporting “In the U.K., one drives on the left side of the road” or “The train arrives at 6 o’clock” need incur no obligation to amend driving habits or improve upon railroad efficiency. To report the train schedule is report that someone is under an obligation to bring it about that the train meets that schedule. When the train fails to arrive on time, it is up to the relevant employees, and not me, to remedy the situation. Consider the analogous case with a statement of fact. When I assert that such and such is the case, I take on an obligation to revise or retract my assertion if the world fails to correspond to that assertion. When I report another’s assertion, I am reporting that someone has taken on the appropriate sort of obligation. So, too, it seems to report a norm is to report that someone is under an obligation to bring the world into line when it deviates.

The last is still too rough, however, because it overly intellectualizes what is required to report a norm by its inclusion of ‘obligation’. Even in the artifactual case, to report a norm need not require that anyone is under an obligation to bring the world, when it deviates, in line with the norm. All that is required is that we have reason to believe that the world, when it deviates, is subject to being brought into line with the norm. Let’s say that, in observing Larry at work in his garden, we have made the following observation: Larry’s daffodils are in rows (or, when in Larry’s garden, daffodils are in rows). A few of his daffodils are out of line in respect to the rows, but enough appear to be in rows that a pattern is generally captured by the claim “Larry’s

[69] For a similar treatment of ‘right’ and ‘wrong’, see Wertheimer’s (1972, chapter 5) detailed discussion.

daffodils are in rows." Our observation is at least an observation of a regularity with some limited number of exceptions. But, we also observe that, when a daffodil grows out of line due to disbursement by squirrels or the division of bulbs, Larry weeds out those exceptions to the regularity "Larry's daffodils are in rows." Larry need be under no obligation to keep weeding or to have so weeded in the first place, because such selective weeding is just what Larry does. Notwithstanding, a standard seems in play in Larry's garden, namely "Daffodils ought to be in rows", because Larry imposes such a standard on his garden.[70]

But why does a standard seem to be in play in Larry's garden? We observed a pattern in Larry's garden, namely "Larry's daffodils are in rows", to which there were exceptions. We observed that only exceptions to "Larry's daffodils are in rows" are weeded out by Larry as he walks through his garden. Part of what explains the observed pattern in the world is that Larry acts to weed those out of line. Had Larry failed to so act, the observed pattern would have differed from that in fact observed. A norm is in play in Larry's garden, because exceptions to "Larry's daffodils are in rows" are subject to elimination such that the world comes to correspond to "Larry's daffodils are in rows" in a way that it would not have done otherwise. This is just where we began: "Larry's daffodils are in rows" is reasonably construed as the statement of a norm, because exceptions to it are subject to revision such that world comes to correspond to it. To report a norm does not, then, require that we have reason to believe that someone or another is under an obligation to bring the world, when it deviates, in line with the norm.

[70] Daily life affords numerous examples of norms that are, so to speak, imposed on the world without obligation: "In Grandma's house, one takes off one's shoes," "The light bulbs are changed every two weeks," "The pictures in the living room are to be straight," "Children in this house eat with their mouth closed," "There is no smoking in this car," "One feeds the chickens and then the pigs," "Tuesday is the day to wear funny hats," etc.

Rather, to report a norm just requires minimally that we have reason to believe that exceptions to the putative norm are subject to modification so as either to become instances of it or cease to exist.

Much rests on these phrases “subject to modification” and “subject to revision”. In the case of a norm, an exception is consequential not to the statement of the norm but to the exception such that the exception is brought into line with the statement. But, it would be too strong to say that every exception to a norm *is* modified so as to correspond with that norm. Norms are more or less efficacious: not every exception to a norm will be the target of criticism nor will every bit of criticism find a responsive target (e.g., neither the murderer that gets away with his crime undetected nor the murderer unresponsive to social sanction show that there is no prohibition on murder). We do, however, in reporting a norm need reason to think that there are practical consequences to there being that norm; this is exactly what allows us to construe a statement as that of a norm. What we need reason to believe in reporting a norm is, then, that exceptions ought to be revised.

That last ‘ought’ need not be, however, in the business of stating some further norm: it expresses what is owing given that we have reason to believe that there are consequences to being an exception. Understand, then, the ‘subject’ of “subject to modification” as it is used in sentences like “When starting daycare, children are subject to colds,” “The elderly are subject to dementia,” or “In their burrows, ground squirrels are to subject to predation by snakes.” Whether our reasons to believe that exceptions will *to some degree* be modified[71] are that there is some further norm, governing how to handle the exceptional case and that itself has some practical consequences, or just that exceptions are, more or less, so acted upon makes no difference to whether there are practical consequences of the appropriate sort, namely that exceptions are subject to modification so as to correspond to the norm.

Perhaps, the last seems too quick. One might think that not only do we need reason to believe that there are practical consequences of the appropriate sort but also that an intentional and psychological agent is, even if not obligated, seeing to it that those consequences ensue. If this is a further conceptual requirement on the reporting of a norm, then it is not just that an intentional and psychological agent is somehow involved in the enforcement of a norm. I am in a sense involved in my knee reflex (it is my knee, after all), but my involvement is not in virtue of or to be explained by my psychology. Rather, it should be the case that such "seeing to it" is explained as the intentional product of the relevant agent's mental states. For example, Larry acts to weed those daffodils out of line, because Larry perceives daffodils out of line, believes that they are out of line on the basis of that perception, and believes further that they ought not be. It should come as no surprise to Larry, if he is "seeing to it", that he holds his garden to a certain standard – that is, it should come as no surprise that he believes daffodils out of line ought to be weeded. However, it is not at all implausible that Larry will be surprised by our observation that he holds his garden to a certain standard. We and Larry could explain his actions by appeal to some subterranean set of beliefs and desires. But, it is because both Larry and ourselves have reason to believe that there is such a rule,

[71] To what degree must a norm be efficacious for its reporting is not something, I suspect, rightly answered in the abstract, no more than it is rightly answered in the abstract at what point do mounting exceptions prompt retraction or revision of a statement of fact (see note 66).

independently of what Larry thought he believed, that we are prompted to offer explanations invoking the apparatus of folk psychology.[72] Or, consider the dean of admissions that we accuse of imposing an admissions standard excluding particular races. He might vigorously and with all sincerity deny that it was ever his intention to impose such a standard. What he is not denying is that his actions have in fact imposed a standard; what he does deny is that that was his intent. We might suppose some subterranean beliefs and desires to explain his actions, but we need not do so. We could, giving the dean the benefit of the doubt, think that the imposition of such a standard was an accident in respect to his psychology. Our reason to believe that there is a standard imposed by the dean's actions does not first presuppose that he is "seeing to it" in some psychological sense that the relevant practical consequences ensue. All that is required for us and the dean to recognize the admissions standard is that the appropriate sort of practical consequences do ensue.

But, it has been pressed on me that the *appropriate* sort of practical consequences can only ensue if the relevant corrector has the authority to bring the deviant back into line. The point has been pressed as an objection that one cannot possess authority or be authorized unless one is an intentional and psychological agent.[73] Authority, the power to enforce obedience, has both a clear *de jure* and *de facto* sense, however. In its *de jure*

[72] I imagine a conversation like the following:

> Observer: "Larry, why do you keep your daffodils all lined up in rows?"
> Larry: "I do?"
> Observer: "Well, yes, you seem to. Look, you are always out here pulling out daffodils, but the only ones you ever pull are those out of line."
> Larry: "Hmm. You're right. I do seem only to pull those out of line."
> Observer: "Do you like them that way, lined up in rows, I mean?"
> Larry: "Well, you know I never really thought about it. But, now that you mention it, I guess that I must like them that way. I can't think of why else I would keep pulling the ones that I do."

[73] Thanks to Dave Landy for this point.

sense, some rule or standard confers permission or obligates enforcement of some *further* rule or standard. For example, the law empowers me to enforce certain standards of conduct on my child's behavior. In its *de facto* sense, one possesses authority to enforce a standard or rule when one possesses the power to enforce obedience, independently of any further rule or standard. The tyrant possesses, through brute force and will alone, the power to enforce obedience. He is the *de facto* authority in such matters. We can ask whether the tyrant ought to act so tyrannical, but, in so asking, we are not questioning that he does enforce rules; that is the presupposition of the question itself. So, it is the case that the relevant corrector must possess the authority to bring the deviant into line for the appropriate sort of practical consequences to ensue, but, given the *de facto* sense of authority, this does not imply that only intentional and psychological agents can be so empowered.

Additionally, it has been pressed on me that in another respect the emphasis on practical consequences places too high a standard on the existence of a norm.[74] Given the above, unless there is some corrective mechanism such that exceptions are subject to modification, there is no norm or standard. (Additionally, unless those exceptions, as subject to correction, are modifiable, there is no norm.) But, we recognize, the objection goes, the existence of norms or standards even in the absence of any corrective mechanism. Stock examples of such norms are those peculiar and strange laws on the books that have not nor will be enforced. For example, there was, to my knowledge, a law in the state of Washington until quite recently that forbade a man cursing in the presence of a woman. No one has or will likely ever enforce this law nor is it the case that, when men do not curse in the presence of women, such men's actions can be

[74] Thanks to Marc Lange for this point.

explained as conforming to the norm (if for no other reason than that they are likely ignorant of it). Such laws are norms for which no exception is in fact subject to modification; therefore, the emphasis on practical consequences is just too stringent a standard on the existence of a norm.

This objection, however, can be just set aside, because there is no reason to think that the stock examples are examples of existent rules or standards. Merely stating that such and such is a rule is not sufficient for there to be a rule. My declaration "all are required to tip their hat to me" does not make it so. Does the fact that the statement of a rule is found in a published legal code change matters? No, it doesn't. Bringing about rules and standards is an act that the legislative body is empowered to perform. But, the success of such an act cannot be the mere publication of a rule in the legal code. Just as illocutionary act unheard or unheeded is a failed illocutionary act, so too a published rule unread or unobserved is a failed act. The basic principle is the same: saying that there is a rule is just insufficient for there to be a rule. What then are the grounds to think that what is printed is in fact a rule? Well, I offer the above proposal. Without the relevant practical consequences, any claim that such and such is a rule degenerates into merely saying that it is so, and merely saying it is so never suffices for it be so. Since there is no reason to think that the "norms" on which the objection relies are anything more than merely apparent, let us set it aside.

III. The field guide

Enough is in place to extract a general moral on how to identify a norm in the field.

What is in the field to observe are various patterns. Following Chaitin (1975) and Dennett (1991), we have a pattern iff there is a more efficient description of a series than a verbatim bit map. That of a series not captured by that more efficient description is considered noise. Importantly, a described pattern, though a regularity, need not be a description of that which is typical, frequent, or average, etc. A series with 80% noise still affords a more efficient description than a verbatim bit map of the series.

Take "When φ, ψ" to be a pattern description. If "When φ, ψ" is a norm, then exceptions in virtue of being exceptions need be subject to modification so as to correspond to it. With that in mind, let's divide patterns into two species: mere regularities and normative regularities. That division reflects the explanatory relevance of noise or exceptions to a pattern. In the mere regularity, exceptions or noise to a pattern are just that, exceptions or noise, and the explanatory projects in respect to such noise are to explain why there are exceptions from the pattern, what it is about the world that makes some of a series instances of a pattern and some not, etc. In contrast, the normative regularity is one where the stability of the pattern is to be explained in part by the fact that exceptions or noise to the pattern (by being exceptions) are subject to modification such that they become instances of it (or are subject to elimination such that the series better approximates the described pattern). That an exception is *subject* to modification or elimination requires at least a further *mere* regularity such that exceptions become non-exceptions or are eliminated, respectively.[75] So, part of the explanation of the stability of the pattern is that a corrective mechanism is in place such that exceptional members of a series are brought into line.

Noise or exceptions need not be present, however, for a pattern to be rightly a normative regularity. From the previous section, what is required is that exceptions be subject to modification so as to correspond to the norm. In the absence of noise, a pattern is rightly characterized as normative if exceptions would be subject to modification if they were to occur. That an exception would be subject to modification if it were to occur is part of the explanation of the stability of the pattern: the pattern is stable, because if exceptions to the series arose, they would be acted against. For example, at a press conference, each member of the press is wearing their press pass. There is a guard who examines each member of the press to see whether they are wearing their press pass and is prepared, if any were to fail to show their press pass, to ask them to show it. We have reason to believe there is a norm in play, even though there have been no exceptions to it: That a guard is in place serves in the explanation of why the observed pattern to date is likely to continue as it has, and that explanation is one that makes reference to the fact that exceptions to the putative norm are subject to modification so as to correspond to it.

While an exception need not have occurred for a regularity to count as normative, it does need to be case that an exception can occur. Absent the possibility of deviation, no part of the explanation of a pattern's stability will be that exceptions are subject to modification. Even with a corrective mechanism standing by, it will not be part of the explanation of a pattern's stability if it is just not possible for there to be any exceptions

[75] Since the relevant mere regularity can itself be exception-laden, not all exceptions to the normative regularity by being subject to modification will always be so modified.

on which it can act. Simply, without the possibility of deviation, there is no reason to think that exceptions are subject to modification, because there can be no exceptions. As such, no norm is in place. Think again of the press conference. Suppose that each member of the press has their press pass indelibly tattooed to their forehead before entering the conference. Our guard has been rendered redundant, though his presence and vigilance might present the appearance of a norm, e.g. "Each member of the press must have a press pass". It would be wrong, however, to read "Each member of the press must have a press pass" as the expression of a rule of conduct, because there is no reason to think that exceptions are subject to correction. Rather, the 'must' is rightly that of a factual sort of necessity: it is just a matter of fact that each must wear a pass. From an explanatory viewpoint, there is just no value in explaining the relevant pattern of behavior as normative, because, despite his presence, the guard has nothing to do with the stability of the relevant pattern.

Having teased out one way in which a pattern might present the appearance of a norm without in fact being normative, let me note three more sorts of cases.

a. *An ersatz norm: the absence of targeted correction*[76]

Suppose the statistically normal life of a car is 20 years, but suppose that in Seattle we notice that cars on the road are less than 10 years of age. It is statistically normal that cars on the road in Seattle are less than 10 years of age, because the heavy precipitation there enhances the rate at which cars rust out. Is "Cars in Seattle are less than 10 years old" the statement of a norm or a normative regularity? Well, no, but it might strike one, given the above, to appear to be. "Cars last 20 years" is a pattern in the

[76] Thanks to John Roberts for this point.

world, so we know how things ought to be if it were not the case, as it is in Seattle, that there is a further regularity, namely rain-induced rusting out of cars. So, if it were not for the heavy rainfall in Seattle, cars would exhibit a pattern different from that they do exhibit in Seattle. What explains why the cars in Seattle exhibit the pattern they do and its stability is the presence of some further causal mechanism, i.e., rainfall. So wouldn't the rain, by the above, be a corrective mechanism explaining the stability of a pattern? If so, "Cars in Seattle are less than 10 years old" looks to be a statement of a norm according to the theory.

Let me try to explain why it is not. The rust-inducing rain responsible for the short-lived cars is not acting on exceptions so as to ensure "Cars in Seattle are less than 10 years old." That is, the rain is not just falling on the 11 or 12 year old exceptional cars and, thereby, eliminating them from the population of cars on the road and ensuring the stability of the pattern. The rain is acting on every car and in so doing speeding up the rate of dissipation. The pattern is as stable as it is, because all the cars, independent of whether they are exceptions or not to "Cars in Seattle are less than 10 years old", are acted against. There are, then, no practical consequences to being an exception, and so, no reason to think "Cars in Seattle are less than 10 years old" is the statement of a norm.

Let's modify things slightly. Let's make the cars under ten years of age immune in someway to the rusting effects of precipitation. So, for example, at the factory, each car is treated with an anti-rusting agent, and that anti-rusting agent wears off at or around the ten year mark independent of whether the car is rained upon. Now, we have rendered rain impotent in respect to the younger cars, and it is only capable of having causal effects on those older cars. So, it might seem that we have got it so that rain acts only on those exceptional cases, i.e., cars over ten years of age, and so achieved a normative regularity. However, it is still only a merely apparent norm, and it is so for the same reason as before. While the rain is now only causally efficacious in inducing rust on the older cars, the rain does not act on the older cars alone. Rain still falls on each and every car. The problem is, as it was before, that the rain does not act on some exceptional case because it is an exceptional case – that is, the rain does not fall on a car because it is over ten years of age. It is this that we need to report a norm.

To think that being an exception is consequential is to think that something is acted against because it is an exception. That a car is one, two, or twenty years old does not make it subject to modification in the scenarios above. What we need is a corrective mechanism that acts on an exception because that exception fails to act or be in a way corresponding to the norm. Larry in his garden does not act on every daffodil, but he acts only on those out of line because they are out of line. The dean does not act on every applicant, but only on those applicants of a certain race because they are of that race. A causal mechanism is not a corrective mechanism (and, thereby, a regularity is not normative) unless the causal mechanism acts on exceptions in virtue of their being exceptions. So, if rain only fell on older cars in virtue of their age, then we might have cause to view the pattern "Cars in Seattle are less than 10 years old" as a normative regularity.

b. *An ersatz norm: random correction*

Suppose "When φ, ψ" to be a pattern description and the explanation of why the series approximates the pattern description as well as it does – that is, why the pattern is as stable as it is – is that historically exceptions have been acted on so as to make them instances of the pattern. However, let us suppose that the occasions in which exceptions have been brought back into line are the result of varied random forces.[77] For example, we observe a pattern in an ant's foraging behavior: whenever the ant starts out to forage, it turns 90 degrees from its initial position before proceeding. We also observe on a few occasions that it fails to turn before proceeding. However, in each exceptional case, a distinct random something occurs to our ant to put him back on the track that he would have been on had he turned, e.g., the wind knocks him back on course, he slips on some sand, a leaf falls in his path forcing him to turn, etc. In such a case, it is right to think that there only appears to be a norm in play and not there is in fact a norm. The pattern to date presents the appearance of stability, but it is not in fact stable because there is nothing causing it to be stable. We can explain what caused each exception to be altered. What we cannot explain is what causes exceptions to be altered; this is just what it means to say that the "corrective" events are random. This brings out the importance of the earlier comment: part of the explanation of the stability of the pattern is that a corrective mechanism is in place such that exceptional members of a series are brought into line. Without such a corrective mechanism in place, we do not have reason to believe that a pattern is stable because exceptions are subject to modification, and so, we do not have reason to believe that there is a norm in play.

c. *An ersatz norm: merely historical correction*

Norms are temporal: they exist at some times and not at others. The norms governing the ritualistic cannibalism of the Fore Highlanders of Papua New Guinea,

[77] Thanks to Dave Landy for this point.

governing the custom of the "Sacred Spring" among the Italian highlanders of the fifth century B.C., or governing verb conjugation among Carthaginians are all clear examples of norms that once were but are no longer – that is, they once had causal influence on the world but no longer do. If a pattern has gone out of existence, then it should be clear enough that there is no longer a norm. But, there is a sort of a case where a pattern was once a normative regularity but no longer is – that is, there was a period in which the stability of the pattern was explained by appeal to exceptions being subject to modification but no longer is. For example, the man that continues unconsciously to lay his coat over puddles for women to cross unsullied would have been conforming to a norm if the relevant norm of social etiquette applied, but in the present case, when the continued existence of such a norm is doubtful, his action neither conforms nor fails to conform. He acts as he does, because there *was* a certain norm but not because there *is* such a norm. If he were now to fail to act as he ought to have in the past, he is not subject to criticism for his failure. The stability in his pattern of behavior is not to be explained by the fact that he is subject to criticism for failing to so act. He then appears to conform to a norm because of the historical connection of his behavior to modifying acts. However, there is no reason to think that exceptional acts on his part are subject to correction, and so there is no reason to think that a norm persists.

IV. The field guide as a guide to the teleological

The above might serve as a guide to identifying norms, but, given the previous chapter, it is sensitivity to such norms that is the issue for teleology. That is, for some entity's activity to be rightly characterized as teleological, that entity need be sensitive to the success or failure of that activity in respect to the relevant *telos*. It might seem that we need some further guide for the identification of such sensitivity to success and failure in addition to the guide for normative identification. However, the above will serve just as well for identifying sensitivity to success and failure. When some pattern of behavior can rightly be described as a normative pattern, it is such exactly because some entity is sensitive to the success and failure of its actions in light of the relevant norm. Alternatively, when some item is subjected to a critical or corrective mechanism such that its pattern of behavior is to be explained in part by such correction, then we have the relevant sensitivity to success and failure.

Think back to von Uexküll's night moths. von Uexküll suggests that the night moth is not acting to avoid predators because it is insensitive to its failures in avoiding predators. That the night moth will always flee a squeaky stopper shows its failures to be inconsequential in respect to its behavior – that is, the night moth does not modify its behavior in light of its mistakes so as to engage in the behavior with more success in the future. The putative *telos* of predator avoidance provides a standard of correct and incorrect action in virtue of which some of the night moth's acts are successes and failures, respectively. Its failures are exceptions to the putative norm of performance, but those failures, those exceptions, are without consequence. The reason to deny the night moth's behavior is rightly teleological is just the same reason, then, to deny that there is a norm of performance. In contrast, the earlier hypothetical night moth that stops fleeing squeaky stoppers but continues to flee bats is sensitive to its success and failure in respect to the *telos* of that behavior. Its failures to avoid predators are consequential to its subsequent predator avoidance behavior. Alternatively, its exceptional or incorrect behavior is consequential to the behavioral pattern the night moth exhibits, because such exceptional behavior is subject to modification so that its subsequent behavior will tend to be in line with the standard. Here too, the reason to think the night moth is sensitive to its successes and failures and, therefore, its behavior rightly teleological is just the same reason to think there is a norm of performance.

It is perhaps easier to concede to this in the case where there has been historical deviant behavior, i.e., we have in the past seen exceptional behavior and that behavior having been subject to correction is part of the explanation of the overall pattern of behavior. However, I allowed above that it need not be the case that there has ever been deviant behavior for the identification of a normative pattern. Instead, all that is required to think that a norm is in place is that there is a corrective mechanism in place and deviation possible. But, that requirement is loose enough that that identification of a norm might seem to come apart from teleological ascription. Thinking of the initial press example above, while the members of the press are subject to a standard, it might seem incorrect to say that their actions have purpose. They happen to be meeting a standard of action of which they are to date ignorant. The guard, as the enforcer of that standard, might know that the purpose of wearing a press pass is to identify members of the press, but the press do not by hypothesis. As such, it would seem that at least in these sorts of cases that something further is required over the identification of a norm to ascribe purpose to behavior.

Notwithstanding the last, I want to maintain that the grounds to identify a norm suffice to identify teleological activity. Let me offer an example that suggests that this should turn out to be the case even in the absence of historical deviance. Suppose that a field of mushrooms grows just outside some village, and suppose that the villagers regularly consume those mushrooms. Assume that those villagers have never taken any action to ensure the presence of those mushrooms nor for that matter are they in any way prepared to take any such action. The field of mushrooms, we should be able to agree, is not an artifact of our villagers: the mushrooms provide nutrition, but they are not for the provision of nutrition. On the other hand, we should also be able to agree that, if our villagers actively tend the field of mushrooms, that the field of mushrooms is an artifact: the field serves the function of providing nutrition to our villagers. In the two cases, we have some pattern of mushroom growth and subsequent consumption, and we can explain the pattern in the latter case by appeal to some standard of performance to which our villagers hold the field. There is an intermediate case, however. The field of mushrooms happens to grow next to the village, our villagers consume those mushrooms, and our villagers have never acted to ensure that the field continues to supply mushrooms. However, unlike the initial scenario, our villagers are prepared to act should the mushroom field not yield sufficient mushrooms (e.g., they have fertilizer at the ready, are ready to haul buckets of water, etc.). This intermediate scenario is now like the case of the press pass. A corrective mechanism is at the ready should the mushroom field fall out of line, but that mechanism has yet to be used. That the villagers are prepared to act should the mushrooms fail to live up to some standard of performance shows us that field of mushrooms is for the provision of nutrition.

The mushrooms in this intermediate case have a function not because they have been subjected to a norm of performance but because they are being held to one. In the case of a normative pattern without historical deviants, arguably an explanation of that pattern to date need not make reference to the fact that there is a corrective device standing by. For example, I could explain the pattern of mushroom growth near the village in the intermediate case without discussing the villagers' preparedness to keep the mushrooms growing, because they have never so acted. However, the normative pattern without historical deviance is a *normative* pattern exactly because the explanation of why the pattern *is* as stable as it is does require that we invoke the fact that a corrective device is standing by. That is, the pattern of mushroom growth along our village is as stable as it is, because were it to deviate it would be brought back into line.

In respect to the earlier press room example, the wearing of a press pass serves a function, even though the members of the press are ignorant of it, because the stability of the pattern is to be explained in part by reference to a guard standing by ready to bring deviation back into line. The reason for the apparent tension between the grounds to identify a norm and teleological ascription is the press members' ignorance of the norm. Their ignorance is not irrelevant, however. That the press members' are ignorant of the norm means the explanation of their actions to date will not make reference to the relevant norm. As a result, it would be wrong to think that they are acting purposively in wearing their press pass. But, that they have not yet acted purposively does not imply that their actions do not have a purpose. The mushrooms in none of the above scenarios can be said to have acted purposively to provide nutrition for our villagers. The mushrooms do not alter their behavior in light of some standard of performance. The mushrooms have a function not because they act in light of some standard but because they are held to a standard. Similarly, when I act purposively, my actions have purpose insofar as I hold my actions to a standard of performance. The members of the press have not acted purposively in wearing their pass, because an explanation of their behavior will not refer to their holding themselves to some standard of performance. However, their actions are held to a standard, just as the mushrooms are or my individual acts are when I act purposively, and it is because their actions are held to a standard of performance that their actions have purpose.

The aim here was to become clear on how to identify a norm from other regulative principles in the field. The different practical consequences of a statement of fact and that of a norm provided a starting place. If the statement of a norm is distinguished from that of a fact by having the consequence that the world, and not the statement, is revised so as to correspond to the norm, then we have reason to think there is a norm when just that sort of practical consequences ensue. The aim of introducing patterns was just to make those intuitive grounds for normative identification more explicit. The question now that we have such a field guide in place is whether anything in the world, beyond the artifactual, exhibits the relevant patterns and so exhibits the presence of a norm. It is to that question that I will turn in the subsequent chapter.

Chapter 6

PHYLOGENIC, ONTOGENIC, AND SOCIOGENIC NORMATIVITY AND TELEOGY

We are now in a position to tie together the preceding chapters and answer whether nonartifactual teleological ascriptions can be literally true. From chapters 2 and 3, we have the biological functional ascriptions whose literal truth is in question, namely those functional ascriptions in the service of explaining the systematic effects of autocatalysis and controlled heritable variance. From chapter 4, we saw that, if those biological functions are genuinely teleological, it needs to be the case that the relevant system is sensitive to its success or failure in respect to that function. The last required that the system be sensitive to a norm of performance, and so in the previous chapter, I developed a field guide to allow us to recognize norms in the world. Given that guide, the question of the literal truth of a teleological functional ascription resolves to whether a system exhibits a pattern of behavior appropriately described as a normative regularity. Let us see then whether the biological functions in question satisfy this requirement.

I. Genuine phylogenic normativity and teleology

It will come as no surprise that I will suggest that biological functions do satisfy that requirement. I want to start with a case that prima facie does not and from there develop which functions and what kinds of systems do satisfy the requirement.

Caridean shrimp (an infraorder of the order Decapoda and containing over 2,800 species) possess two sets of antennae. The first of these, the antennules, are divided into two branches, and the outer of these branches, the aesthetascs, have cuticular hairs used in the detection of dissolved substances. (Ache 1982) The antennules are waved or fluttered frequently and, thereby, circulate water through the aesthetascs. Dissolved substances diffuse through the thin cuticles of the aesthetascs and come into contact with sensory cells. The activity of the antennules and the action of the aesthetascs are part of the olfactory system of the shrimp. The olfactory system's contribution to the autocatalytic system of the individual shrimp is the detection of food and members of the opposite sex. (Bauer 2004, 15 & 79) The antennules' contribution to the olfactory system is to bring relevant chemical stimuli to the attention of the relevant processing subsystems. So, let's say that the task of the antennules is the reception of chemical stimuli. (I say "reception" and not "detection", because the antennules are not involved in the processing of the stimuli.)

Is the teleologically characterized function of the antennules genuinely teleological or merely analogous in virtue of the contributory description? "Antennules receive chemical stimuli" is a pattern description of the contributory behavior of the antennules to the olfactory system, and so we need to ask whether the shrimp as a whole or some one of its component systems is sensitive to the success or failure of that task – that is, is "Antennules receive chemical stimuli" reasonably construed as a norm of performance?

"Antennules receive chemical stimuli" is a pattern of behavior for which we can offer a straightforward causal-mechanical explanation. We can explain how reception is performed by decomposing that system into a functional flowchart that includes the waving motion of the antennules, the uptake of chemicals by the cuticles of the aesthetascs, etc. It is the case that “Antennules receive chemical stimuli” is at least a regularity for which we can offer a functional-mechanical explanation.

But, here are a few further facts that should make that explanation seem lacking:

> Aquatic environments are “dirty worlds” with abundant suspended particles of sediment and detritus that can cover and foul surfaces. There is a constant influx of fine sediment from terrestrial erosion into the sea via the rivers. Detritus is particulate matter derived from decomposing organisms, fecal particulars, and mucus sloughed off fish and many invertebrates. Both sediment and detritus are stirred into suspension by the action of currents, waves, and the activities of organisms. These particles settle back under the force of gravity and will cover surfaces, including those living organisms, unless they are removed in some way. (Bauer 2004, 78).

Given the waving of the antennules, the thin comb of the aesthetascs is an ideal collector of the sediment and detritus in the aquatic environment. Further, “the aesthetascs are ideal locations for microscopic fouling organisms to grow in.” (Ibid. 79) Fouling of the aesthetascs by either sediment and detritus or microbial growth will interfere with the functioning of the aesthetascs (remember, they operate by allowing chemical stimuli to pass through the cuticles, and a dirty comb of cuticles is not one through which chemical stimuli can pass.) Under normal conditions, the fouling from microbial growth alone by diatoms is so intense that within two weeks the cuticles will break off under the weight of the infesting diatoms. (Bauer 1975 & 1977)

Given the constancy and intensity of fouling in the caridean shrimps’ environment, we should expect that very few of them have working aesthetascs. The facts, however, are just the opposite. Caridean shrimp with working aesthetascs is the norm (quantitative sense) and not the outlier case. The functional-mechanical explanation of how antennules receive chemical stimuli does not explain this latter fact. It explains how, when the antennules operate, they do so and, given the last, under what conditions we might expect that they fail to operate. The conditions in situ are exactly those in which they ought to fail to operate, and so the functional-mechanical explanation of antennular chemical reception does not explain the fact that the pattern of operating antennules is as stable as it is.

What explains the stability of that pattern is a further fact about shrimp behavior. The third maxillipeds (a leg-like appendage) of caridean shrimp are covered in rows of stout serrate setae (i.e., bristle or hair), and the shrimp with some constancy clamps its third maxillipeds around an antennule, draws the antennule through the rasps of the setae, and strips off the debris and microorganisms on the aesthetascs. The result is that aesthetascs are kept free of fouling. (Bauer 2004, 78-80) What explains the stable operation of the antennules is that a further regular pattern of behavior (i.e., grooming by the third maxillipeds) ensures that antennules will continue to receive chemical stimuli.

Given that the shrimp is engaged in this further activity to ensure the successful operation of the antennules, we might think it plausible to say that the shrimp is sensitive to the success and failure of antennular operation. And so, reception of chemical stimuli is a genuine teleological function of the antennules. However, that thought is too quick, because we do not have in this case any reason so far to think that there is a norm to which the individual shrimp is sensitive. While it is true that the third maxillipeds ensure that the regularity "Antennules receive chemical stimuli" is as stable as it is and in so doing ensure the antennules' contribution to the olfactory system, it is not the case that what explains the stability of the pattern of antennular chemical reception is that exceptions are subject to modification or correction. The conditions in which caridean shrimp find themselves are intensely and consistently "dirty". Unsurprisingly, the grooming activity is constant and untied to the present condition of the aesthetascs. A shrimp placed in clean water and with no environmental fouling of the aesthetascs will continue grooming just as before. (Alternatively, the addition of fouling by the experimenter will not generate increased grooming.[78]) The grooming behavior of the individual shrimp is not prompted by a failure of the aesthetascs to perform their function. The stability of the pattern "Antennules receive chemical stimuli" is explained by the grooming behavior. But, since the grooming behavior is not prompted by the antennules' failure to operate properly, it is wrong to think that the grooming behavior of the third maxillipeds is a corrective mechanism. As a result, it is wrong to think of "Antennules receive chemical stimuli" as a norm of performance to which the system is sensitive. Simply, since it is not in virtue of its failure to achieve reception that the shrimp cleans its aesthetascs, the individual shrimp is insensitive to the success or failure of chemical reception.

The above is a lot of detail on shrimp only to reach a negative result. I have spent so much time on that example, because it provides some cautionary as well as positive morals worth keeping in mind as we proceed.

The above is a clear example of homeostasis. Antennular chemical reception quickly degrades in performance under normal water conditions. However, we have a system of the shrimp dedicated to maintaining the performance of the antennules in the face of external disturbances. For someone such as Nagel, homeostasis should suffice for genuine teleological function. But, if the above is correct, homeostasis does not suffice

[78] There is some observational evidence of this later option. (Bauer, R., personal communication.)

for literal teleological ascription. So, as a cautionary note, we should not think that the matter of nonartifactual teleological ascription is decided by the presence of homeostatic systems within organisms. It is worth cautioning against, because on first blush it might seem that homeostasis offers just what was required for normativity. After all, in such a case some causal-mechanical system keeps some other system stable in its operation. But, that on its own is not yet enough to conclude that there is a norm of performance to which the system is sensitive.

What would be required to turn an example of homeostasis into one involving literal teleological function? Well, it would have to be the case that the homeostasis is maintained by the corrective device in virtue of degrading performance or failure. That is, what explains the stability of the behavioral pattern is that exceptions are consequential by prompting modification to the output of the relevant system. With our shrimp, if it were the case that the grooming behavior was causally connected to the degraded performance of chemical reception by fouling – that is, if we had some feedback loop between, say, the water flow into the cuticles and the initiation of grooming behavior, then we would have grounds to think that what explained the stability of the pattern "Antennules receive chemical stimuli" was that exceptions to that statement were subject to modification.

That such a simple feedback loop within shrimp would apparently suffice to render the function of chemical reception genuinely teleological explains why Rosenblueth et al. (1943) see teleology as synonymous with feedback. Additionally, it helps explain the intuitions of, say, Dretske or von Uexküll that, when the individual is insensitive to some phylogenetically determined function, that function cannot be genuinely teleological. That said, and despite all that has come before, it should appear somewhat odd that the addition of a simple feedback mechanism would suffice to provide the antennules with genuine teleological function.

Let me see if I can explain just what is odd here. It is not all that difficult to imagine an evolutionary scenario in which some present population of shrimp might come to possess just such a feedback mechanism. Assume conditions alter such that the rate of fouling is no longer constant and the shrimp experience only limited periods of intense fouling. Grooming behavior is not free; it comes with metabolic and other costs (e.g., perhaps, increased detection by predators). That said, failing to groom when the aesthetascs are fouled is very costly to the shrimp. So, even in variable fouling conditions, the best strategy for the shrimp might be continuous grooming. However, the shrimp has a highly reliable indicator available to it that fouling is occurring, namely that water flow through the cuticles is diminished. Assuming that indicator is reliable enough to overcome the cost of error (failing to detect in a fouled state), it would not be unreasonable to expect our population of shrimp to develop some simple feedback mechanism prompting and halting grooming behavior.[79] So, assume that conditions are such that some population comes to have some feedback mechanism governing grooming behavior. Now all of a sudden the chemical reception of the antennules has become genuinely teleological. But, just as suddenly, we might see that very same population give up that feedback system. For example, some novel form of diatom starts to inhabit the cuticles of shrimp without restricting water flow. As a result, the detection mechanism becomes less reliable. Assume that it is sufficiently unreliable that the benefit

[79] For the formal model of this sort of evolutionary cost-benefit model for signal detection, see Godfrey-Smith (1996 [chapters 7 & 8] & 2002). Independent variants of that model can be found in Moran (1992) and Sober (1994).

of flexible grooming behavior is now outweighed by the costs of failing to detect fouling. As a result, it would not be unreasonable to expect the shrimp revert to their initial strategy of constant grooming. Of course, once reverted, the shrimp lack an internal feedback mechanism, and their antennules would supposedly no longer possess the genuine teleological function of chemical reception.

Now, it should be somewhat clearer what is odd about thinking that the presence or absence of a feedback mechanism suffices for genuine or ersatz teleological function, respectively. In each of the three historical shrimp populations (initially absent a feedback mechanism, with a feedback mechanism, and again without a feedback mechanism), the antennules play exactly the same contributory role in respect to the olfactory system, namely the reception of chemical stimuli. Further, it is because of the value of that contribution to the olfactory system that each historical shrimp population has the grooming strategy it has or comes to alter that strategy. And yet, even though the antennules contribute in exactly the same way in each of the three scenarios and it is exactly because of that contribution that each shrimp population has the grooming strategy that it has, only in the second case is it true that the antennules are genuinely teleological, if the presence or absence of feedback suffices to affirm or deny genuine teleological function, respectively. This is a legitimately odd result.

The oddity, however, reflects the fact that there are two different patterns here; both of which are legitimate targets of explanation. The first is the pattern of behavior exhibited by instances of a system type, namely individual caridean shrimp exhibit a pattern of chemical reception via antennules. That pattern of behavior is explained by characterizing the contribution of the various working parts of the individual shrimp, including, of course, the contribution of grooming by the third maxillipeds. In so doing, we explain how it is that a system type operates to exhibit the relevant behavior. And, of course, that system type in the case of caridean shrimp is insensitive to its success or failure in respect to chemical reception, thereby generating the conclusion that chemical reception is not a genuine teleological function of the antennules. The second sort of pattern is the pattern of antennular chemical reception of the clade of caridean shrimp. The clade, as a group of ancestrally related autocatalytic collectives, exhibits a stable pattern of antennular behavior. This cladistic pattern was appealed to in order to make the case that the shrimp might modify their grooming strategy in the face of changing environmental conditions. In fact, the cladistic pattern seems to present the requisite sensitivity to success and failure; it is, after all, to ensure the success of antennular chemical reception that modifications to grooming behavior are explained. This would seem to give some reason to consider antennular reception genuinely teleological.

I will show, first, that the cladistic pattern does support genuine teleological ascription (so, in each of the three cases above, the antennules possess genuine function), and, then, make a suggestion about how to view that result in relation to that generated by the first pattern of individual system operation.

If "Antennules receive chemical stimuli" is a norm, it is not a norm instantiated in this or that individual shrimp's behavior but is, rather, a norm instantiated by the clade of caridean shrimp. To think that it is a norm, we need reason to think that the fact that exceptions to it are subject to modification or elimination is part of the explanation for the observed pattern. Exceptions to that statement include both those shrimp that do not receive chemical stimuli through their antennules as well as those with degraded performance, i.e., poor reception of chemical stimuli. Olfaction is important to an individual shrimp's success in finding food as well as, importantly, with some species in finding potential mates. Given olfaction's importance to the persistence and reproduction of an individual shrimp, exceptions (including cases of degraded performance) should fare worse than non-exceptions in respect to reproductive fitness. The diminished reproductive fitness of the exceptional case explains the stability of the pattern of antennular chemical reception within the clade of caridean shrimp. So, that exceptions (or, better, the lineages of exceptions) to the statement "Antennules receive chemical stimuli" are subject to elimination and that it is in virtue of their being subject to elimination that the cladistic pattern is as stable as it is provides reason to view "Antennules receive chemical stimuli" as a norm of performance governing the clade of caridean shrimp.

The last might seem too quick, however. This sort of phylogenic case looks awfully close to a case from the previous chapter, namely the short-lived cars of Seattle, that I discounted as a case of a merely apparent norm. All the shrimp, both exceptional and non-exceptional, are subject to fouling by debris and microbial growth. It is that something has gone wrong with the exceptional cases, e.g., a failure to groom, that explains why they have reduced fitness. But, that reduction in fitness is the result of fouling, and fouling acts on all the shrimp, not just the exceptional case. So, it might seem that the appearance of a norm here is merely apparent for the very reason that I had earlier treated the putative norm in the car case as merely apparent.

However, the phylogenic case is importantly different from the earlier car example, and it is that difference that allows us to construe "Antennules receive chemical stimuli" as a norm of performance. While fouling does affect all the shrimp, it is not because of fouling that exceptions are eliminated from the population. Suppose that all the shrimp are exceptions. Caridean shrimp, as a clade, might fare quite poorly, but there is no reason to suppose they will be wholly wiped out. Though less successful than existent shrimp, our hypothetical wholly exceptional population would likely carry on. The fouling of these shrimp, though hampering that hypothetical population, is not causing its elimination. Let us introduce a population of non-exceptional shrimp now to our hypothetical population. These non-exceptional shrimp should quickly take over in respect to the frequency distribution of the overall population. Additionally, one would expect, as the overall population grows and pressure increases on resources, that exceptional lineages would increasingly be eliminated from the population as they just are unable to efficiently acquire resources in a competitive environment. What is eliminating the exceptional lineages? Well, it is not the fouling. What is eliminating the exceptional lineages is the presence of non-exceptional lineages. Importantly, the pressure against the exceptional lineages is there precisely because they are exceptions. This is just what the earlier car case lacked to be counted as a normative regularity. So, we should in this case be able to see "Antennules receive chemical stimuli" as a normative regularity of the clade of caridean shrimp.

But, if the cladistic pattern "Antennules receive chemical stimuli" is a normative regularity, then we should be able to say that reception of chemical stimuli is a genuine teleological function of the antennules of caridean shrimp. How are we to square this with the earlier result that, given the individual pattern of behavior exhibited by the system type, there was no sensitivity to a norm of performance and only, thereby, ersatz function? Here is the suggestion. The failure of the individual shrimp to be sensitive to the norm of performance shows us that the individual shrimp is not acting purposively. That is, it would be wrong to say that this or that individual shrimp acts to receive chemical stimuli. Just as earlier, it was wrong to say of Sorabji's herbs that such herbs acted purposively to ward off the smell. Sorabji's herbs did possess a function, i.e., were for warding off the smell, because they were held to a standard of performance. It is the same, I want to suggest, for our shrimp. The individual shrimp, or better the individual antennule, does not act purposively to receive chemical stimuli. The antennules have a function, nonetheless, exactly because they are held to a standard of performance within the clade.

But, the analogy with Sorabji's herbs seems to carry with it a strange implication. The herbs, though they do not act purposively, possess a purpose, it seems, because of the purposive action of the villagers. Or, alternatively, I said earlier that my individual acts possess a purpose, because I am sensitive to their success or failure; so, again, it seems that the possession of purpose is tied with something else acting purposively. Do we then need to think of the clade as an individual acting purposively via its sensitivity to the success and failure of chemical reception and that by so acting this or that individual antennule possesses a function? No, I do not think that we do. Though some authors[80] have suggested that we understand biological lineages as historical individuals, I do not think that we should be forced to take on that position to see how individual shrimp's antennules can possess function. The possession of function requires being held to a standard, but there is no reason that anything else, the clade, Mother Nature, God, or what

[80] See, for example, Ghiselin (1974 & 1997), Hull (1978), Kluge (1990), and Baum (1998).

have you, be acting purposively for that to take place, because the conditions on a normative regularity just do not presuppose anything like antecedent purposive action.

I do not intend the above to suggest that any bit of contributory language used in respect to biological systems is genuinely teleological. Instead, there is reason to think a phylogenic function is genuinely teleological just when a cladistic pattern is rightly construed as a normative regularity. Such cladistic patterns will be normative just when the fact that exceptions are subject to elimination (in virtue of being exceptions) explains the stability of the pattern. What the above shows is that we should expect this to be the case whenever a selective regime explains the stability of a cladistic pattern. That this is the case has the benefit of vindicating the appeal of etiologically-based realist accounts, because it turns out that such realists are right to look to selective regimes as instantiating norms of performance and, as a result, grounding genuine teleological function.

Clear cases where contributory language fails to be genuinely teleological are those explanations of some environmental or external feature's contribution to system operation. In the second chapter, I suggested that the decompositional/ 'reverse engineering' model will lead one to characterize the contribution of background or external conditions to the system. For example, cyanobacteria as phototrophs utilize photons as a source of free energy, and so in characterizing that system, one could characterize the contribution of sunlight to the operation of cyanobacteria. Or, for a chemotactic organism such as E. Coli, one might describe the chemical gradients formed in liquids as serving to direct E. Coli toward food or away from toxins.[81] Exercises in comparative biology often generate such functional or contributory language of environmental or external features. For example, slime molds live as free-living

[81] See Vogel (1981) for a collection of wide ranging examples along these lines.

terrestrial amoebae, but, when local nutrients are exhausted, they join together with other amoebas to form a slug-structure in order to move to new regions. Free-living aquatic amoebas do not form such structures and lack flagella, so how is it that they are capable of moving to new regions when local resources are exhausted? The answer is just that the currents in the aquatic environment serve the same motility function that the terrestrial amoeba accomplishes by forming collective slugs. (Gross 1994) Though such functional or contributory descriptions of external or environmental features are a useful part of systematic analysis and explanation, they are not genuinely teleological.

Let's take the cyanobacteria example. The sun, we might say, contributes the free energy requisite for the functioning of cyanobacteria. The statement "The sun supplies free energy to cyanobacteria" is not, however, a norm of performance. It is true, of course, that the sun can fail to supply its contribution to the metabolism of cyanobacteria: excessive cloud cover, atmospheric debris from comet impact, or nuclear winter will inhibit or remove the relevant contribution. Is the stability of the pattern captured by "The sun supplies free energy to cyanobacteria" explained by the fact that exceptions to it are acted against in virtue of their failure so as to ensure the stability of that pattern? Well, the answer should clearly be "no". The sun is not subject to modification in virtue of its failure to supply free energy to cyanobacteria. Contrast this with the antennules of the shrimp. Antennules that fail at their olfactory work are subject to correction, because those shrimp with lousy antennules will be eliminated from the population. Of course, if the sun is blotted out for some period of time, cyanobacteria (and most of the biosphere for that matter) will suffer tremendously. The sun, however, will not, and it will not, of course, modify its behavior. "The sun supplies free energy to cyanobacteria" is not, then, a norm of performance, and, since it is not, it would be incorrect to view the contributory language mobilized here as signifying anything genuinely teleological.

We want to be careful not to assume that every contributory description of external features or conditions fails to be genuinely teleological. Niche construction is a pervasive feature in the life histories of organisms. The beaver that builds a dam in order to flood a stream is altering the world to serve some function. Its service of that function is subject to the same sort of considerations that apply to phylogenic traits more generally, and so such constructed niches can serve genuine teleological functions.

II. Genuine ontogenic normativity and teleology

The above applies to phylogenic traits, but we might also distinguish a class of distinctively ontogenic traits. Such traits play a contributory role in the persistence and subsistence of an organism but are not heritable. A clear example would be contributory functions that result from individual learning. For example, a squirrel has discovered how to acquire food from a bird feeder and consequently enjoys enhanced fitness. Such behavior, however, might not be heritable by its offspring, so the contribution of such behavior to the autocatalytic collective that is our squirrel is not phylogenic. Now, learning is itself a teleologically characterized activity, and I will take up learning and how to understand such acquired functions in the next chapter. But, the behavior might not be result of learning but the result of some peculiar or individual accident of fate. Our squirrel, for example, has lost its tail early in life due to a run-in with a hawk, and it is the absence of a tail that allows the squirrel to make its way past the squirrel detection system guarding my bird feeder. The squirrel will be more successful reproductively than its brethren, but the reason for its success is not something that it will pass on to its progeny.

Since I will take up learning in the next chapter, let me just address this latter type of ontogenic contributory effect. The accidental modification case is just the sort of case that Wright had in mind with the accidental screw and that was to supposed to show the importance of etiology to individuate genuine from ersatz function. But, given what has come before, etiology is not the criterion on its own to decide whether the contributory effect is genuinely teleological. What is important is whether the individual shows some sensitivity to the success and failure of that contributory effect. Alternatively, does the organism following the accidental modification act in such a way so as to maintain the contributory effect? With our squirrel, it is unlikely that the contributory effect afforded by the shortness of its tail is a genuine teleological function. First, its tail just is not likely to grow back, so there is no possibility of an exception to the putative norm of performance involved with its contribution. But, second, even supposing the tail did grow back at some point, there is no reason to think that the squirrel would act to chew its tail back down to a nub to maintain the contributory benefit of a short tail. That said, if our squirrel did act in such a way as to keep nibbling down its growing tail due to the benefit of a short tail, then we would have cause to think that, despite its accidental origin, the short tail was serving a genuine teleological function; the squirrel would be demonstrating a sensitivity to the success and failure of the contributory effect of a short tail.

III. Genuine sociogenic normativity and teleology

A further class of biological traits is what I will call "sociogenic traits". By 'sociogenic', I intend a trait distinctive of one or some communities of conspecifics that persists not through reproductive heritability[82] nor is it to be accounted for by variances in local environmental conditions alone. The intent is to recognize that some traits in populations can be heritable without being inherited through genetic or epigenetic factors. Biologists often refer to what I am calling "sociogenic" as "traditions",[83] where a tradition is

> … a behavioral practice that is relatively enduring (i.e., is performed repeatedly over a period of time), that is shared among two or more members of a group, and that depends in part on socially aided learning for its generation in new practitioners. (Fragaszy & Perry 2003, 12)

"Socially aided learning" is not as conceptually packed nor as narrow in its application as one might think. Instead, the intent is just to capture that some traits are heritable in a population without being reproductively so. So, for example, though communities of cricket frogs (*Acris crepitans)* and the northern bobwhite (*Colinus virginianus)* exhibit distinct vocalization dialects, these are not considered sociogenic, because genetic variance accounts for the distinct dialects. (See Ryan & Wilczynki (1991) and Baker & Bailey (1987), respectively.) But, even the absence of genetic or epigenetic explanation is not sufficient to think that a distinctive trait of a community of conspecifics is heritable in some alternative sociogenic manner. Distinctive community traits might just reflect

---

[82] I use the more cagey phrase 'reproductive heritability' versus 'genetic heritability', because it is increasingly clear that the genome is just one locus of reproductive heritability. So, for example, there is a small population of domesticated sheep with the economically desirable trait of enlarged buttocks. Such callipyge sheep, however, are not the result of a genotypic difference but are the result of an epigenetic difference, namely variations in maternally inherited chromatin. (See Gibbs 2003)

[83] I favor 'sociogenic' over 'tradition', because 'tradition' has too frequently lead to the leap that the sociogenic is sufficient for culture and in turn generated needlessly distracting disputes about the latter term ('culture').

ecological considerations. For example, the maneless lion of Tsavo has a radically diminished mane compared to other populations of lion, but that phenotypic difference does not reflect a genotypic difference. Instead, it is the result of a physiological response to sustained exposure to high temperatures. (West & Packer 2002) Alternatively, the presence of algae fishing in some communities of chimpanzee is not considered sociogenic, because its absence in other communities is more easily explained by the rarity of algae in those communities. (Whiten et. al. 1999) The latter sort of case is important, because it stresses that we should not confuse instances of individualistic learning or behavioral acquisition with socially aided learning. The former is not a trait that one acquires by being a member of a community, while the latter is. So, in the explanation of a trait's presence in a population, we have three gross explanation types: reproductive heritability, phenotypic plasticity (including individual rediscovery), and socially-aided learning.

For example, ground finches (*Geospiza difficilis)* of the Wolf and Darwin islands of the Galapagos' Archipelago have the interesting and distinctive behavior of feeding on the blood of live boobies. They land on the tail feathers of masked and red-footed boobies, peck at the base of the booby's feathers, and feed on the blood from the resultant wounds. Such behavior is not found in other communities of *Geospiza difficilis*. It might be the case that this behavior reflects a genetic or epigenetic difference. Alternatively, it might be a case of repeated individualistic learning events. Bowman and Billeb (1965) suggest that, during the dry season, the typical fare of *Geospiza difficilis* becomes scarce but boobies at that time are infested with black flies conspicuous on the booby's white plumage. Initially, the finch might be going after flies but in so doing draws blood and subsequently learns to extract the blood alone. We would have then a trait distinctive to a community, but one that is repeatedly reacquired and not in any inherited by interaction with conspecifics. Lastly, it might a sociogenic trait where the explanation of its prevalence in these populations of finches is something like imitation learning, etc. What fills out that 'etc.' is left intentionally open-ended, where it can be any sort of learning process propagated by social interaction.

So, let's look to some sociogenic cases[84] that exhibit norms of performance. A quick caveat is in order. Sociogenic traits can be socially directed or not. For example, tool-use in chimpanzees, an often hypothesized instance of a sociogenic trait, is not socially directed, whereas variations in chimp hand-clasping is a socially directed trait insofar as it used in recognizing members of a social group, establishing social cohesion, etc. (Whiten et. al. (1999) & Nakamura (2002)) Since sociogenic traits cover both socially directed and non-socially directed behaviors, we should not think that finding a sociogenic norm of performance is to be equated to finding a social norm. (Whatever the latter amounts to, it is at least clear that it is not homologous with a sociogenic norm of performance.)

Norwegian rats (*Rattus norvegicus*) provide a nice example of such sociogenic traits and one with clear experimental data supporting the status of those traits as sociogenic. Norwegian rats are an extremely successful mammal with a range covering much of the globe. Much of that success is attributed to their ability to adapt foraging preference and method to a range of ecological conditions. (Galef 2003) Despite the

[84] The most celebrated cases of such sociogenic traits are reflected in the extensive work on chimpanzee "culture". (See Whiten et. al. (1999) and Whiten (2005) for reviews of that work over the last century.) I avoid such cases below, because they often seem to lead to the overly quick inference that something like advanced cognition is required for such traditions to get established.

plasticity in foraging preferences and methods, interestingly wild rats are extremely hesitant to ingest novel food sources, and, once they have tried some novel food, they are reluctant to eat anything else. (Barnett 1958, Galef 1970, & Galef and Clark 1971). The variations in food preferences and methods that make the rat so successful are not the result of genetic or epigenetic variations among wild rats. Nor are those variations accounted for by various ecological considerations alone. For example, some communities of rats living along the Po River of northern Italy dive and feed on mollusks, while other nearby communities with equal access to the same resource do not. (Gandolfi & Parisi 1973) Additionally, Galef and Allen (1995) have shown that rat colonies can be trained to eat one of two equally palatable and available food sources and, once trained, the rats continue with the trained food preference despite the availability of the alternative food source. So, the mere presence or absence of ecological considerations will not explain food preference (in contrast with the rarity of algae in the case of chimpanzee algae fishing). Moreover, the resistance to novel food sources limits the likelihood that communities of rats with the same food preference are an example of individual rediscovery of the food source (as might be the case with the vampiric finches above).

Instead, a combination of mechanisms serves to establish, for example, community specific food preferences. Prenatal and nursing pups are exposed to food odors in the amniotic fluid and breast milk, respectively, and consequently develop strong food preferences for the associated food. Adult rats will develop food preferences associated with the food particles discovered on the bodies of other rats or with their odors when associated with rat breath. Neither of these latter mechanisms are cases of imitation learning. Rats will develop food preferences even if the food particles are found on the bodies of dead rats and even if they are just exposed to the chemical signature of rat breath in correlation with the odor of a food source. (Galef 2003) Importantly, these social mechanisms for the transmission of food preferences establish persistent preferences in communities. It has been shown that food preferences will last even after an originally trained population of rats is long since gone. (Galef and Allen 1995).

Assume a community of rats foraging on thistle seeds and that food preference is rightly sociogenic. Let's see if we can construe the statement "Rats forage only on thistle seeds" as that of a norm. The relevant functional ascription in this case is the contributory effect of thistle seed foraging to the subsistence and persistence of the community of rats. So, if there is a norm of performance in respect to that, then it would need to be the case that exceptional behavior, namely foraging activity that resulted in the ingestion of something other than thistle seeds, was subject to modification such that the stable pattern of thistle seed foraging was maintained. Suppose a plant source reasonably close to the thistle such that the sociogenic mechanism leading to the consumption of thistle seeds has some likelihood of leading our rats to consume this alternative food source.

Initially, assume that our alternative seeds are just as nutritious as thistle seeds and assume that, nonetheless, the community of rats remains almost exclusively thistle seed consumers. In such a case, it would be reasonable to reject that there is any specific sociogenic norm of performance in the community governing thistle seed consumption. Even if our rats are lead "accidentally" to consume the alternative seeds, no negative repercussions issue from such exceptional behavior. As such, it would be wrong to think that the contributory description of foraging as for the acquisition of thistle seeds was literally true (though it might be appropriate to think that such foraging was literally for the acquisition of seeds generally; that would, of course, depend on whether there was some more general norm of performance).

Let's make our alternative seeds somewhat toxic; they need not be directly lethal. The sociogenic mechanisms that get our rats to thistle seeds in the first place are not sensitive enough to keep our rats from eating other food sources, such as our alternative seeds. Previously absent toxicity in the alternative seeds, if the pattern of thistle seed foraging was more stable than ought to be given the strength of the sociogenic mechanisms, then this is likely the result of chance. However, the presence of toxic seeds alters the scenario. The stability of thistle seed foraging is explained in part by the negative repercussions of eating toxic seeds. The sociogenic mechanisms do not suffice to explain why rats forage on thistle seeds over alternatives, but the fact that those that eat alternatives suffer reduced relative reproductive fitness does explain why the pattern of foraging thistle seeds is as stable as it is. Exceptional behavior (eating the alternative seeds) is subject to elimination in virtue of being an exception and by being so subject explains the stability of the pattern. So, we could say in such a case that there is a standard of performance in respect to such foraging behavior, namely that the rats ought to forage for thistle seed, and so say that the particular foraging activity is literally for the acquisition of thistle seeds.

We should be done. Phylogeny, ontogeny, and sociogeny are all capable of exhibiting patterns rightly construed as normative regularities, and, as a result, it is right to think that phylogenic, ontogenic, and sociogenic traits can be genuinely teleological. There is, however, a particular distinctively teleological activity not captured by what has come so far. That activity is learning. The characterization of an acquisitional process as learning requires a species of the normative, namely the ideal, that our previous field guide and the comments in this chapter just fail to capture. In the next chapter, I will show how one might extend the earlier field guide to include this species of the normative and show, given this extension, how we can understand whole lineages and individuals as acquiring by learning. We will then be in a position to return to the question of the mind and its place in nature and see how the possibility of literal teleological ascription, underwritten by the presence of normative regularities, can be put to work in resolving that question.

Chapter 7

THE FIELD GUIDE EXTENDED: THE IDEAL AND PROGRESSIVE ACQUISITION OF FUNCTION

Assuming the preceding has been persuasive, phylogenic, ontogenic, and sociogenic traits of biological systems can possess genuine teleological functions. However, the possibility of genuine nonartifactual learning (i.e., the possibility that a system is capable of learning without presupposing intentional and psychological states) has not yet been gained from the previous considerations. In short, what is required for the process of learning to be genuinely for the acquisition of some skill or function is the ideal, a species of the normative. What has come before just will not suffice to capture that species. The aim here, therefore, will be to extend the previous field guide to cover the ideal and subsequently show how and when the acquisition of phylogenic and ontogenic traits can be construed as instances of genuine nonartifactual learning.

I. The problem of progressive acquisition for the earlier field guide

To describe a process as acquisitional is to provide a teleological characterization of that process. It is simply to describe the process in virtue of its terminus or end.

"The wall is developing a crack."

"The fetus is developing lungs."

"The lion is growing a mane."

"The neurons are forming a network."

"The infant is learning to walk."

Given the end of the process, we can at times further describe the contribution of the composite stages of such a process to the production of the end. So, not only is it the case that the whole of the process is teleologically characterized, but various component stages of the process are also teleologically characterized. And, as before, the explanatory value of such contributory descriptions need not involve the commitment that the process or its component stages literally serve some function. The expansion of ice in the crack is contributing to the formation of the crack in the wall, but we need not think that the expanding ice has that as a function.

Clear cases of acquisitional processes that are genuinely teleological are, I think, developmental processes in organisms. Take, for example, the process of cardiogenesis. That process, at least in zebrafish, is the following: some undifferentiated fetal cells differentiate into a mass of endothelial cells; that mass forms a primitive contractile vessel; the subsequent pumping by that vessel induces shearing[85] by blood flow over the composite endothelial cells; that shearing force affects gene expression in the endothelial cells and produces vortexes strong enough to rearrange the cytoskeleton of those cells, thereby altering the shape of the contractile vessel; the shape of the contractile vessel continues to alter until shearing force is minimized; at that point, the zebrafish has achieved a mature heart. (See Hove et. al. 2003 & Summers 2005) Within zebrafish, there is a stable pattern of sequential events, i.e., cardiogenesis, that terminate in the acquisition of a mature heart. The stability of that pattern, as with the phylogenic traits in the previous chapter, will be explained in part by the fact that those zebrafish that are

[85] The shearing force of a fluid is the force exerted by that fluid flowing over a solid object.

exceptions to that pattern (i.e., those do not reach the proper terminus of the process) will be acted against in virtue of failing to generate a well-formed heart. Cardiogenesis can genuinely possess the function of generating hearts, because, just as before, a normative regularity can be in place governing the process via a selective regime. Further, the contribution of each stage of cardiogenesis looks to be a plausible candidate for genuine function. For example, the contractile vessel pumps in order to produce a shearing force over the composite endothelial cells. Here too, we have a pattern of contractile vessel behavior in zebrafish, and the stability of that pattern, given the value of cardiogenesis, is likely to be explained by the fact that exceptions (e.g., contractile vessels producing insufficient or too much shearing force) will be acted against in virtue of being exceptions. (If the force is altered, then a malformed heart will be generated. This sets up the possibility of normative regularity by a way of a selective regime.)

A class of acquisitional processes is not captured by this model of phylogenic functions, however. Such acquisitional processes are those, I think, that one would normally associate with learning. The category or genus of learning contains a varied assortment of species from simple passive forms, such as stimulus-response learning, to more active complex forms, such as intentionally setting out to acquire a skill or talent. I do not want here to sort out what is properly in that somewhat mongrel category from what might properly be within the category of development and growth. What I do want to notice is that the sort of processes typically called “learning” will not suffer the previous phylogenic treatment via selective regimes.

The bear learning how to bust open a log for grubs, the child learning to walk, or the squirrel learning how to get into my birdfeeder are all in the process of acquiring a skill or function. Those processes are teleologically characterized, and such processes look to be plausible candidates to be genuinely teleological. The bear in learning how to bust open logs and the child learning to walk respond to their successes and failures in respect to the end. That the log failed to break when it was merely rolled explains why the bear adopts some new strategy, or that ripping at the log shattered it and brought forth a meal explains why the bear continues to engage in that behavior in the future. Or, the fall resulting when the child leaned too far forward is part of the explanation for why the child subsequently balances over its hips. The activity of the bear or the child look to be plausible candidates to genuinely possess the functions, respectively, of learning to bust open logs or to walk, because the bear and the child are sensitive to the end of the respective process. In contrast, the individual zebrafish in process of developing a heart is insensitive to its success or failure. (Add an obstacle to blood flow, and the contractile vessel does not compensate for the diminished shearing force. Instead, such an obstacle leads to a deformed heart.) Typically learning involves an ontogenic sensitivity to the end of the process, whereas development and growth typically involve a phylogenic sensitivity to the end. For example, think of the vampiric finch from the previous chapter. Excluding a sociogenic explanation, the process of acquiring the behavior of blood-drinking could either reflect some facts about reproductive heritability or about individual rediscovery. Even though within the community of vampiric finches there is a stable pattern of acquiring the blood-drinking behavior, it is an open question whether the stability of that pattern reflects the sensitivity of the phylum or of the individual. In the former, the individual finch develops the behavior but is individually insensitive (as the zebrafish is) to its successes and failures, whereas in the latter the individual finch learns the behavior through its sensitivity to its successes and failures at acquiring an additional nutritional source.

Individual rediscovery of vampirism by a finch as well as other typical instances of ontogenic learning present a problem for the earlier offered model, if such ontogenic learning is genuinely teleological. The learning of the bear, the child, and the finch, if genuinely teleological, reflect an individual sensitivity to the end of the acquisitional process – that is, their sensitivity to their successes and failures in respect to the end is part of the explanation for trajectory of the process. The end of that process provides a standard in virtue of their early attempts to break logs, walk, and drink blood are relative failures, and it would seem that the fact that those early attempts are failures is part of the explanation for why each continue to move on in progressing toward the end. (Of course, some of these attempts along the way are relative successes to other attempts, and that fact as well would be important to the explanation for why certain behavioral aspects of those early attempts are retained.) As these attempts in the process of learning are failures relative to the end, no behavioral instance in the process of learning is an instance of the standard comprising the end. However, our earlier field guide required for the identification of a norm that there be instances of the norm. We described some stable pattern in the world. If exceptions were subject to modification (in virtue of being exceptions) so as to correspond to that description, that description could be understood as the statement of a norm. Yet, that statement of a norm rests on the description of observed instances of the norm, and these sorts of typical ontogenic learning cases will not provide any instances of the norm. So, the earlier method of identifying a norm in the world will not do.

I will call this sort of norm "an ideal", where an ideal is just an existent rule or standard for which there are no existing instances.[86] Clear cases of such ideals are those where one intentionally sets out to acquire a skill or capacity. For example, I recently set out to acquire the skill of a slapshot. In setting out, I know what a slapshot should look like, namely that a properly made slap shot should produce a high-velocity, elevated shot on goal by a wide swing of the hockey stick. As I am learning, I have not yet produced, at least a replicable, instance of that ideal. I can judge of my various attempts at producing a slapshot which are more or less relative successes in respect to that ideal. I seek to retain those behavioral features (e.g., weight over the skates, position of the stick, etc.) that are part of those relative successes, while I seek to eliminate those features that are part of the relative failures. That process is genuinely for the acquisition of the skill of a slapshot, because the explanation of the sequence of events comprising that process relies on my sensitivity to the end. It would be implausible, however, to think every case of individualist learning (e.g., the finch learning to drink blood) involved some representation of the goal of learning. Consequently, if nonartifactual learning is genuinely learning – that is, a process genuinely for the acquisition of a skill or function, then there should be nonartifactual ideals to which the relevant entity is sensitive. In the subsequent section, I will extend the previous field guide in such a way to accommodate this further species of the normative.

Before turning to that, I want to address a possible source of confusion. In the example of the vampiric finch, the juvenile finch learning how to drink blood from

[86] Since I am interested in those standards that do exist, I am setting aside one sense of "ideal", namely that sense in which we recognize that there is not a standard but that there ought to be one. For example, in response to some offense, the offended exclaims, "there ought to be a law." Such a law is itself an ideal in that it would be ideal if there were such a law. That is not the sort of ideal that I am interested in here, however.

boobies is surrounded by a community of adults that have that skill. Those adults provide, it seems, instances of the standard toward which the juvenile finch is aiming in learning. If the community of finches exhibit instances of the standard, then that would seem to show that there isn't and doesn't need to be an ideal in place. That said, that there is a community of adult finches with the skill of blood drinking does not show that there is anything like a community-wide norm governing the acquisition of vampirism in this or that individual finch.

Initially, assume that our juvenile finch is insensitive to its successes and failures in acquiring blood from boobies in just the way that the individual zebrafish is insensitive to the success or failure of the process of cardiogenesis. (In such a case, I suspect that we would not say that the juvenile finch is learning but would say rather that it is developing. But, I do not want to hang anything on that.) We do have a stable pattern of acquisition of this skill in the community as a whole. The stability of that pattern, if there is a selective regime in place for blood drinking, would be explained by the sensitivity of the clade to both the behavior and the process leading to its acquisition. In that sort of case, it would be right to think that there was some community-wide norm governing vampiric behavior and the acquisition of vampirism. Now, assume that our juvenile finch is sensitive to its successes and failures in acquiring blood from boobies. If there was a sociogenic mechanism in place, such as imitation learning, then here too it would be right to think that there was some community-wide norm governing the juvenile's acquisition of vampirism. As with the phylogenic case, the correction that would take place in respect to the individual finch's behavior would be the result of its behavior failing to be like the instances of the norm exhibited in the community as a whole. Lastly, assume that our finch is sensitive to its successes and failures, but there is no sociogenic mechanism in place. So, when it is learning to drink blood, it is not doing so through imitation learning. In that sort of case, that there are instances in the population of what the juvenile is aiming for is irrelevant to what it does in acquiring the skill. Whatever it is that explains the juvenile's behavioral transformations in acquiring vampirism, it will not be that others possess the skill. Its sensitivity to the end of that process, by hypothesis, is not a sensitivity tied to the behavior of others. The adult vampiric finches are not providing examples of the end for the juvenile. In respect to that individual finch and its behavior, there are just no instances of the norm while it is learning. If the juvenile finch is sensitive to the end of the acquisitional process, that end provides the ideal to which it is aiming.

II. The field guide extended: the ideal

In order to account for the ideal, the field guide needs to be extended to include cases where there *is* a standard and yet there have not been instances of it. In extending the field guide, I will continue to maintain that the distinction between a statement of fact and that of a norm is in respect to practical consequences. Previously, we had reason to report a norm when we observed a pattern and observed that exceptions, by being subject to modification, were part of the explanation of the stability of that pattern. That will not do, however, without instances of the norm. We need a different procedure, then, to recognize an ideal in the field, and that is what I will develop below.

To give some intuitive sense to a standard existing without instances, I will begin with two stories in which, I think, an ideal plays an explanatory role. From these, I will develop a more general suggestion for how to recognize an ideal in the field.

Here is the first. Fred wants to build a water pump to get water from a nearby pond to his crops. Fred sets to work and comes up with a series of prototype pumps. The first few fail to pump water at all, but he does come up with one with some pumping action. Until Fred gets a chance to do further work, he puts this early water pump to work. As free time presents itself, he continues to tinker away. Some of what he comes up with is no advance over his earlier operable prototype, but, when he does come up with a pump better at drawing water, he adopts that newer variant. What we observe over time is that some of the results of Fred's efforts never make it out of his toolshed as well as that others are adopted for some period before being replaced with newer and better models.

In this kind of case, an ideal of sorts is in play. If we want to explain why it is that some variants find their way to the field, why others do not, as well as why some variants replace others, then we might appeal to some standard of water pumps (e.g., maximizing pumping at minimal energy cost) in virtue of which Fred judges the varied products of his efforts. None are true instances of the standard, because each pump is subject to further replacement by a still better model. Yet, the replacement of one variant for another is explained by its better approximation to the ideal. Alternatively, why some variant does not make it out of the toolshed is explained by its failure to be a better approximation of the ideal relative to existing variants.

Like before, a standard is in play because of the practical consequences which ensue. Fred judges some variant to be a more or less better approximation of the standard and adopts or rejects that variant in light of his evaluation. What is subject to modification is not this or that water pump, because newer and better water pumps replace older and worse units; the individual units are not modified. What is subject to modification is what is used in the field to pump water. Different variants on the water pump will be used in the field, and the ideal that Fred has in mind explains why what is used in the field changes over time.

An ideal plausibly has an explanatory role to play above, because Fred has some standard in mind that informs his adoption and rejection of certain variants. I want to tell one further story that, I think, plausibly involves an explanatory ideal, even though no one has that ideal in mind.

Over several centuries in the life of some isolated village, assume some 18 variants of the wheelbarrow have been produced at some time or another. Only three of these variants enjoy widespread use during this period. Further, these three enjoy widespread use in distinct, but partially overlapping, contiguous time periods.

One target of explanation is why this or that variant came into existence. An alternative target of explanation is why this or that variant persisted, replaced another, or never came to be widely adopted. This latter target of explanation need not attend to the various reasons why some variant or another came into existence. For example, with Fred and his water pumps, the explanatory role of the ideal was in respect to why some water pump or another was or was not adopted. It needed to play no role in explaining how some water pump came into existence. Fred, even when madly at work in his toolshed trying to build a better water pump, could still produce one by luck alone. What was relevant for the ideal to play its explanatory work was that, once a water pump came into existence (however it did so), Fred evaluated its promise as a water pump and adopted or rejected it based on that evaluation. Similarly, whether each and every new wheelbarrow comes into being via the explicit intentions of its maker or via mere accident or blind luck is irrelevant to the explanation of its further adoption or rejection by our villagers. So, let's just set explicit intention aside. Assume that our villagers are a fairly uncreative lot who never strive to build a new type of wheelbarrow, and so, each new variant is the result of some happenstance or another.

There are any number of reasons why we might have a history of some variants faring better than others in respect to their adoption and relative persistence in use. One of those is that those wheelbarrows with widespread use are just better at being wheelbarrows than their temporal peers. The history of wheelbarrows in our village might be explained by appeal to some standard of wheelbarrows. Such an explanation would seem appropriate and plausible insofar as our villagers, though an uncreative lot, judge and evaluate which wheelbarrows to use in virtue of that standard.

However, I want to disallow that the villagers evaluate the quality of wheelbarrows and see whether an ideal can rightly, nonetheless, play an explanatory role. To remove such evaluatory judgments from the equation, all we need to suppose is that our villagers do not have a choice in respect to which wheelbarrow to use. For example, we might add to our story the following: there is no artisan in the village making wheelbarrows; the temporal and material expenditure to produce a wheelbarrow limits individual villagers to at most a single wheelbarrow at a time; and, villagers, in building their own wheelbarrows, only tend to copy the design passed down from one of their parents. (Given the last, the origination of wheelbarrow variants might be explained as mistakes in the copying process.) So, our villagers cannot shop around for a better wheelbarrow; they cannot produce a variety of wheelbarrows and test, as Fred did, which is best; nor can they look around to see how their neighbors are faring for a better wheelbarrow design.

In such a case, what might prompt us to think that the history of wheelbarrows in the village can be explained by appeal to an ideal? Well, let's add a few more facts to the story before answering that question: wheelbarrow types vary in respect to the ratio of carrying load to energy expenditure; those types that maximize carrying load relative to energy expenditure make the various other projects in which our villagers engage easier relative to those wheelbarrows that do not; these auxiliary projects are relatively important to the livelihood of our villagers, e.g., bringing crops in from the field, removing debris from the field, etc.; and so, those villagers with wheelbarrows that maximize carrying load relative to energy expenditure (relative to other existing variants) will fare better in their lives compared to their less lucky peers. Since wheelbarrow design is copied from one's parents, those villagers, whose parents were more successful than their peers, are likely to fare better than those with less successful parents.

Assume the following as a putative norm and ideal: "In our village, a wheelbarrow maximizes carrying load while minimizing energy expenditure." As before, we want to see that exceptions to that statement are subject to modification so as to realize or better approximate it. Individual wheelbarrows, given the story, are not subject to modification; what is, however, subject to modification over time is use of this or that wheelbarrow type by the population. It is the pattern of use over time that is our explanatory target, because we want to explain why it is that some variants came to be adopted, rejected, or replaced by other variants in the history of the village.

Call our three dominant variants of wheelbarrows "α", "β", and "χ". Assume that their respective periods of dominance are ordered: α, β, χ. Take the initial time period when α's dominate. During that period and limited to the other variants in existence, α's are as close as we get to an instance of the standard. In fact, given the variants in existence, we can treat α's as if they are instances of the standard. During that period, what explains α's dominance in use is that exceptions to the standard, i.e. non-α's, are subject to reduced or eliminating use over time. They are so subject to reduced or eliminating use, because such wheelbarrows reduce the prospects of their owners and their owners' descendents relative to those α-owners and their descendents. We have during that period just the conditions required to report a norm. The emergence of β in the village alters the situation. Now, β's are the closest approximation to the hypothesized standard, and, relative to β's, α's are now exceptions to that standard. The increasing dominance of β's over α's and other variants is again to be explained by reference to the standard. Exceptions to that standard, i.e., non-β's, are subject to modification in respect to use within the population such that wheelbarrow use comes in line with that standard. The emergence of χ's will again cause a transition in dominance, and like the transition to β's, that transition will be explained in part by the standard on wheelbarrows.

Internal to each time period, a recognizable norm is in play: "In our village, a wheelbarrow maximizes carrying load while minimizing energy expenditure." Across those three periods, it is the same norm in play, and it is that norm which explains why the pattern of use alters in the way that it does over time. Like with Fred's water pumps, what explains why some variants come to be adopted or rejected over time is that they are relatively better or worse approximations to the ideal wheelbarrow. That each dominant variant is itself subject to elimination with the emergence of a still closer approximation to the standard governing wheelbarrows provides the sense in which this standard is an ideal.

Notice that the above suggestion for how we might recognize an ideal works just as well when our villagers are explicitly judging and evaluating wheelbarrows. What has been altered in removing the villagers' explicit evaluations of wheelbarrows? Well, given the additional facts, all that has been altered is how the ideal carries out its work in the world. We have in effect offloaded the cognitive operation from the minds of individual members of the population to the environment. The environment carries that workload just because it happens to be so structured to do so; no further fancy metaphysical thesis need be entailed. However, if the environment is to carry that workload, it does need to be the case that there is an approximate instance of the standard throughout. That approximate instance serves as an in fact representative of the standard in place of the mental representation of the standard within judgment. That is, instead of a villager judging in light his represented ideal that some variant is a worse wheelbarrow, it is in virtue of the presence of a better representative wheelbarrow that those worse wheelbarrows are acted against.

Let's generalize the above. It is right to report an ideal when and only when the trajectory of pattern development is in part explained by the fact that exceptions to the ideal are subject to modification so as to correspond to it. Like above, we might have some initial pattern that we have reason to believe is a normative regularity. Over time we might see some further pattern develop that is also rightly described as a normative regularity. When and only when the norm in play in both patterns is one and the same and when and only when that same norm explains the transition from one pattern to the next do we have reason to report an ideal. It is the ideal that explains in part the trajectory of pattern development, because it explains why one pattern should transform into another over time. And importantly, given the ideal, those transitions can be explained as literally improvements with respect to that ideal.

III. Phylogenic ideals and progress

The motivation to look for the nonartifactual ideal was to account for the ontogenic acquisition of function as typified by non-imitative learning. However, the wheelbarrow case above should suggest that the applicability of the ideal in explaining the progressive acquisition of function is wider than that. Since the villagers in the wheelbarrow case are, I assume, human, it would be wrong to explain the production of wheelbarrows by appeal to reproductive heritability. Children copying wheelbarrow design from parents is, rather, likely to be a sociogenic mechanism, just as the rats from the previous chapter copied the feeding behavior of other rats in their community. What we are observing in the wheelbarrow case looks to be social evolution and within that social setting the progressive acquisition of function. Now the wheelbarrow case is just make believe on my part, and it is was designed to present the appearance of progressive acquisition of function. So, I make no claim that human social structures do progressively develop in that way. That said, the wheelbarrow case does demonstrate that it is plausible to think that whole groups – either social or cladistic – can progressively acquire function in the way typified by individualistic learning.

I want to take up the last point just in respect to phylogeny. It is, of course, a mistake to think that evolution as a whole is progressive. The reversibility of trait polarity[87] or genetic drift are both easy counterexamples to the thought that evolution as a whole is a progressive process. That said, certain tracts of evolutionary history in which there is the gradualistic acquisition of function are, I think, rightly construed as progressive. That is, such periods of phylogenic acquisition can be understood as literally for the purpose of acquiring some phylogenic function or trait. They are so, because the clade is sensitive to the end of that acquisitional process. In such cases, assuming the conditions for a nonartifactual ideal are met, it is plausible to view the clade as a whole as genuinely learning.

Take, for example, the fennec fox, a desert-dwelling fox of northern Africa with a number of desert-oriented adaptations. One of these adaptations is its overly large ears relative to other foxes. The increased size of the ear contributes to the fennec fox's capacity to cool itself in the scorching deserts where it makes its home. (See Lariviere 2002 & Sheldon 1992) It is not unreasonable to suppose that the history of the fennec fox in northern Africa (from the non-desert adapted, progenitor fox entering northern Africa in the late Pleistocene to the present population) is one of a series of gradualistic increases in ear size and, consequently, a series of gradualistic improvements in the thermoregulatory function of the ear.

[87] See Brandon (1990, 171-4) for some nice examples of the reversal of trait polarity.

The thermoregulatory function of the progenitor fox's ear was suboptimal. That is, when the progenitor fox entered northern Africa, there was a possible phylogenic trait, namely an increased ear, that would have maximized the benefits of the thermoregulatory function relative to other costs. That optimal phenotype was possible in that there was an available sequence of reproductive events that would take the progenitor fox to a descendent fox with the genomic structure that would produce the overly large ear. That there was, say, some initial gradualistic improvement in ear size available to the progenitor fox population did not cause the progenitor fox population to start producing foxes with that improvement. Such a gradualistic improvement was available given the facts about the genomic structure of the progenitor fox population and its reproductive strategy. The explanation why the improvement came into existence reflects those facts alone. Once, however, such a gradualistic improvement comes into existence, the increased fitness afforded by the improved thermoregulatory function does explain why the population distribution of the fox shifts over time in the direction of the improvement. Further, once the initially improved fox has come into existence, there is from that fox a further available improvement in function. That is, the facts about the newer fox's genomic structure and reproductive strategy make available through a reproductive event a fox with a genomic structure that would produce a further still improved ear. Again, once a further improved fox came into existence, the subsequent shift in the population distribution toward that improvement would be explained by the fact that the further improvement is an improvement. We can have, then, a series of gradualistic improvements from the progenitor fox to the present population and can explain the series of transitions in respect to dominant phenotypes by the increasing improvement in the thermoregulatory function of the ear. During that history, there would have likely been at certain stages the emergence of foxes with diminished or less optimal ears relative to their temporal peers. The explanation for why these foxes do not become dominant or generate a shift in population distribution will reflect the fact that they are relatively functionally worse off than their contemporaries in the population.

We could have within the history of the fennec fox a series of transformations over time leading to the present phylogenic trait, and that series of transformations in the dominant phenotype would be explained by improvements in the thermoregulatory function of the ear. That presents at least the appearance of the progressive acquisition of function. From what we know so far, however, we cannot say that the process of acquiring the present phylogenic trait was genuinely progress. That is, what we cannot yet say is that the process was for the acquisition of the present phylogenic trait. To say the latter, we would need some reason to think that the fennec fox through its history was sensitive to the end of that process, and we do not yet have that.

Though the sequence of shifts in dominant phenotypes is explained by the increased functionality of the ear, that could be true when the history of the fennec fox reflected genetic drift alone. For example, take some intermediate stage in the history of the fennec fox with the recent emergence of a fox with an enhanced thermoregulatory function relative to its contemporaries. That enhanced thermoregulatory function by hypothesis increases the fitness of this novel fox relative to its contemporaries. Given that this novel type of fox has a higher fitness than its contemporaries, we would expect to see the frequency distribution to shift over subsequent generations in the direction of the novel fox type. This would be true whether there was a selective regime in place or just genetic drift. Whether selection or genetic drift were involved, it would be true that the frequency distribution shifted over time due to the higher fitness of the novel type of fox. If the history was one of genetic drift, it would not be true, however, that the quantity of members of the older type at some generational point is to due the presence of the novel type in the previous generation. Even if there were no novel type foxes around, the older type foxes would have produced the same number of descendents as they did when the novel type was present in the previous generation. In the genetic drift case, the presence or absence of the novel type is irrelevant to the explanation for why there is a certain number of the older type at a particular generational point. That quantity of older type foxes at some generational point reflects facts about the fitness value of the older type foxes. And, in the genetic drift case, the fitness value of those older type foxes is not affected by the presence or absence of the novel type.

If the history of the fennec fox was one of genetic drift, then it would present the mere appearance of progress. Whether we are considering an ideal or a non-ideal norm, the basic feature that being an exception is consequential is the same. Assuming that the older type fox is a relative exception to a putative ideal in comparison to the novel type, it will not be true in the genetic drift case that being an older type fox is consequential. That is, it will not be true that the older type fox is acted against in virtue of the fact that it is a relative exception to the putative ideal. The population of older type foxes is what it is irrespective of the presence of the more approximate instances of the putative ideal. It is not, then, in virtue of the fact that exceptions are acted against that the transition toward the novel type takes place. If the history of the fennec fox was one of genetic drift, we would not have the grounds for a normative regularity (ideal or otherwise) and, consequently, ought to consider the apparent acquisitional progress of the fennec fox as merely apparent. Alternatively, we ought not to think that the acquisitional process of the fennec fox was for the purpose of gaining the present phylogenic trait.

In contrast, if the history of the fennec fox was a selectional history, then I think that it would be right to say that the historical process was literally for the acquisition of the present phylogenic trait. That historical acquisitional process would be for the purpose of generating the present phylogenic trait, because the clade, assuming a selective regime in place, would be sensitive to the end of that process. Unlike genetic drift, if selection on the ear's thermoregulatory function is place, the fitness value of the older type fox is affected by the presence of the novel type fox; so, the quantity of the older type foxes at some generational point reflects the fact that there were novel type foxes in the previous generation. (The transformation from the genetic drift case to selection just requires that there is some competition for resources among fox types as well as that the likelihood of being a successful competitor reflects the difference in thermoregulatory function.) When selection is in place, simply we have the wheelbarrow case with foxes. The progenitor fox provides a normative regulatory governing the thermoregulatory function of the ear. However, that fox is functionally suboptimal, because further functional improvements on that ear are available. When they emerge, the transition or shift in the dominant phenotype is explained by the fact that the novel phenotype is literally better at the function. It is literally better, because the fact that it is a closer approximate to the ideal or optimal functional type explains why relatively less optimal types within the population are acted against. As such, the fennec fox, as a clade, is sensitive to the ideal or optimal phenotype, and the sequence of transitions in its history would be explained by appeal to that ideal. The fennec fox, as a clade, would be genuinely learning how to improve its thermoregulatory function.

IV. Ontogenic ideals and learning

Let's return to ontogeny. Ontogeny seems to present acquisitional processes in which the individual is sensitive to the end of those processes. Such sensitivity need not require anything like the explicit representation of the goal: that the individual finch might be learning vampirism does not require that the finch is representing the goal of vampirism. (The theoretical value of stimulus-response models of learning lies partially in the fact that they provide a model of learning not dependent on representation.) Further, such ontogenic acquisition need not be imitation but can, instead, be a discovery or rediscovery on the part of the individual. Without a representation of the end or a causally relevant instantiation of the end in the behavior of others, such ontogenic sensitivity to the end required a normative ideal. Let me show then how we can have such an ideal in these sorts of ontogenic cases.

The key to get the previous examples of the wheelbarrow and the fennec fox to apply in the ontogenic case is to replace differential fitness with differential amplification. The differential fitness of the members of a population explained, whether within a selective regime or genetic drift, subsequent shifts in frequency distribution. Switching to the ontogenic case, we do not have populations of replicating individuals. Instead, we have a single individual, our learner, and the behavior in which the individual engages. Whereas fitness is roughly the probability of organismic reproduction, amplification is roughly the probability that a tokened behavior will be subsequently repeated by the individual. Different behavioral types can be assigned differing amplificatory values, just as we can assign different fitness values to differing phenotypes. Differential amplification can explain then the frequency distribution of tokens of differing behavioral types for an individual in just the way that differential fitness can explain the frequency distribution of individual organisms of differing phenotypes.[88] Taking advantage of differential amplification, the previous case, say, of the fennec fox is fairly easy to map to ontogeny.

In mapping it to ontogeny, the aim is to offer an explanation for a pattern of behavior in the organism. Any organism will be engaged in a wide range of various behaviors for very different purposes. The entirety of behavior exhibited by the organism is not the target of explanation. Instead, the target of explanation is a range of behavior that the organism appears to use for some purpose. Since the plausible candidate for genuine *nonartifactual* learning will be stimulus-response learning, the range of relevant behavior will be that exhibited by an organism in the face of a particular stimulus. Further, since the explanatory interest is the acquisition of a function or skill, that behavioral range exhibited in response to some stimulus should make some contribution to the economy of the organism. The description of that contribution both provides that the behavior within the range all falls under some general functional type as well as provides the putative norm or ideal.

It is reasonable to assume that an organism that has learned some new skill or function will have learned that skill or function over time and could have done so through

[88] I take the suggestion that differential amplification with respect to behavior can do the same work that differential fitness does with respect to populations of organisms from Edelman (1987 & 1992). Edelman uses differential amplification as the backbone for his suggestions concerning acquired immunity as well as neural group selection.

some process of gradualistic improvement. The beginning learner could be, then, like the progenitor fox. The beginning learner will respond to a stimulus with a behavioral type that falls within a contributory description, just as the progenitor fox used its ear for thermoregulation. Like the progenitor fox, the beginning learner's response is a suboptimal response to that stimulus. It is suboptimal, because there is a possible behavioral type in response to that stimulus with increased amplificatory value. With the progenitor fox, there was a possible trait with increased fitness value, because that trait was available to it through a series of reproductive events. That another behavioral type is possible for our beginning learner can reflect something similar. That optimal behavioral type is available to our learner insofar as a series of possible behavioral transformations terminate in the optimal behavioral type. The transformations to the fox reflected transformations to its genomic structure, and the possibility of those transformations reflected both its initial genomic structure and reproductive strategy. Alternatively, those facts reflect an instability in the copying of the ancestral genome to descendents, and that instability makes room for the possibility of descendents with a differed genome. For the beginning learner, the analogue of organismic reproduction is behavioral reproduction – that is, the ability to reproduce behavioral tokens of the same type. It would suffice to get the ontogenic model up and running if behavioral reproduction was instable in the way that organismic reproduction is – that is, the reproduction of some behavioral type is sufficiently instable that it can on occasion lead to the production of a behavior under another type. Further, as with organismic reproduction, we can think that a new behavioral type is replicable, and that the emergence of a new behavioral type (in virtue of the instability of behavioral reproduction) can lead to the emergence of still further new behavioral types.

The learning individual will exhibit over time various behavioral responses to some stimulus. We can ask for an explanation for why this or that behavioral type first emerged or was generated by our individual. But, in respect to learning, why this or that behavioral type first emerged can be wholly irrelevant. With Fred and his water pumps, the villagers and their wheelbarrows, or the fennec fox and its ear, the explanation for why some novel water pump, wheelbarrow, or ear size emerged need not be that it was an improvement in function. Similarly, for our learner, it need not be true that the emergence of some novel behavior is for improvement. Like the previous cases, the target of explanation for which the ideal might play an explanatory role is the pattern of use of some behavior by our individual in response to some stimulus.

Assume some stable pattern of behavioral use by an individual, and assume that we can form a contributory description of that behavioral type. That stable pattern of use presents a normative regularity if exceptional behavioral types would be acted against upon their emergence in virtue of the fact they are exceptions. As behavioral types can differ in amplificatory value, behavioral types with lower amplificatory values will be acted against in virtue of the fact that they are exceptions to the dominant behavioral pattern. They are acted against in virtue of being exceptions for just the same reason that a deviant phenotype is acted against under a selective regime. The behavioral types are in direct competition for the resources of the individual, because given some stimulus on one occasion the individual can only produce one behavioral token.[89] Which behavioral

---

[89] I suspect that this fact counts out an analogue in the ontogenic case of the apparent progress produced by genetic drift in the phylogenic case. Genetic drift can occur not only in the absence of differential fitness

token is produced will affect, given its amplificatory value, the future replication of that behavioral type by the individual. Since those behavioral types with higher amplificatory values will tend to beat out those with less for the resources of the individual, they will tend to increase their likelihood of future replication at the expense of those behavioral types with lower amplificatory value. We can have, then, with the beginning learner an initially stable of pattern of use that meets the conditions of a normative regularity.

Since the beginning learner is, however, operating at a suboptimal level, there is an available behavioral type with a higher amplificatory value. Upon the emergence of such a novel behavioral type (however that takes place), that novel behavioral type as a closer approximation of the putative ideal (captured in the contributory description of the behavior) will become the dominant behavioral type tokened by our individual. The transition in dominance will be explained by the fact that the predecessor dominant type is now a relative failure in respect to the novel dominant type. We can, then, in explaining the trajectory of behavioral transitions in the acquisition of some skill mobilize just the sort of explanation that earlier made for the ideal. We can, then, say of such ontogenic acquisitional processes that they are genuinely learning – that is, they are literally for the purpose of acquiring the relevant skill or capacity, because the individual is sensitive to the end of that acquisitional process.

With the nonartifactual ideal in hand, we have seen how acquisitional processes typical of ontogenic learning are genuinely teleological without the presumption of

---

but also in its presence absent competition among phenotypes. It was the latter sort that produced the earlier apparent progress. In the ontogenic case, since the behavioral types are always in competition for the resources of the individual, an analogue to phylogenic case should not present itself.

intentionality and psychology. I will in the subsequent chapter turn at last to intentionality and psychology. I will show that nonartifactual normativity as the ground of genuine teleology can and ought to serve as the theoretical foundation for the project of naturalizing the mind.

Chapter 8

NATURALIZING THE MIND: THE NORMATIVE FOUNDATION

With nonartifactual norms underwriting teleo-function and nonartifactual ideals underwriting the progressive acquisition of teleo-function, I want to return to a claim I made at the outset. I said that I saw no prospect for naturalizing the mind unless nonartifactual teleo-functional ascriptions could be literally true. The reason simply is that mental states are essentially teleologically characterized states. Those teleologically characterized states cannot themselves be artifice without an obvious vicious regress. If there is mentality, then there must be nonartifactual teleology. Consequently, the value of the preceding exercise to my mind lies in shoring up a theoretical pillar requisite for naturalizing the mind.

The aim here is two-fold. First, I want to make a case, albeit briefly, for the claim that mentality is essentially teleologically characterized. The aim of naturalizing the mind is to explain how mentality is part of the natural world, and, unless one supposes mental states are metaphysical primitives like the fundamental particles of physics, naturalizing the mind is to explain mentality by appeal to non-intentional and non-psychological facts. Given the traditional restriction of teleology and especially normativity to the province of the psychological and the intentional, the naturalist's explanation of the mind had to appeal to non-teleological and non-normative facts. But, we have seen that teleology and normativity are not merely intentional and psychological phenomena, and consequently, the naturalist can make use of teleological and normative facts in the explanation of mentality. And, if mentality is essentially teleological and given the connection between genuine teleology and normativity, the naturalist will have to rely on normative facts in the explanation of the mental. This brings me to the second aim. I want to show that the prior account of nonartifactual normativity can provide the theoretical basis required for naturalizing the mind.

The naturalist faces two sorts of general theoretical questions: 1) how can a physical state be a mental state; and, 2) how can physical facts determine the representational content of a mental state? The first reflects the traditional mind-body problem, and the second reflects the problem of intentional inexistence. These questions (or problems) are not wholly distinct, and it is difficult to answer one, I think, without addressing the other. That said, what I propose to do here will occur in two acts, one for each of those questions. In respect to the first, I aim to show that the gross individuation of mental types is teleological – that is, what a mental state type is is given by the specification of a certain job or work type.[90] Physical states are mental states the extent to which they literally perform the work of a mental state. In respect to the second, the aim is similarly to show that content is determined teleologically – that is, to specify the content of an intentional icon[91] is to specify its purpose or function within the cognitive

---

[90] Dennett (1978 & 1987) and Lycan (1981) are probably the primary sources of this view within philosophy of mind, and much of what I have to say below is heavily influenced by those works.

[91] Millikan re-introduces Peirce's notion of "icon" as the technical term "intentional icon":

> Intentional icons are devices that are "supposed to" map *thusly* onto the world in order to serve their direct proper functions; that is, Normally they do map so when serving those functions. And they are devices that are supposed to be used or "interpreted" by cooperating devices. Thus they exhibit a sort of "ofness" or "aboutness" that one usually associates with intentionality. (Millikan 1984, 95-6)

I use her term here with a slightly less technical meaning. An intentional icon is whatever possesses semantic value but itself is not a complete representation. Tokened logical and nonlogical terms of inner economy.[92] The physical facts that determine the content of an intentional icon are the normative facts binding that intentional icon.

I. The gross architecture of mentality

a. *Mentality as essentially teleological*

Let's start with the first question "How can a physical state be a mental state?"[93]

As Ryle (1949) nicely pressed, the standard mentalistic terms (e.g., 'belief', 'desire', 'hope', 'expect', 'stupidly', 'intelligently', 'cunningly', 'foolishly', etc.) are primarily used to make explanatory-cum-predictive assertions, and the target of such assertions is the behavior of physical entities. That explanatory-cum-predictive mentalistic framework can be brought to bear on any physical behavior, however. E.g., the sun rises, because the Earth likes to rotate; or, my chair is stationary, because it believes that it would end up in a zoo were it to move. That said, that explanatory-cum-predictive mentalistic framework is only of particular explanatory value in respect to certain sorts of physical behavior.

---

episodic thought, spoken language, and written language are all examples of intentional icons. It will turn out that the best way to understand the possession of semantic value is that the intentional icon possesses the teleo-function of mapping thusly because its consumption by a cooperative device is subject to a standard of performance. So, intentional icons will turn out to be pretty much what Millikan says that they are, but I want that to be the result of argument and not the stipulation of a term.

[92] Teleosemantics is now a developed philosophical industry: see, for example, Millikan (1984, 1993, 2004, & 2005), Papineau (1987 & 1993), McGinn (1989), Neander (1995), Godfrey-Smith (1996), Rowlands (1997), and Price (2001). What I suggest below is most directly influenced by Millikan (1984) but also by what I find to be the recurrent teleo-functional theme in Sellars' analysis of the meaning rubric. See, for example, Sellars (1963, 1967, 1969, 1979, & 1981).

[93] The dominant answer to that question until the last century had been Descartes' assertion that physical states just cannot be mental states, because the latter were essentially nonphysical. Descartes' dualism, like eliminativism and fictionalism, is a theory of last resort – that is, it is a theory worth countenancing only when the naturalizing project has clearly failed. As such, I will not discuss dualism or its theoretical counterparts, eliminativism and fictionalism, here.

James, in setting out to describe the scope of psychology, nicely describes the certain sorts of behavior which rightly seem to occasion the mentalistic framework.[94] James writes,

> If some iron filings be sprinkled on a table and a magnet brought near them, they will fly through the air for a certain distance and stick to its surface. A savage seeing the phenomenon explains it as the result of an attraction or love between the magnet and the filings. But let a card cover the poles of the magnet, and the filings will press forever against its surface without it ever occurring to them to pass around its sides and thus come into more direct contact with the object of their love. Blow bubbles through a tube into a pail of water, they will rise to the surface and mingle with the air. Their action may again be poetically interpreted as due to a longing to recombine with the mother-atmosphere above the surface. But if you invert a jar full of water over the pail, they will rise and remain lodged beneath its bottom, shut in from the outer air, although a slight deflection from their course at the outset, or a re-descent towards the rim of the jar when they found their upward course impeded, would have easily set them free.
>
> If now we pass from such actions as these to those of living things, we notice a striking difference. Romeo wants Juliet as the filings want the magnet; and if no obstacles intervene he moves towards her by as straight a line as they. But Romeo and Juliet, if a wall be built between them, do not remain idiotically pressing their faces against its opposite sides like the magnet and the filings with the card. Romeo finds a circuitous way, by scaling the wall or otherwise, of touching Juliet's lips directly. With the filings the path is fixed; whether it reaches the end depends on accidents. With the lover it is the end which is fixed, the path may be modified indefinitely.
>
> Suppose a living frog in the position in which we placed our bubbles of air, namely, at the bottom of a jar of water. The want of breath will soon make him also long to rejoin mother-atmosphere, and he will take the shortest path to his end by swimming straight upwards. But if a jar full of water be inverted over him, he will not, like the bubbles, perpetually press his nose against its unyielding roof, but will restlessly explore the neighborhood until by re-descending again he has discovered a path round its brim to the goal of his desires. Again, the fixed end, the varying means!
>
> … with intelligent agents, altering the conditions changes the activity displayed, but not end reached; for here the idea of the yet un-realized end co-operates with conditions to determine what the activities shall be. (James 1890/1950, 6-8)

---

[94] The second chapter of Ryle's *The Concept of Mind* (1949) equally contains wonderful descriptions of the sorts of behavior to which the mentalistic framework is particularly apt. A more contemporary example within evolutionary reasoning is seen in Godfrey-Smith (1996, chapters 7-9, & 2002).

James' suggestion is that the mentalistic framework is rightly brought to bear to explain a certain sort of plasticity in the behavior of physical entities. That plasticity is the adaptation or modification of behavior to changing circumstances to bring about some consistent result or end.

The mentalistic framework, James suggests, is rightly brought to bear in respect to compensatory behavior, because that framework provides some explanatory headway into how that behavior is possible. If such compensatory behavior is not accidental, then the modification of behavior should not only fit the changed conditions but it should be because those conditions obtain. If James' frog and Romeo can know what those present conditions are (that is, if they have some ability to form indicative representations), then they would be capable of altering or modifying their behavior due to changing conditions. But, varied behavior under changing conditions alone is not compensation. Compensation is to bring about the same end. If James' frog and Romeo can not only form indicative representations but imperative representations as well, then, by being aware of both world and their ends, they would be able to modify that behavior to bring about some consistent end. However, awareness of the world and one's end does not generate behavior. James' frog and Romeo will need a further ability to use those representations to generate the appropriate action. If James' frog and Romeo were further capable of inference over their representations, then we would now seem to have made some headway into how they are able to produce compensatory behavior: "for here the idea of the yet un-realized end [i.e., imperative representation] cooperates with conditions [i.e., indicative representation] to determine [in inference] what the activities shall be." (James 1890/1950, 8) In contrast, James' iron filings or the Earth's rotation do not fit their behavior to changing conditions, and there is just no need consequently to think that they form representations of the world to accomplish what they do.

That said, the mentalistic framework has no special claim to best explanation in respect to compensatory behavior. The compensatory behavior that James describes could equally well reflect the operation of a homeostatic system. The varied homeostatic and regulative systems of the body do not seem rightly to occasion the mentalistic framework. While it is true of homeostatic systems that "altering the conditions changes the activity displayed, but not the end reached", it is not true that "the yet un-realized end cooperates with conditions to determine what the activities shall be." The latter is false, because the explanation of a homeostatic system need attribute no causal efficacy to the end. A homeostatic system, despite its plasticity, is just like the iron filings in the sense that its path is fixed given particular conditions. Importantly, it is not just that the homeostatic system is not itself to be explained by appeal to the mentalistic framework, but it is, rather, that homeostasis provides an alternative or competing explanation to the mentalistic explanation. If some compensatory behavior is best explained by some underlying homeostatic structure, then there is just no positive reason to invoke the mentalistic framework.

Physical entities often exhibit compensatory behavior – that is, they bring about consistent ends by varying means to coordinate with changing conditions. One general mechanistic hypothesis of that compensatory behavior is that the entity adjusts its behavior in light of its representations of present conditions and its ends.[95] An alternative

---

[95] To supply a mentalistic explanation is importantly to supply a mechanistic hypothesis. Remembering Brandon (1984) from the second chapter, to provide a mechanistic hypothesis is just to offer a model of the process underlying some effect of interest. The effect of interest, as James points out, is compensatory behavior, and his suggestion is precisely to offer a model of the mechanism underlying that behavior: "for

general mechanistic hypothesis is that a homeostatic system is in place in the entity generating differential responses to particular inputs. These are both hypotheses for the same general phenomenon, namely compensatory behavior. Deciding between them, deciding which is the best explanation, is not to be done by trying to find some special type of compensatory behavior appropriate to one and not the other. If there is a difference between them, it is not in what is produced but in how it is produced. There is an important difference, I want to suggest, in how these mechanisms produce behavior, and that difference is substantive enough to distinguish which empirical conditions will favor one or the other hypothesis as the better explanation.

Roughly, the homeostatic hypothesis involves hypothesizing a set of pre-built, preset, or otherwise hardwired connections between particular inputs to some single output. In contrast, the mentalistic hypothesis does not hypothesize a set of pre-built, preset, or otherwise hardwired connections, because what a representational system allows is for those connections to be built on the fly. The representation is a complex: its content is a function (in the mathematical sense) of, at least, mood (indicative or imperative), the semantic value of its composite terms, and the syntactical and compositional rules governing the assembly of that composite. Given some set of composite terms and syntactical rules, the possible range of representation is in a sense preset or hardwired. While the range might be in a sense preset or hardwired, what is not preset is the tokening of this or that representation from the range. When a representational system is working, the indicative and imperative representations tokened

---

here the idea of the yet un-realized end co-operates with conditions to determine what the activities shall be." As Ryle pressed, even the Cartesian, who separates the mental from the physical on the principle that minds are non-mechanical, adopts a "para-mechanical" explanation of the mental. The Cartesian retains all the familiar mechanical language in respect to mental processes; he just insists that mental mechanisms are somehow not run-off-the-mill physical mechanisms, whatever that amounts to. (Ryle 1949, chapter 1) are those appropriate to the particular conditions that obtain and the ends of the system. But, the representations do not generate practical action. It is the further ability of inference over those representations that generates practical action. The connection between particular conditions, means, and ends is constructed on the fly, because it is in response to the tokened representing that the representer generates compensatory behavior through the activity of inference. In contrast, the track from input to output is already set in the homeostatic system: the creation of that track is not something generated in response to particular conditions; it is already there.

Importantly, that difference between the mechanistic hypotheses allows the specification of empirical factors that would favor one over the other hypothesis. One such factor, for example, is temporal. Building the connection between particular conditions and means on the fly occurs in real time and requires, then, time for inference to establish the connection to practical action. The homeostatic system can avoid such temporal costs in virtue of the fact that the connection is already preset. The extent to which compensatory behavior is time-sensitive can favor one or the other hypothesis. For example, where the response to predator detection needs to be sufficiently quick and predator avoidance behavior is sufficiently consistent (e.g., dart forward erratically), the homeostatic hypothesis is the more plausible.

A further factor is the relative cost of error in fitting the means to particular conditions. The representational system introduces a number of places where error can occur that are not present in the homeostatic system. For example, things can go awry in the formation of a representation or in the inferences generated. The construction of these connections between conditions and means on the fly requires more internal systems to be operating correctly than does a system where the connection is preset. As a result of this increased complexity within a representational system, there will be generally a higher probability of internal breakdown or error than within a homeostatic system. Similar to the temporal factor, a reason to favor a homeostatic system over a representational system is when the costs of error in compensatory response are relatively high. For example, where the costs for failing to engage in appropriate predator avoidance behavior are fairly high and where some set of behavioral means to dealing with that threat are available, the system might be better prepared to deal with those threats by forgoing the costs of error implicit in a representational system.

A further factor still is the extent to which, to use Clark's (1997) phrase, the world is "sufficiently unruly". When the range of conditions to which the entity need compensate is sufficiently wide and when the range of means required for compensation is sufficiently wide, a representational system enjoys a clear advantage over a homeostatic system. When the world is sufficiently unruly, the entity is faced with, as Clark (Ibid.) calls it, "a representationally-hungry problem". It must fit its behavior to highly variable circumstances with variable means, but predicating what in that range will present itself is far more difficult than just building the required connections on the fly.[96]

I don't intend the above to be a complete list of all the factors involved; I intend, instead, only to give a sense to the sorts of considerations that operate to favor one or the other hypothesis. The mentalistic hypothesis is the more plausible hypothesis and has some claim to best explanation, when it is implausible to believe that the compensatory

[96] See Godfrey-Smith (1996, chapters 7-9) for a nice example for how one might work out precisely the sorts of conditions that favor one or the other hypothesis.

behavior of an entity would be best explained by a system of pre-built or otherwise hardwired modes of response. Whether an entity is or is not mental – that is, whether it is rightfully a target of mentalistic explanation – is a wholly empirical matter. Historically, we, the human race, have frequently found ourselves to be James' savage, having wrongly attributed mentality because the relevant entity, whether a tree, a mountain, or the weather, failed to compensate or modify its behavior under varied conditions. Alternatively, we might discover that, though the entity compensates for changing conditions, it does not require any representational or inferential abilities to do so. That is, we can discover that its compensatory ability is best explained by the presence of a homeostatic system and not a representational system.

For example, the immune system's ability to develop defenses against novel invaders might seem on its face to require the mentalistic framework: the immune system recognizes novel threats as threats, acts to eliminate those threats, and, having learned of a new type of threat, remembers that type for future occasions. What we see with acquired immunity is compensation to bring about the same end (i.e., defense against invading organisms) under a range of circumstances and by varied means (i.e., varying its eliminative actions). Given the continuous mutation of pathogens, the world ought to be presenting on a regular basis threats that no human body has previously encountered. So, it would be implausible that some preset mode of response was in place to recognize these novel threats as threats. But, despite its appearance, the immune system is not faced with a representationally-hungry problem. The lymphocytes of the immune system recognize a foreign agent when an antigen of that foreign agent binds to an antibody on the surface of a lymphocyte. An antigen fits like a key to an antibody's lock. But, the key to lock fit allows for some degree of wiggle room, so many different molecular keys will fit the same molecular lock. It has been calculated that a hundred million different antibodies generated at random would cover the whole set of possible antigens, and it turns out that human bodies can make about a hundred million different antibody molecules. (Kauffman 2000, 11-3). Burnet's theory of clonal selection hypothesizes that human bodies generate a diverse repertoire of antibodies innately. When a foreign molecule on a bacterium or virus binds with a lymphocyte, that lymphocyte divides repeatedly, increasing its frequency in the overall population of lymphocytes. Somatic clonal selection within the population of lymphocytes is all that is required to account for the acquired response to some new invading pathogen. (Edelman 1992, 75-9). Since the immune system has built within it in effect a preset response to whatever state of the world it might encounter, there is no explanatory reason to invoke the apparatus of the mentalistic framework.

Or, in a similar vein, the elaborate "cathedral"-like structures of termite mounds would seem to require the attribution of mentality to individual termites. Lacking any foreman directing the construction of these mounds, individual termites in the colony would seem to have to coordinate their activity: they would seem to need to keep track of what has been done so far, what is needed next, and so on. On its face, this looks like a representationally-hungry problem. The massive number of small steps in the process of building the mound coupled with the limited contribution of any individual termite would seem to make it unlikely that any individual termite has the whole set of possible contingencies pre-built in its head. From an explanatory viewpoint, it would seem far simpler to allow that the termites can represent how they are situated within the scheme of things as well as the overall plan, thereby adjusting their constructing behavior appropriately.

However, there is a third explanation and one that is far simpler than postulating either a hardwired set of responses to every possible contingency or a representational system. Clark writes,

> All the termites make mud balls, which at first they deposit at random. But each ball carries a chemical trace added by the termite. Termites prefer to drop their mudballs where the chemical trace is the strongest. It thus becomes likely that new mudballs will be deposited on top of old ones, which then generate an even stronger force…. Columns thus form. When two columns are fairly proximal, the drift of chemical attractants from the neighboring column influences the dropping behavior by inclining the insects to preferentially add to the side of each column that faces the other. This process continues until the tops of the columns incline together and an arch is formed…. At no point in this extended process is plan of the nest represented or followed. No termite acts as a construction leader. No termite "knows" anything beyond how to respond when confronted with a specific patterning of its environment. The termites do not talk to one another in any way, except through the environmental products of their own activity. (Clark 1997, 75-6)[97]

A simple recursive rule set will do to explain how the collaborative interactions of termites result in the complex architecture of the termite mound. Further, that rule set does not require that individual termites be prepared to recognize some massive range of contingencies to which they fit their behavior. The mentalistic framework, consequently, is not on empirical grounds the best explanation of termite mound construction.[98]

[97] Clark intended this passage to be an example of how simple recursive action among collectives can add up to complex behavior without some central processing or planning component. That point for Clark's purposes is important in resisting a Cartesian intuition infecting classic AI, but I think that the passage works equally well in the present context. What we see here is a way to expose empirically an apparently representationally-hungry problem as merely apparent. For more on termite architecture and the power of simple recursive rules with respect to complex behavior, see Beckers et. al. (1994) & Resnick (1997).

[98] See Brooks (1997 & 1999) for a number of examples of nonrepresentational systems handling scenarios that on their face would seem to require a representational system.

But, in contrast to those negative cases, we can also find ourselves forced to attribute mentality to previously considered non-mental entities based on empirical evidence. Darwin, in investigating the life history of the earthworm, found himself faced with just such a case. Recognizing that attributing mentality to earthworms will strike everyone as improbable and conceding that he himself found it a surprising result, Darwin felt compelled to recognize that earthworms exhibit intelligent behavior. Darwin observed that earthworms plug their holes with leaves and that they appeared to modify their behavior in respect to differing types of leaves so as to make the task as simple as possible (e.g., drawing a leaf with a narrow tip and a wide base in by its tip or, for a leaf with a more uniform base to tip ratio, seeking out the foot stalk of the leaf). Darwin subjected his worms to a number of experiments with different native leaves, foreign leaves, and artificial paper leaves, and repeatedly his worms compensated for the variances to bring about the same end. What pushed Darwin to recognize worm intellect is that earthworms successfully engaged in such compensatory behavior in novel conditions. For example, Darwin writes,

> In this case the worms judged with a considerable degree of correctness how best to draw the withered leaves of a foreign plant into their burrows; notwithstanding that they had to depart from their usual habit of avoiding the footstalk. (Darwin 1881/1985, 70)

Darwin attributes some capacity for judgment and representation to the earthworm, because its compensatory behavior went beyond what plausibly could have been hardwired by instinct alone given the ability of the earthworm to compensate for novel types of leaves.[99] To exhibit such plasticity in the face of novelty suggests a

[99] As some of his further evidence for earthworm intellect, Darwin noted that, in building protective basketlike structures over the mouths of their burrows, his earthworms had pressed the sharp points of the Scotch pine needle into the lining of voided earth to avoid being damaged or trapped by the ends of the

representational system, because novelty suggests that the connections are not pre-built in the system but are rather being constructed on the fly. It is not surprising then that Darwin felt compelled given his evidence to conclude that the worm must have some representational and inferential capacity.[100]

The point of having gone through the above is to stress that the mentalistic framework can provide a legitimate empirical mechanistic hypothesis. It provides a mechanistic hypothesis by providing a model through the tripartite mentalistic core (indicative representation, imperative representation, and inference) of the process underlying a type of physical behavior, namely compensatory behavior. The mentalistic framework provides for a legitimate empirical hypothesis, because there are empirically specifiable conditions to favor that hypothesis over alternatives. I have suggested that the mentalistic framework is of distinctive explanatory value in respect to compensation conducted on the fly. That distinctive explanatory value just rests in the fact that it can provide a model of that sort of physical behavior where the competing mechanistic hypothesis does not. Now, the suggestion tying the explanatory value of the mentalistic framework to compensation on the fly is just that a suggestion. I have not provided above

---

needle. What Darwin found surprising and noteworthy about this behavior was that the Scotch pine was not a native plant. (Darwin 1881/1985, 112) Again, what is cementing the deal for earthworm intellect is that such compensatory behavior is presented in the face of novel circumstances. The novelty of the introduced leaves and needles made it seem implausible to Darwin that straightforward instinct could generate the resulting behavior. For commentary on Darwin's work on earthworms, see Ghiselin (1969), Graff (1983), Gould (1983), and Crist (2002).

[100] von Uexküll later rejects Darwin's conclusion that earthworms exhibit intelligence, because he claims to show that they do not discriminate the shape of leaves in order to draw them in most efficiently. He suggests that the behavior Darwin observed is, instead, a simple tropistic mechanism responding to the different tastes at different parts of the leaf. von Uexküll fails to explain away, however, Darwin's work on artificial leaves. What is interesting here is von Uexküll's empirical grounds to reject Darwin's hypothesis, namely that earthworm compensation can be accounted for by a simple tropistic mechanism keyed to taste and not shape. If von Uexküll was right, then the apparent novelty presented by foreign leaves could be merely apparent. Consequently, there would not be a reason to think that instinct alone could not perform the relevant work. (von Uexküll 1934/1957, 37-40)

any knockdown argument to that effect. I do think that suggestion tracks the empirical investigations of mentality typified by cognitive ethology and psychology, but I have not made that case here. What is important to my purposes here is that we can delimit a type of empirical phenomenon for which the mentalistic framework has some clear explanatory value and that value rests in the provision of a distinctive mechanistic model.

What's the metaphysical status of the mechanistic hypothesis generated from the mentalistic framework? The individuation of its core components looks to be straightforwardly teleological, because the development of that model follows standard decompositional functional reasoning. James suggests the mentalistic framework as an explanation of the systematic effect of compensatory behavior. James' suggestion of the tripartite core of mentality reflects what is required for the organism to pull of this compensatory feat. Since the organism fits its behavior to present conditions even when those conditions alter over time, it needs some subsystem that informs the balance of its system what those present conditions are. Further, since the organism compensates by bringing about some consistent end, there ought to be a subsystem informing the balance of the behavioral system of the end to be generated. Lastly, the organism, as a generator of compensatory behavior, requires some behavioral generation system that is responsive to the information about present conditions and ends. The tripartite core of the mentalistic framework is a description of the various feats that need to be pulled off if the organism is rightly compensating for its behavior. Indicative representation, imperative representation, and inference are each just descriptions of the functional capacities that James' frog and Romeo must have in order to produce such compensatory behavior.

My suggestion that an alternative mechanistic hypothesis could explain compensatory behavior is a further bit of standard functional reasoning. A system with pre-built responses to varied conditions could pull off compensatory behavior without the need of a representational system. The work that a representational system can do for an entity isn't needed for compensation if that entity possesses an alternative functional architecture. The subsequent suggestion that the mentalistic framework can explain compensation on the fly is a further instance still of standard functional reasoning. The work of inference is to build the relevant connections between conditions and means to an end on the fly. Further, it can conduct that work only if there are further systems capable of representing the present conditions and ends of the system. Again, the mentalistic framework is utilized as a description of certain functional capacities. I proceeded, then, to suggest that we could delimit the sorts of empirical conditions that would favor attributing to a system those functional capacities over alternatives.

It is at least the case that the mentalistic mechanism can be teleologically characterized, and it is at least the case that the mentalistic framework can perform explanatory work when it is so characterized. But, I made a stronger claim than those at the outset, namely that the mental is essentially teleologically characterized. Here's how, I think, we get that stronger claim: given the sort of empirical considerations that favor a mentalistic hypothesis as best explanation, nothing other than a teleological characterization of the hypothesized mechanism is warranted. To assume that the gross architecture of the mechanistic model is type identical to a structural/compositional type, dispositional type, or machine functional type is to introduce an assumption not supported by the empirical grounds for the mentalistic hypothesis. Only the teleological characterization of the mental mechanism is warranted given those empirical grounds.

The Identity Theory of Place (1956) and Smart (1959) (i.e., psychological types are identical to neural types) is generally rejected on the grounds that it is overly chauvinistic. The theory seems to restrict unreasonably the attribution of mentality to entities that share our neurological structure. That Identity Theory is too chauvinistic or unreasonably restricts the attribution of mentality reflects the fact, I want to suggest, that it involves an assumption that goes beyond or is unwarranted given the empirical grounds to attribute mentality. Let me offer, first, what I think is an analogous sort of case. In observing a biological lineage over time, we observe that certain traits of ancestors reappear in offspring and do so even under varied external ecological conditions. A plausible hypothesis based on that empirical evidence is that there is something that performs the work of passing on traits from ancestor to descendent. That is, there is an underlying process or system that performs the feat of heredity. The sort of general empirical considerations that support thinking that there is such an underlying mechanism do not warrant thinking that that mechanism is of any particular structural or compositional type. To think that mechanism must be identical to some structural or compositional type, e.g., DNA, RNA, chromatin, and so on, is to go beyond what is warranted by the empirical evidence for heredity. All that is warranted is that something performs the relevant work. Similarly, the empirical grounds to think that an entity is minded warrant minimally that some underlying set of processes account for compensatory behavior. But, those empirical considerations do not warrant thinking those underlying processes are of any particular structural or compositional type. The empirical considerations in favor of the mental hypothesis, like those for the above heredity hypothesis, only get us as far as that something is doing the relevant work of indicative representation, imperative representation, and inference. They do not warrant thinking that those mechanisms are of any structural or compositional type.

That Identity Theory is overly chauvinistic reflects the fact, it is commonly said, that mental states can be multiply realized by differing structures of differing compositions. That they can be multiply realized, I have suggested, just reflects what we are warranted to think about the mentalistic mechanism given the empirical considerations favoring the attribution of mentality. Further, these multiple realizability considerations work just as well against dispositional accounts such as Ryle's (1949) or against machine functional accounts such as Putnam's (1967).

As Millikan (1999b) and Shapiro (2000) stress, multiple realizability of function by variance in composition and structure is a weak form of multiple realizability. For example, it is reasonable to think that corkscrews that differ in composition can all, nonetheless, be corkscrews. It is reasonable, because whether a corkscrew is made out of steel or carved granite does not seem to make any difference to how it removes corks. Such a form of multiple realizability is considered to be weak, because the difference in composition is irrelevant to how those corkscrews do their work. Strong multiple realizability requires that the differences between items are relevant to how they do their work. Think of the difference between a traditional corkscrew, which is screwed into the cork, and the newer vacuum-operated corkscrew, which just sucks the cork from the bottle. While both types are different compositionally and structurally, the important difference between them is the way in which they operate.

Dispositional and machine functional accounts of mentality equally fall prey to multiple realizability considerations, because there are often varied ways to perform the same work. Varied dispositional sets across individuals can just be different ways to do the same thing as with our corkscrew. Both a deer and a sea hare exhibit the ability to avoid predators, but their avoidance behavior in the face of predation reflects radically different dispositional sets. The deer in response to particular visual, aural, and olfactory cues is disposed to sprint forward erratically. The sea hare in response to tactile cues draws its appendages inward and ceases movement. These varied dispositional sets are just different ways to avoid predators. The same sort of consideration applies to machine tables. Again, think of the deer and the sea hare. Between the two of them, there are radically different inputs and outputs, and, given the very different physiologies involved, there will be distinct sequences of state transitions connecting those inputs to outputs. The machine tables for the deer and the sea hare are both machine tables of the same thing, namely predator avoidance, but the machine tables are radically different.

Just as with predator avoidance behavior, it is reasonable to think that there are radically different ways in which a physical system could perform the operation of, say, an indicative representation. For example, if I want to indicate to you that a dangerous monster is lurking behind you, I can call out, write out, or tap out in Morris code, "Monster!" or even just widen my eyes while inhaling sharply and staring over your shoulder. We could distribute these various ways across different individuals. Now, in each of our different individuals, we will have very different dispositions and very different machine tables that all do the same thing. And, if there are very different ways to warn you, there is no reason to suppose that there are not very different ways to warn myself, i.e., to notice the presence of a threat. Again, each will reflect a different disposition or different machine table, but, despite those differences, they will all just be ways to do the same thing.

Like Identity Theory, the applicability of multiple realizability considerations to dispositional or machine functional accounts of mentality reflects the fact that they enjoin us to take on a metaphysical commitment unwarranted by the empirical grounds to attribute mentality. For example, the empirical grounds to think that an organism is engaged in predatory avoidance behavior warrant thinking that there is some underlying mechanism accounting for that systematic effect. They do not warrant thinking that that mechanism is of any particular dispositional or machine table type. We have reason to think that some underlying process is generating predator avoidance, but we have no reason to think that the process is of any particular structural, dispositional, or machine functional type. Similarly, in respect to mentality, the empirical grounds to think some entity minded do not warrant thinking that the underlying process is a particular structural, dispositional, or machine functional type. What is warranted by the empirical evidence for mentality is the teleologically characterized model of the mental mechanism. That is, the empirical grounds for the mentalistic hypothesis provide some reason to think that something is doing the work of representation and inference. Any further metaphysical claims beyond the teleo-functional characterization of this gross architecture will go beyond what is warranted by the empirical evidence for mentality.

The last can be bolstered by noticing that the inverse of multiple realizability applies to structural/compositional, dispositional, and machine functional typing as well. Some single function is multirealizable if differing structures, compositions, or means can perform the same work. But, the inverse relation holds as well. Any compositional, dispositional, or machine functional type can do very different things. For example, think of a single screw. That screw can perform a variety of jobs: it can fix two objects together; it can support a picture hung from it, it can fill in a hole in a pipe; it can indicate ambient temperature; it can be a pawn in a chess game, and so on. In doing each, the screw is acting in exactly the same way. Whatever physical description we have of the screw and its varied dispositional properties (e.g., how it changes in response to varied temperatures, how it will respond to changes in pressure, etc.), it will be the same description for each of the above jobs. Similarly, think of a computer chip for which we have formed some machine table description – that is, a complete list of the inputs, state transitions, and outputs of that chip. We can put the chip in a lawnmower in order for it to regulate the carburetor, put in it a toy dinosaur to control motion, put it in an irrigation system to control flood gates, etc. That single chip with its single machine table can perform a variety of different jobs. Or, think of the deer that responded to certain sensational cues with erratic forward sprinting. We could form a dispositional set or machine table of those responses, but those will not explain what the deer is doing when it so responds. It might be that it is engaging in predator avoidance. It might not be. Those same sensational cues and the erratic sprinting in a somewhat different environment might be part of a mating display.

Structural/compositional, dispositional, or machine functional types can perform a variety of work, because the conditions for a structural/compositional, dispositional, or machine functional type need not be identical to a work type. For example, the conditions for something to be a screw are different than what is required for something to be a pawn in a chess game. This tokened screw can be a pawn when it meets the further conditions on being a pawn, but being a screw does not suffice to be a pawn. Or, the conditions for this computer chip to be a token of a machine functional type do not suffice for it to be a regulator of a carburetor. That chip, as a token of a machine functional type, is a regulator of a carburetor when it provides that contribution to the system in which it is embedded.

With respect to mentality, the empirical grounds for an entity to be a minded entity are not identical to the conditions for something to be a token of a structural/compositional, dispositional, or machine functional type. For example, the facts that secure that some token is a token of a structural/compositional type, e.g., a neurological structural type, are not the facts that secure that it is a token of a mental type. This token is a token of neurological type given facts about neurology and spatial distribution. That token is a token of a neurological type independent of whether it is housed or not within a body or whether it is hooked up to my digestive system or my visual system. Whether this tokened neurological structure is a mental type depends on further facts about how it is used within the body. Or, that a bit of my neurology realizes some neurological disposition or, more generally, that a bit of my physiology realizes some physiological disposition depends on neurological or physiological facts, respectively. Those facts are independent of facts about, say, compensatory behavior. That I am disposed to act in a particular way might reflect a cognitive disposition, but whether it does depends on how that disposition is used in respect to, say, compensatory behavior. The same point applies again to machine functions. That a bit of my neurology, say, realizes a machine table reflects facts about inputs, outputs, and state transitions.

Those facts can obtain whether or not that neurological structure is within my body. (That is, we could remove the neurological structure from the body and within a laboratory setting effect the same machine table by properly stimulating the neurological structure.) The facts that determine whether this tokened machine function is a token of a mental type will be further facts about its involvement in my capacity, say, to generate compensatory behavior.

Given the explanatory purpose to which the mentalistic framework is put, that framework provides an empirical, mechanistic hypothesis. That mechanistic hypothesis, I have suggested, is put forth to explain the systematic capacity of compensatory behavior conducted on the fly. That mechanism looks to be straightforwardly individuated on the basis of the varied feats required to pull off that systematic effect. The description of the mental mechanism is at least a teleologically characterized mechanism insofar as it is a description of the cooperative causal interactions required for compensatory behavior on the fly. And, it is at most a teleologically characterized mechanism, because the empirical grounds for the mentalistic hypothesis do not warrant a characterization of that mechanism as any particular structural/compositional, dispositional, or machine functional type. Further, the grounds for a given structural/compositional, dispositional, or machine functional type are not the grounds for mental typing. The mentalistic framework is, consequently, an essentially teleo-functional framework.

b. *A normative foundation for mental functions*

Let's return finally to the initial question "How can a physical state be a mental state?" Well, the answer is not that mysterious if mental states are essentially teleological states.[101] Physical states are mental states when they perform the work of a mental state – that is, when they provide the sort of causal contribution to the operation of a system that is a representing or an inferring.[102] Physical states are mental states in just the same way that physical states can serve the functions of flying, running, stalking, eating, etc. That some physical state literally performs some work – that is, that some physical state literally possesses a teleo-function – requires that it be subject to a standard of performance. With respect to mentality, this requires that, if an entity is a minded entity, it needs to be sensitive to the successes and failures of its putative mind.

---

[101] There is, of course, a large gap between the gross architecture of the mentalistic mechanism and, say, the neurological structure. That gap is crossed by further decompositional or homuncular analysis of the gross functional architecture. That there is such a gap is not a special fact about mentality. The gross architecture of heredity, reproduction, flight, or any other higher level biological function will present a similar gap, and those gaps are crossed via further decompositional analysis. Lycan (1981) suggests that it is this distance between the gross functional architecture of mentality and neurology that gives rise to the perception that there is a mind-body problem.

[102] If we construe the representing function broadly enough such that it includes any intentional function, then that a physical state literally serves a representing function does not suffice for it to be a mental state. Let me offer an analogy. The fennec fox's ear and my heat pump both serve thermoregulatory functions. However, that my heat pump serves a thermoregulatory function does not render that function biological. And, that the fox's ear serves a thermoregulatory function does not suffice to render that function artifactual. The fox's ear serves a biological function given the sort of system in which it makes its thermoregulatory contribution. The suggestion that I have been operating with is that the mind reflects a system whose functional decomposition involves three main gross architectural components. Further, those components are themselves functionally characterized. For a state to be a mental state, it needs not only to perform one of the three functionally characterized roles, but it needs to do so within the relevant system, i.e., one with the balance of the mental architecture. If representing is construed broadly enough that any intentional function is a representing, then some representings will not be part of mentalistic systems.

For example, biology provides numerous cases of systems with the functions of detecting, indicating, or responding to some state of the world. For example, E.Coli has a subsystem dedicated to the detection of the location of food by measuring changes within the chemical gradient. Its internal states detecting food are intentional states. Those states are, given their function, literally of or about the location of food, and like any intentional context, that the individual E.Coli detects that food is that way does not imply that food is that way. Why are systems like E.Coli not minded? What they seem to lack is one of the functional units of minds, namely inference. And, the reason to think that E.Coli lacks inference is that it simply has no need for it – that is, given the way E. Coli operate, there is no reason for them to compensate for the changing location of food on the fly, a simple tropistic mechanism will do. I think that it is not unreasonable to withhold mental attribution in the absence of inference. Inference both seems intimately bound up in traditional conceptions of the mental as well as aids in marking off a distinctive mode of operation, namely compensation on the fly.

The primary function of the mind, I have suggested, is the production of compensatory behavior on the fly. If an entity is in fact minded, then it should be the case that a teleo-function of that entity is to compensate for changing conditions. Consequently, that entity should be sensitive to its successes and failures with respect to such compensation. Given the conditions for normativity, there should be a pattern in place of the following sort: there is a pattern of compensatory behavior, and the best explanation of the stability of that pattern is that exceptions are subject to modification in virtue of being exceptions. So, for example, think of Darwin's worms and their drawing leaves into their holes. There is a pattern of worms varying means to fit differing conditions to achieve the same overall result of efficiently drawing a leaf into their hole. Exceptions to that pattern will be worms failing to draw a leaf into their hole in the most efficient manner. The more efficient ways of acting are metabolically less expensive than the less efficient ways of acting. As we saw in chapter 6, we can turn to the phylogenic level to get those metabolic costs to make a difference in respect to the sorts of behavior engaged in by the overall population. The same sort of moves made with the shrimp earlier can secure a norm of performance here with our worms. The stability of the relevant compensatory behavior will reflect the fact that those which fail to engage in that compensatory behavior sufficiently will be eliminated from the population. Under the right conditions, we can have, then, norms of performance governing this general function of the worm to engage in compensatory behavior.

That norm, however, only establishes that a genuine function of the worm is to fit its behavior to changing conditions. It does not establish that the worm is a minded entity. There are two general mechanistic hypotheses (i.e., the homeostatic and the mental) for how the worm might be engaging in that work. The norm governing compensatory behavior only shows that compensatory behavior is a genuine function of the worm. It does not establish how it is engaging in that work. However, there should be different standards of performance governing the different ways to fit behavior to conditions.

Why should there be different standards of performance? Consider the earlier corkscrew examples. There is a general standard of performance that governs corkscrews, whether the old-fashioned screw type or the newer vacuum type. Whatever type, corkscrews are supposed to remove corks. But, given that a corkscrew is the old-fashioned variety, there is also a standard governing how that type is supposed to get the work done of cork removal, i.e., that item should be able to pierce deeply the cork without shattering it. Similarly, since the two mechanisms in respect to compensatory behavior operate in different ways, there should be differing standards governing their operation. In the mental case the fitting of means with conditions for some end is occurring on the fly, whereas in the homeostatic case it is not occurring on the fly and is in some sense pre-built or preset. This isn't a difference like the difference between the corkscrews, however, where are there are clear different standards of correct operation. It is more difficult, as a result, to specify the distinct norms of performance.

Let's take a very simple homeostatic case. Assume there are three different conditions to which the entity responds with three respectively distinct means to accomplish the same end. There are three different causal paths in place in our entity, and that complete set is what explains the entity's compensatory behavior. Each has its own contribution to make to that compensatory behavior. If the contribution of each is a genuine function, there will be three distinct norms of performance, one for each causal path. So, for example, there is a pattern fitting some means-α to condition-β to bring about some end, and the stability of that pattern is to be explained by the fact that exceptions to that pattern are subject to correction due to their failure to connect means-α to condition-β. Since exceptions in a normative pattern are subject to modification in virtue of being exceptions, the exceptional circumstances for each of the three norms will be distinct. So, in the case above, the exceptional case is failing to fit means-α to condition- β, but in the remaining two cases, the exceptional case would be, say, failing to fit means-γ to condition-δ for one norm and failing to fit means-ε to condition-η for the other. The only point so far is that there are three distinct and separable normative patterns in place and that each is keyed to fitting some *particular* means to some *particular* condition.

Norms governing the connection of *particular* means to *particular* conditions will be missing, I think, in the mental case. The basic operating model of the mental is to forgo pre-established connections between means and conditions and, instead, generate those connections on the fly. As a result, the standard of performance governing the mental will not govern this or that connection between conditions and means but will govern how to form those connections well. (How those connections come to be is irrelevant in the homeostatic case. Instead, with homeostasis, the norm governs the stability or the maintenance of an established connection.) Given the compensatory function that the mind serves, a well-formed connection is just one that serves that compensatory function – that is, a well-formed connection is one that connects means appropriate to conditions in such a way to achieve the end. The standard of performance for a well-formed connection is just then the same standard of performance governing compensatory behavior generally. The standard of performance governing the mental function abstracts then from particular conditions and particular means, as it is just the specification of the general norm governing compensatory behavior. Exceptions to that norm are specified as just that which fails to bring about an appropriate means to a condition to achieve the end.

Here is the situation. In both cases, there is in effect the same general norm of performance governing compensatory behavior. In the homeostatic case, the reason for that general norm of performance reflects the fact that there are a set of distinct norms governing particular behavioral responses under particular conditions. In the mentalistic case, that general norm of performance does not rest on any set of underlying norms of performance governing particular behavioral responses under particular conditions. So what are the considerations that favor one or the other case? Well, just the sort of considerations suggested earlier. For example, what prompts Darwin to concede that his worms are minded entities is that they are able to engage in compensatory behavior under novel circumstances and that the latter could not be accounted for by instinct alone. In other words, what is maintaining this stable compensatory function in worms is not a set of particular selective pressures on particular responses to particular conditions. Rather, what is maintaining or keeping stable this compensatory function in worms is just the more general selective pressure targeted at orienting leaves. Or, the initial plausibility of termite intelligence reflected the enormous range of conditions to which the individual termite would have to adapt its behavior in order to contribute to mound construction. Given the enormity of that range, it would be an extremely rare event that any particular response to particular conditions took place. The rarity of those events would make it difficult for there to be selective pressure for any particular sort of event. Instead, it seemed more plausible that selective pressure is operating higher up at the general compensatory ability. If so, then the norm of performance for the compensatory ability does not rest on a set of norms for each particular response set. Again, it would be more plausible that the particular action was the result of a general ability, and not vice versa. (Of course, the attribution of intelligence to termites didn't work out, because the original assumption about adapting to a wide range of conditions didn't work out.)

So, if the general norm of performance governing the compensatory ability is not to be accounted for by a set of norms governing particular responses to conditions, then the entity must be compensating on the fly. Compensating on the fly does require these general representational and inferential abilities, and so, if the entity is genuinely representing and inferring, there will be norms of performance governing those abilities. These abilities are themselves the product of a further subset of functional capacities. The representational ability is at least the product of mood, various syntactical rules, and the semantic value possessed by intentional icons. The inferential ability will be the product of each of the general formal and informal inference types available to the entity. Instead, of developing a fuller picture for each of these further downstream functional capacities, let me just mention how to develop the norms of performance governing the general abilities of representation and inference. (I will take up semantics in the next section.)

Let's start with inference. The overall general function of inference is to create the connections between indicative and imperative representations in order to generate practical action. Inference is the mechanism responsible, then, for building the connections between conditions and means on the fly. If Darwin's worms are in fact engaged in judgment, there is a pattern of connections being formed in the process of generating compensatory behavior. This pattern of connections is not, of course, specified between particular conditions and particular means. Instead, the pattern of connections is like that of formal logic where we specify the shape of an inference without reference to the particulars of any single inference. Exceptions to that pattern would be those inferences that generate connections of a differing shape or form. If those exceptions result in failed compensatory behavior, then there will be clear metabolic costs; therefore, we can develop a phylogenic story for such a norm just as we have before for other phylogenic traits. If there was such a normative pattern for Darwin's worms, we would have a nonartifactual norm of performance grounding inference.

Let's turn to general representational function. The representational ability contributes to compensatory behavior by generating states usable by the inferential ability in its formation of connections. From the perspective of the inferential ability, we can think of these states as analogous to the uninterpreted formula of formal logic. The inferential ability does not need to know or be sensitive to semantics to do its work. However, if the representational ability is to contribute to compensatory behavior, these states do need to map onto the conditions in which the entity finds itself as well as onto its ends. That is, the representational ability, when working correctly, should generate states for the consumption of the inferential system that will produce behavior appropriate to the present circumstances. As circumstances alter or internal ends of the entity change, the representational ability is performing its function insofar as it produces states for the consumption of the inferential system such that the latter system produces appropriate behavioral modifications.[103] Failures of the representational system to map should generate, then, failures in compensatory behavior. If an entity is genuinely a representer, then it will be sensitive to its successes and failures with respect to such mapping. It will be so sensitive if there is a norm of performance governing the representational ability. We can form such a nonartifactual norm in the following way. There is a pattern of tokened and variable internal states the consumption of which by the inferential system generates behavior appropriate to circumstances. Exceptions to that pattern will be tokened internal states consumed by the inferential mechanism but that produce failed compensatory behavior. There will be a normative regularity in place just in case the best explanation of the stability of the pattern of internal states involved in the

[103] One might be tempted, at this point, to think that we can forgo talk of mapping functions as well as content more generally and, instead, just adopt a syntactical theory of mind like that offered by Stich (1978 & 1983). Bluntly, a syntactical theory of mind is not a theory of mind. As Dennett writes,

> The alternative of ignoring the external world and its relations to the internal machinery (what Putnam called psychology in the narrow sense, or methodological solipsism, and Gunderson has lampooned as black world glass perspectivalism) is not really psychology at all, but just at best abstract neurophysiology – pure internal syntax with no hope of a semantic interpretation. Psychology "reduced" to neurophysiology in this fashion would not be psychology, for it would not provide an explanation of the regularities it is psychology's particular job to explain: the reliability with which "intelligent" organisms can cope with their environments and thus prolong their lives. Psychology can, and should, work toward an account of the physiological foundations of psychological processes, not by eliminating psychological or intentional characterizations of those processes, but by exhibiting how the brain implements the intentionally characterized performance specifications of subpersonal theories. (Dennett 1987, 64)

As Dennett suggests, adopting methodological solipsism or a pure syntactical theory of the mind will fail to explain the regularities that a psychologist (and, I would add, a cognitive ethologist, as well as in fact any biologist who is not merely a biochemist) ought to explain. Here is an example of just such a regularity. The ground squirrel of Northern California will at times stand upright and start tossing soil through the air. What is the ground squirrel doing? From the methodological solipsistic standpoint, that question is meaningless. But, it is clearly not meaningless for the cognitive ethologist. Is the ground squirrel drawing the attention of a predator to save its young, showing its sexual prowess, indicating its unwillingness to mate, building a mound to hide seeds, or scaring a predator? Actually, the answer is none of these. The ground squirrel is tossing soil at a rattlesnake, and it is doing so to cause the snake to start rattling. The ground squirrel will then use the amplitude and frequency of the snake's rattle to judge the size of the snake and its body temperature. The ground squirrel will use the information about the size and temperature to assess the level of threat that the rattlesnake poses to itself and its young. (Owings 2002). To explain what our ground squirrel is doing is precisely the work of cognitive ethology, and that work will not be accomplished by looking to the internal syntax of our ground squirrel. The answer involves facts about ground squirrel ecology. And, importantly, that answer is through and through semantically laden.

production of compensatory behavior is that the exceptions are subject to modification and are so in virtue of being exceptions. Let's stick with Darwin's worms. There should be, if Darwin was right about worm intellect, a pattern of events internal to worms used in generating on the fly appropriate responses to different shaped leaves. Exceptions to that pattern will be cases where the tokening of these internal events leads to inappropriate responses to different shaped leaves. Such exceptions will involve metabolic costs, and, just like the cases in chapter 6 or inference above, those metabolic costs could suffice to act against those worms engaging inappropriate behavior in virtue of the fact they engaged inappropriate behavior. This would suffice for there to be a norm of performance governing the tokening of these internal states. As these states contribute to the production of appropriate behavior under variable circumstances and are subject to a norm of performance, it is reasonable to conclude that these states genuinely serve that contributory function of mapping the worm's behavior to the relevant circumstances or conditions.

Nonartifactual norms of performance can serve, then, the roles of grounding representing and inferring functions generally as genuine teleo-functions of an entity. I have relied on selectionist history not because there is anything essential about selectionist history to grounding these sorts of norms. Instead, I have done so because mentality and the gross abilities of inferring and representing are more likely phylogenic traits than strictly ontogenic. Selectionist history does provide at least one clear way to establish norms governing phylogenic traits. I want to turn now to the question of intentional content, show that such content is essentially teleological and can be genuinely so through nonartifactual norms of performance.

II. Semantics

a. *Semantics as essentially teleological*

Let's turn to the second question "how can physical facts determine the representational content of a mental state?"

Given the functional work of mental representation and its analogy to natural language, roughly four factors collaborate to determine the content of a tokened representation: its mood, the semantic value of its composite intentional icons, the syntactical and compositional rules, and context. For example, the content of the tokened thought "the bat is flying" is a function (mathematical sense) of its mood (indicative or imperative), the semantic value of 'the', 'bat', 'is', and 'flying', the syntactical and compositional rules composing those icons into a single thought, and the context of the thought (e.g., if I was looking a little to the left at a different bat, the content of my thought might differ.) The naturalist's job, in short, is to explain how non-intentional and non-psychological facts determine these varied factors.

That said, representational content enters the picture in virtue of the fact that intentional icons possess semantic value. Consequently, the primary theoretical hurdle for the naturalist is to account for how intentional icons possess their semantic value. And, that hurdle has at times seemed to present an in principle stumbling block to the naturalistic project. For example, impressed by the fact that "the cow is in the barn" requires for its truth a cow to be in the barn where "Fred believes that the cow is in the barn" does not, Chisholm writes, "the point of talking about "intentionality" is not that there is a peculiar type of "inexistent" object; it is rather that there is a type of psychological phenomenon which is unlike anything purely physical." (1957, 170 ft. 2)[104] Unlike anything purely physical, the semantic value possessed by intentional icons, for Chisholm, cannot suffer a naturalistic treatment. I will concentrate here on just this particular theoretical hurdle (how can physical facts determine the semantic value of intentional icons) and suggest that it can only be crossed by adopting a teleo-functional account of meaning.

As Sellars repeatedly stressed, our primary analogy for inner episodic thought is thinking-out-loud.[105] Thoughts-out-loud are those candid, overt, and spontaneous linguistic acts which do not function as a means of communication (they are not cases of talking to anyone at all, even to oneself in soliloquy). Lacking a communicative function, thoughts-out loud are best understood as functioning as expressions of thought. They are not expressions of thought in the sense of being the causal product of thought. We need not suppose any antecedent intention or other bit of mental machinery leading to their production. Instead, they are expressions of thought in the logical sense of expressing a proposition.[106] It is in virtue of the fact that a thought-out-loud expresses a proposition in

[104] Chisholm's arguments against the possibility of naturalistic explanations of semantic value are largely inspired by Brentano's understanding of intentionality. (See Brentano 1874/1995 & 1930/1966).

[105] 'Thinking-out-loud' and its role in Sellars' verbal behaviorism are a constant throughout his works from "Empiricism and the Philosophy of Mind" (1956) through "Mental Events" (1981). Sellars saw himself consistently misunderstood as asserting that, since such linguistic activity is our primary analogy for inner episodic thought, that such inner episodic thought was in someway parasitic on natural language. Instead, Sellars intended, and I intend to follow him here, that

> In the domain of the mental, language is prior in the order of knowing…. I am not claiming that thoughts-out-loud and propensities pertaining thereto are what thinking primarily is *in the order of being*. I am saying, rather, that the *concept* of thinking-out-loud is our primary *concept* of thinking and is, therefore, our conceptual point of entry into the domain of the mental; as our *concepts* pertaining to the middle-sized objects of the perceptual world are our conceptual point of entry into the domain of the physical. (1981, 326-7)

[106] For a detailed analysis of the difference between the logical and causal sense of the term 'express', see Sellars (1967 & 1969)

the logical sense that such linguistic acts provide some analogy for inner episodic thought. (Lest we fall into a vicious regress, at least some inner episodic thought must express propositions only in the logical sense of the term 'express'.) Further, such thinking-out-loud is an activity within a linguistic system and, as a result, provides an analogy of a representational system as a whole capable of expressing propositions.

To take advantage of that analogy, I want to draw attention to two linguistic platitudes. First, different tokened sentences with differing physical descriptions can express the same proposition. 'Het is een hond', 'C'est un chien,' and 'It is a dog' are each tokened sentences that express the same proposition but differ in their physical description. It is standard to go across languages to make this point, but that move is unnecessary. 'It is a dog', 'That's a dog', and 'Lo, a dog' each express the same proposition. As does "It is a dog," whether spoken, written, signed in American Sign Language, or felt in Unified English Braille Code. This point holds as well one step below complete tokened sentences for their composite intentional icons: different tokened linguistic items with differing physical descriptions can possess the same semantic value. 'hond', 'chien', and 'dog' each mean dog. And, if one is skeptical of inter-linguistic synonymy, then the intra-linguistic synonymy of writing or speaking 'dog' should suffice to make the point. Second, different tokened sentences with the same physical description can express differing propositions. This can be the obvious consequence when indexicals or demonstratives are involved: 'He is here,' 'That is wonderful', 'She will be late,' etc. Again, the point can be made with respect to the composite intentional icons: different tokened linguistic items with the same physical description can possess different semantic value. Indexicals and demonstratives are not the best way to make the case for this point, however. A shift in the proposition expressed by different tokened utterances of 'He is here' is often best explained by appeal to a shift in context and not a shift in the semantic value of the relevant indexical or demonstrative. Synchronically, homonyms make the point, however. Diachronically, the semantic value possessed by linguistic terms is instable. The extensive and relatively rapid schisms in linguistic communities prior to the development of modern media testifies to the instability of the semantic value of linguistic terms. Notice that, since different but descriptively identical tokened linguistic items can differ in semantic value, it is not trivial that a 'cat' in one sentence and a 'cat' in another possess the same semantic value.[107] Putting those two platitudes together for linguistic intentional icons, the physical description of a tokened intentional icon is neither necessary nor sufficient for it to possess a particular semantic value. Descriptively different linguistic tokens might, nonetheless, be synonymous, and linguistic tokens descriptively the same need not be synonymous.

The point of noting those linguistic platitudes is to apply them by analogy to inner episodic thought. Inner episodic thought like thought-out-loud is an activity. Thoughts-out-loud express propositions but are not identical to the proposition expressed, because descriptively varied tokened linguistic activities can have the same content and descriptively identical tokened linguistic activities can possess varied content. (Again, the 'express' of "thoughts-out-loud express" is not its causal sense but its logical sense.) Similarly, inner episodic thoughts express propositions but are not identical to the proposition expressed. The 'similarly' is by force of the analogy: if the physical descriptive character of whole sentences and their composite linguistic intentional icons

[107] Ignoring figurative and colloquial uses as well as slang, 'cat', according to the *OED*, can refer to a feline that is a common household pet, a vessel formerly used in the coal and timber trade, and a shrub of Arabia or its derived narcotic or, when used as a verb, can mean to raise the anchor.

is neither necessary nor sufficient for the possession of a particular content, then so too the physical descriptive character of inner episodic thought is neither necessary nor sufficient for the possession of a particular content. For example, Sellars writes,

> Inner sentence episodes can differ in their descriptive character and yet express the same proposition, just as can overt sentences episodes.
>
> And just as the generically specified character of the shapes and motions and relative locations demanded of chess pieces *must* have determinate embodiment in actual games, so the generically specified character of pieces, positions, and moves, which is common to determinate ways of playing the same conceptual game, must be determinately embodied in the natural order. In other words, while a mental act which expresses the proposition that it is raining is *ipso facto* an •it is raining•[108], it must belong to a specific variety of •it is raining•, just as a token of the corresponding English sentence is not only an •it is raining• but has the specific empirical character by virtue of which it sounds (or reads) like *that*.
>
> The fact that conceptual "pieces" or "role-players" *must* have determinate *factual* character, even though we don't know what that character is save in the most generic way, is the hidden strength of the view which identifies mental acts with neurophysiological episodes. (Sellars 1967, 136-7)

The inclusion of dot-quotes above shows that Sellars has already made the move here to identifying linguistic and mental content functionally. But, for the naturalist to take his point to heart, the naturalist need not yet take on that functional commitment with respect to content. For the naturalist, thoughts are physical events just as linguistic activities are. Unless the naturalist is possessed by a hyper-chauvinism, the naturalist ought to allow that physically different inner episodic events can possess the same content, just as is the case with language. Further, rejecting any hyper-chauvinism, the naturalistic ought to allow that inner episodic events with the same physical description, say, in differing

[108] Sellars uses the device of dot-quotes to provide the function of an expression by means of an illustrating sortal. So, •it is raining• is the function of the sentence "it is raining". See, for example, Sellars (1974, 431).

organisms or at differing time periods, can vary in content, just as is the case with language.

Since the physical description of tokened intentional icons (whether linguistic or mental) is neither necessary nor sufficient for their possession of a particular semantic value, the challenge to the naturalist (as well as to anyone who is not a full-out skeptic about meaning) is to explain in virtue of what are tokened intentional icons tokens of a semantic type.

An obvious strategy for a naturalist is to flesh out the "in virtue of" by appeal to causal mechanical relations. My inner tokening of some intentional icon is caused by a extra-mentalistic event, and that extra-mentalistic event, as a result, determines the content of that tokened intentional icon. For example, I see that a bat is flying, because a bat is flying. If the bat had not been flying, I would not have seen it. The bat caused me to have the perception that I had. The content of non-perceptual representation could afford a similar treatment. While the semantic values possessed by the intentional icons composing a belief need not themselves be directly caused by a present state of affairs, they do possess their semantic value through some indirect causal chain tied to perceptual states. This should have a familiar empiricist ring: the content of non-perceptual representation is causally dependent on perceptual states and is so through some associative or otherwise causal route.

Even with such a thin sketch of that strategy, it has some clear advantages. First, it provides a straightforward naturalistic explanation of the "in virtue of" by relying on causal mechanical facts. Second, it relies on a fact that seems patently true: I see a dog, because there is a dog. If there is perception, then it ought to be the case that my perceptions of objects are causally linked to those objects. Third, my beliefs about empirical matters ought to be linked to perceptual experience, and the above strategy provides a straightforward answer to how they are so linked. Fourth, given the previous, we have a metaphysical explanation for the epistemological worry about how there can be empirical knowledge: in a traditional empiricist voice, the possibility of empirical knowledge is through perception and that perception is the causal product of what is perceived.

Despite those advantages, there are notorious difficulties with such a strategy. The most notorious is that such a causal strategy seems to count out the possibility of error, e.g., misperception, misbelieving, etc. As Chisholm (1957) pushed against earlier versions of this strategy,[109] if the content of an intentional icon is determined by its cause, then the possibility of intentional error is excluded. Simply, if the content of a tokened intentional icon is determined by what caused it to be tokened, then it would not be possible to perceive a cat as a result of looking at a cow; that cow would determine the content of the perception to be of a cow and not of a cat. But, of course, intentional error is possible.

The clear response by the causal strategist is to restrict the causal claim to a claim about normal conditions. That is, under normal conditions, perceptual content is causally determined by the objects of perception. Error results from perception under abnormal conditions. Obviously such a response requires a way to flesh out in naturalistic terms "normal conditions". One might treat the 'normal' of "normal conditions" in its

[109] The two central targets of Chisholm's criticism were linguistic dispositional accounts (such as that of Ayer (1947) and of Carnap (1955) that identified content with a subject's linguistic dispositions when confronted by the represent-ed) and causal accounts (such as Reichenbach's (1947) that straightforwardly identified content with the cause of a particular neurological state).

quantitative sense. As a result, the causal strategy becomes the following: the semantic value possessed by an intentional icon is the statistically normal cause of tokens of that intentional icon type. But, if the semantic value of an intentional icon is tied to what reliably causes tokenings of it, then those tokens are more reliably caused by including both the non-error and error cases. Consequently, an appeal to statistical normalcy has not advanced the ability of the causal strategy to include error cases as error. Further, even assuming sufficient ingenuity to avoid the latter problem, appeal to statistical normalcy suffers a more fundamental problem: the statistically normal cause can be the error case. For example, when the cost of failing to detect some feature is relatively high but the cost of responding to false positives is relatively low, the error case can be the statistically normal case. A deer that engages in predator avoidance behavior in response to each snapping twig is more often than not mistaking the snapping twig to be the sign of a predator. But, on the off-chance that the snapping twig is caused by a predator, the deer by having responded to every snapping twig might be in a better position than if it had failed to do so.[110]

The appeal to normal conditions reflects a sort of tempting mistake when faced with concerns about intentional error. That mistake is to think that all the causal theorist needs is a supplementary theory of error. But, as Fodor (1990, 231) suggests, intentional error is only illustrative of the causal strategy's general inadequacy to provide a theory of content. We can see this by looking to Fodor's formulation of the "disjunction problem":

> Since there are B-caused tokenings of 'A'; it follows that the causal dependence of 'A's upon A's is imperfect; A's are sufficient for the causation of 'A's, *but so too are B's*. If, however, symbols express the properties whose instantiations reliably cause them; it looks as though what 'A' must express is not the property

[110] See both Matthen (1988 & 1989) and Godfrey-Smith (1996 & 2002) for examples of occasions when the error case turns out to be the statistically normal case.

> of *being A* (or the property of *being B*) but rather the *disjunctive property of being (A v B)*. But, if 'A' expresses the property (A v B), then B-caused 'A' tokenings are veridical after all. They're not misrepresentations since, of course, B's *are A v B*. But if B-caused 'A' tokenings are true of their causes, then we don't yet have a theory of misrepresentation.
>
> That's what I call the 'disjunction problem'. We can put it that a viable causal theory of content has to acknowledge *two* kinds of cases where there are disjoint causally sufficient conditions for the tokenings of a symbol: the case where the content of the symbol is disjunctive ('A' expresses the property of *being (A v B)*) and the case where the content of the symbol is not disjunctive and some of the tokenings are false ('A' expresses the property of *being A*, and B-caused 'A' tokenings misrepresent). The present problem with the Crude Causal Theory is that it's unable to distinguish between these two cases; it always assigns disjunctive content to symbols whose causally sufficient conditions are themselves disjoint. (Fodor 1987, 101-2).

The two cases that the causal strategy is unable to distinguish are not the cases of error and non-error but are rather the cases of disjunctive and non-disjunctive content. The crude causal theory, as Fodor calls it, is inadequate as a theory of content not merely because of the possibility of error but, importantly, because it is unable to assign any non-disjunctive content or semantic value to intentional icons. Crude causal accounts can serve to underwrite a theory of natural information, but what the disjunction problem demonstrates is that natural information is fundamentally different than content. As Fodor explains,

> Information is tied to etiology in a way that meaning isn't. If the tokens of a symbol have two kinds of etiologies, it follows that there are two kinds of information that tokens of that symbol carry. (If some "cow" tokens are caused by cows and some "cow" tokens aren't, then it follows that some "cow" tokens carry information about cows and some "cow" tokens don't.) By contrast, *the meaning of a symbol is one of the things that all of its tokens have in common, however they may happen to be caused.* All "cow" tokens mean *cow*; if they didn't they wouldn't be "cow" tokens.
>
> So, information follows etiology and meaning doesn't, and that's why you get a disjunction problem if you identify the meaning of a symbol with the information that its tokens carry. Error is merely illustrative; it comes into the disjunction problem only because it's so plausible that the false tokens of a symbol have a

> different kind of causal history (and hence carry different information) than the true ones. (Fodor 1990, 231)

Fodor suggests that the causal strategy can be salvaged, nonetheless. What is needed is a way to break, in nonintentional and nonsemantic terms, the symmetry between cow caused 'cow' tokens and non-cow caused 'cow' tokens. The key to breaking that symmetry is to recognize, Fodor thinks, that the error cases are ontologically parasitic on the non-error cases.

> It's an observation – as old as Plato, I suppose – that falsehoods are *ontologically dependent* on truths in a way that truths are not ontologically dependent on falsehoods. The mechanisms that deliver falsehoods are somehow *parasitic on* the ones that deliver truths. (Fodor 1987, 107)

That ontological dependence shows, Fodor suggests, that the counterfactual properties of the causal relations of the error and non-error cases are not the same. (If so, then the apparent symmetry that drew us into the disjunction problem is merely apparent.) Fodor suggests that we retain the basic feature of the crude causal theory, namely 'cow' tokens mean *cow*, because cows cause 'cow'-tokens. This does not, he suggests, count out falsehood but rather has nothing to say about it, because it is statement of what is required for a non-error token. Since error tokens are ontologically dependent on non-error tokens, non-cows cause 'cow's insofar cows cause 'cow's. Non-cow caused 'cow' tokens are asymmetrically dependent on the existence of the semantic setup between cows and 'cow' tokens.

> If B-caused 'A' tokenings are wild – if they falsely represent B's as A's – then there *would be* a causal route from A's to 'A' even if there *were no* causal route from B's to 'A's; but there would no causal route from B's to 'A's if there were no causal route from A's to 'A's. (Fodor 1987, 108)

Since error tokens are dependent on the prior semantic setup of the non-error cases, we can specify a non-disjunctive content by specifying that semantic setup (say, between cows and ‘cow’ tokens). We maintain error, because tokens in error do not get their semantic value from their causes. They get their semantic value, because of the causal set up between, say, cows and ‘cow’ tokens. Pressed by a variety of issues that need not concern us here,[111] Fodor does later recharacterize the thesis as one about the nomic relation obtaining between the property of being a cow and the property of being a ‘cow’ token. (See Fodor 1990)

Fodor’s thesis will not do, however. I want to suggest that Fodor is faced with a dilemma. On one horn, Fodor must presuppose the semantic value that his thesis purports to explain. If so, Fodor has failed to do just what he intended, namely to characterize in nonsemantic and nonintentional terms the circumstances which fix the content of an intentional icon. On the other horn, if Fodor does not presuppose semantic value, then he does so at the cost of eliminating error. In short, Fodor has, despite appearances, made no progress over the earlier causal accounts criticized by Chisholm.

The essence of Fodor’s suggestion is the following: ‘Cow’-tokens mean *cow*, because cows cause ‘cow’-tokens and, if non-cows cause ‘cow’ tokens, that latter nomic relation is asymmetrically dependent on the fact that cows cause ‘cow’ tokens. What, however, makes it the case that some tokened intentional icon is a token of the type “cow”? The obvious answer is that tokened intentional icons are tokens of the type “cow”, because they all mean *cow*. Fodor says exactly that: “All “cow” tokens mean *cow*; if they didn’t they wouldn’t be “cow” tokens.” (Fodor 1990, 231) Fodor is here quite reasonably individuating intentional icon types in virtue of their semantic value. But, if

[111] There are a variety of issues that need to be resolved here, and Fodor tries to work through a number of them, e.g., the content of theoretical intentional icons or the content of terms like ‘unicorn’. (See Fodor 1987 & 1990) I will concentrate just on this core statement of his thesis, because that thesis is at its core already unworkable independent of these further issues.

he is in fact doing so, then he has failed to express in nonsemantic and nonintentional terms the circumstances that determine content. The nomic relation obtaining between the property of being a cow and the property of being a "cow" token is a relation involving a semantic property, namely the property of being a "cow" token. Whatever nomic relation obtains between cows and "cow" tokens, that relation cannot determine the content of the "cow" token, because for the relation to obtain we already have to have intentional icons with the semantic value of cow: "All "cow" tokens mean *cow*; if they didn't they wouldn't be "cow" tokens."

To escape that problem, Fodor does not need to abandon the claim that all "cow" tokens mean cow. Fodor just needs to recharacterize his thesis in such a way that the relevant intentional icon type standing in a nomic relation with cows does not presuppose semantic value. Suppose some nonsemantic principle of individuation for intentional icon types, we could reformulate Fodor's thesis as follows: cows cause tokens of type 'x', and if non-cows cause tokens of type 'x', that latter nomic relation is asymmetrically dependent on the fact that cows cause tokens of type 'x'. Tokens of type 'x' possess their semantic value in virtue of the relevant nonsemantic and nonintentional nomic relation, and consequently such tokens are "cow" tokens. Further, the possibility of error is retained, because asymmetric dependence is retained. But, what principle of individuation for intentional icon typing would do for the nonsemantic reformulation?

As naturalists, we must say that intentional icons have, as Sellars said, some determinate embodiment – that is, these mentalistic intentional icons are physical states. So, we might individuate intentional icon types in virtue of their physical description. As with the sign design of written linguistic items, we might individuate various internal physical state types in virtue of descriptive facts, e.g., individuating differing patterns of neural activity and so on. To avoid worries about what makes some pattern of neural activity or otherwise internal event a tokening of an intentional icon,[112] let's just adopt the convention, again relying on the analogy from language, that certain arbitrary strings of letters provide the physical description of an inner intentional icon type. For example, 'efft', 'dfft', and 'shp' are each types of inner intentional icons typed in virtue of their mental sign design. Fodor's thesis would become: cows cause tokens of 'efft', and if non-cows cause tokens of 'efft', that latter nomic relation is asymmetrically dependent on the fact that cows cause tokens of 'efft'. However, just as with language, inner intentional icon tokens of differing descriptive types can possess the same semantic value. Fodor's thesis should be reformulated again:

> cows cause tokens of 'efft', and if non-cows cause tokens of 'efft', that latter nomic relation is asymmetrically dependent on the fact that cows cause tokens of 'efft'.
> *and*
> cows cause tokens of 'dfft', and if non-cows cause tokens of 'dfft', that latter nomic relation is asymmetrically dependent on the fact that cows cause tokens of 'dfft'
> *and*
> cows cause tokens of 'shp', and if non-cows cause tokens of 'shp', that latter nomic relation is asymmetrically dependent on the fact that cows cause tokens of 'shp'
> *and so on.*

Fodor should not, I think, object to this last reformulation. Tokens of 'efft', 'dfft', and 'shp' each possess the semantic value of cow, because the 'efft', 'dfft', and 'shp' types stand in the appropriate nomic relation to cows. Tokens of 'efft', 'dfft', and 'shp' are all then "cow" tokens, because each means cow. A token of , say, 'efft' caused by a non-cow

[112] Such worries ("why think that neural pattern x or what have you is an intentional icon in the first place?") are not trivial. In fact, I think the only way to answer them is to adopt a teleo-functional account of the mental. I suggest that we set them aside here, because the problem with Fodor's thesis runs far deeper than those concerns alone.

is an erroneous tokening, because such tokens asymmetrically depend on the fact that cows cause ‘efft’s.

However, not only can linguistic tokens of differing sign design types possess the same semantic value, but linguistic tokens of the same sign design type can vary in semantic value (e.g., a tokened ‘bank’ with the semantic value of a financial institution and another tokened ‘bank’ with the semantic value of a river’s edge). Similarly, differing tokens of an inner sign design type might differ in semantic value. So, it can be the case, say, that cows, pigs, and chickens each suffice to cause tokens of the type ‘dffts’. A tokened ‘dffts’ need not be semantically ambiguous or disjunctive, no more than a tokened ‘bank’ need be ambiguous or disjunctive. Some tokens of ‘dffts’ possess the semantic value of cow, others of pig, and others of chicken. None of these tokened cases caused by their respective cows, pigs, or chickens need be errors. Lest we fall directly back into the disjunction problem, we cannot continue to formulate Fodor’s thesis as a nomic relation obtaining between the property of, say, a cow and some physical description type like ‘dffts’. If we are to maintain Fodor’s thesis, we need a different principle of individuation for the intentional icon types standing in the appropriate nomic relation to extra-mental bits of the world. Excluding an appeal to semantic value or physical descriptive types, there is, as far as I can tell, one further principle of individuation open to us.

Here is the situation as I see it. Tokens of a semantic type, say, “cow” can be tokens of different descriptive types, say, ‘efft’, ‘dffts’, ‘shp’, etc. Tokens of a descriptive type, say, ‘efft’ can vary in semantic value and thus are tokens of different semantic types, say, “cat”, “pig”, “chicken”, etc. Fodor’s aim is to state in nonsemantic and nonintentional terms a metaphysical relation sufficient to determine semantic value. Further, that relation should be a nomic or causal relation between the represent-ed and the represent-ing. A semantic type is the type that it is, because of a nomic relation that obtains between some extra-mental bit of the world and some nonsemantically characterized type of intentional icon. In virtue of that nomic relation, tokens of that nonsemantically characterized intentional icon type are thereby tokens of the semantic type. The only nonsemantic principle of individuation that seems to remain, having dropped an appeal to the physical description, is to type intentional icons in virtue of their cause. So, returning to the earlier formulation "cows cause tokens of type 'x', and if non-cows cause tokens of type 'x', that latter nomic relation is asymmetrically dependent on the fact that cows cause tokens of type 'x'," type 'x' are just those intentional icons caused by cows. However, if membership in type 'x' is just in virtue of being at the tail-end of a causal route from, say, cow, then there just cannot be non-cow caused tokens of type 'x'; hence, no token in the semantic type "cow" can be tokened in error. Asymmetric dependence is impossible given the last principle of individuation.

What has gone wrong for Fodor is what must inevitably go wrong for any version of the causal strategy.[113] The project is to provide a naturalistic account for the possession of semantic value by some tokened intentional icon. Given the analogy to language, we can understand that project as the provision of a naturalistic explanation for why tokened intentional icons are all tokens of a semantic type or, alternatively, what is the naturalistic

[113] Ironically, Fodor himself diagnoses what is the heart of the trouble for his own strategy: "If the tokens of a symbol have two kinds of etiologies, it follows that there are two kinds of information that tokens of that symbol carry." In suggesting asymmetric dependence, Fodor seems to forget his own cautionary advice that error is only illustrative, because an obvious way in which there might be two kinds of etiologies and, thus, two kinds of information is with homonyms.

basis for semantic typing. The causal strategy, with or without the amendment of asymmetric dependence, tells us that semantic type "x" is an intentional icon type-α standing is some causal relation with x's. Intentional icon type-α clearly cannot itself be semantically characterized given the aim of the causal strategy. The causal strategist can provide some descriptive and nonsemantic characterization of intentional icon type-α . If that description is not just the statement of causal dependence (e.g. "type-α is whatever is caused by x"), then sign design provides a clear analogy for what those descriptions would be like. Suppose type-α is so nonsemantically characterized. Even if a causal/nomic relation obtains between type-α and x, that relation cannot be necessary for content, with or without Fodor's asymmetric dependence thesis, because of descriptively varied synonyms. Further, even if a causal/nomic relation obtains between type-α and x, that relation cannot suffice for content, with or without Fodor's asymmetric dependence thesis, because of homonyms. Coupling homonymy and a nonsemantic description of type-α within the causal strategy straightforwardly produces a disjunction problem. Unless homonymy is not possible, that a causal/nomic relation obtains between type-α and x cannot suffice for content. If the description of type-α is causally dependent (i.e. "type-α is whatever x causes"), then there cannot be erroneous tokens of the semantic type. The causal strategy is a dead end.

Up until now, I have pushed the point that, from the analogy with language, different "cat" tokens might be members of differing physical descriptive types and tokens of the same physical descriptive type might be members of differing semantic types. But, I have ignored the fact that some tokens of the same physical descriptive type might possess a semantic value while others do not. To push past the causal strategy and start to get things right, it is to this latter fact that I think we ought to attend.

A phonological tokened ‘cat’ might or might not be a member of a semantic type. A phonological tokened ‘cat’ produced by a breaking ice shelf or pneumatic door opening is no intentional icon, though it is a token of the same descriptive type as my spoken ‘cat’ in “There is a cat.” Or, take a pattern of neural activity of, say, descriptive type ‘efft’, some tokened ‘efft’s might be intentional icons and others not. For example, removal of the relevant neurons from the brain and activation of the pattern in a laboratory setting would produce a token of ‘efft’. It would be difficult to maintain that this ‘efft’ is an intentional icon for the same reason that it would be difficult to maintain that the phonological ‘cat’ produced by a breaking ice shelf is an intentional icon. Unless we fall prey to mysticism, each tokened intentional icon has a determinate factual character and is a token of some descriptive type, but that descriptive type will not suffice to individuate which are intentional icons and which are not. (No particular descriptive type, it should be clear, will be necessary for intentional icon typing either.)

Obviously, the mode of production of the different phonological ‘cat’ tokens above varies, but it would be wrong to think that some mode of production suffices to render some descriptive token into an intentional icon. For example, let’s assume that Larry is stranded on a deserted island and is apt to make linguistic observation reports. His reports are not reportings for the purpose of communication but are examples of thinking-out-loud. Larry at times suffers bouts of deafness. He will utter, for example, ‘there is an apple’ and at times hear what he has said and at other times not. In either case, the mode of production or causal route leading to his utterance ‘there is an apple’ is the same. When Larry has heard what he has said, a number of further events or activities might occur with respect to Larry, e.g., he might go on to engage in a bit of practical reasoning terminating in his reaching for and consuming the apple. There is a positive reason to think that his utterance is an example of a thinking-out-loud and that its composite sounds are intentional icons, because the best explanation of Larry's subsequent cognitive activity is that his utterance is a thinking-out-loud. When Larry fails to hear what he has said, his utterances are in an important respect inert. When suffering deafness, the production of a sequence of sounds is the end of things with respect to Larry. There is no positive reason to think that that sequence of sounds is a sequence of intentional icons. That the same mode of production is in place when Larry is and is not deaf is no reason to think his utterances while deaf are thoughts-out-loud. The reason to think that some of his utterances are thoughts-out-loud is their potential causal contribution to further cognitive activity. How those thoughts-out-loud are produced is irrelevant to why they can be rightly understood as thoughts-out-loud. What is relevant is not the mode of production but the potential subsequent consumption of those thoughts-out-loud by Larry in cognitive activity. The shared mode of production in the deaf and non-deaf cases is not, then, relevant to the representational status of those utterances made while Larry was deaf. The very reason to consider the non-deaf utterances to be thoughts-out-loud is what is missing when Larry is deaf. His utterances in those deaf cases are inert. They are just sounds at the end of a causal chain, just as the earlier phonological 'cat' token produced by an ice shelf is just a sound at the end of a causal chain.

Nothing about a descriptive type nor about the mode of production of some token of a descriptive type suffices to render type or token into intentional icons. What does suffice, we just saw, is when the downstream effects of that token are best explained by the fact that the token is an intentional icon. What is fundamentally wrong about the causal strategy and any production side account is that semantic typing is robbed of its explanatory value. Just as when Larry is deaf, there is just no positive explanatory reason, no matter the causal mechanical or production-side facts that obtain, to think the generated descriptive token is an intentional icon. It is in the explanation of the consumption or use of some descriptive token within the economy of a system that semantic typing performs explanatory work. Just as with Larry when he is not deaf, we have a positive explanatory reason to think his tokened 'apple' is an intentional icon given the sorts of subsequent cognitive behavior it enables. More generally, semantic attribution or ascription to some descriptive token performs explanatory work in explaining the pattern of tokens of that descriptive type within the cognitive behavior of the system. There is a pattern of phonological 'apple' tokens within a range of behavior exhibited by Larry; it is the explanation of that behavior and explaining the role of those phonological 'apple' tokens within that behavior that provides reason to think that they possess semantic value. Or, there is a pattern of phonological tokens of 'cat' among English speakers. That pattern of tokens is not inert but is involved in a range of behavior. The reason to think that those tokens are tokens of a semantic type "cat" is that their being members of that semantic type aids to explain the sorts of behavior in which such tokens appear or appear to enable. It is in the explanation of such patterns of consumption or use that semantic value finds explanatory purchase, and it is for this reason that attending to production-side facts alone will just fail to provide reason to think that the relevant descriptive token is an intentional icon at all.[114]

Given the explanatory role of semantics, it is at least true that semantic typing is teleo-functional. Semantic attribution or ascription is in the service of explaining how some token of a descriptive type contributes to cognitive behavior. A descriptive token is a token of a semantic type given the contributory effect that tokens of that descriptive type have within some system's cognitive economy. Further, for reasons similar to those in the first section with respect to the gross architecture of the mental, semantic typing is at most teleo-functional. Synonymy demonstrates that semantic types are multiply realized by differing physical structural types. A semantic type is not type identical to any structural type. The inverse of multiple realizability holds as well. Homonyms are a case in point. A type of spoken term or neurological state can perform different semantic jobs. Moreover, a token of, say, a phonological type or neurological state type need not be a token of a semantic type at all – that is, there is nothing about its physical character as such that requires that it perform any semantic work. Again, a semantic type is not type identical to any physical type. And, the reason these multiple realizability considerations hold here for semantics are the same reasons that they operated with respect to the gross architecture of mentality. The empirical grounds to think some item is an intentional icon and has semantic value are not the empirical grounds to think some token is a member of a physical type.

Now, Fodor, for example, does not deny that semantic typing is teleo-functional. Instead, that it is so teleologically characterized is what he views as part of the problem. Fodor (1987 & 1990) repeatedly emphasizes that the naturalist is not successful until we

[114] The above claims are largely inspired by Millikan's (1989a) criticism of production-side accounts.

have a nonsemantic, nonintentional, *nonnormative,* and *nonteleological* account. But, part of the point of going somewhat painfully through all the preceding steps is to make it clear that we have no handle on content that isn't teleological. We do have a straightforward handle on semantics from teleology: to specify a semantic value is to specify the contributory effect of a token of some descriptive type to the cognitive economy of the system. If nonartifactual teleology was merely ersatz teleology, then Fodor would be quite right that we would need to clear away this heuristic to get at the naturalistic facts. Though if it were merely ersatz teleology, the failure of the causal strategy shows that we ought to concede that talk of semantics is merely a heuristic. But, the point of the preceding chapters has been to show that we can have just the sort of nonartifactual teleology required. So, the simple answer that 'cat', 'cat', 'cat', and 'cat' are all instances of the same word insofar as they perform the same work of representing *cat* should no longer be seen as part of the hurdle that the naturalist needs to get past.

b. *A normative foundation for semantics*

Back then to the question at hand: how can physical facts determine the semantic value of intentional icons? Well, if a semantic type is teleo-functional, then some tokened physical state has some particular semantic value insofar as it serves the relevant function. That is, a tokened phonological 'apple' means apple, when its function is to represent apple. The question concerning semantic value is just a more specific instance of the more general question: how can the physical facts determine the function of some item? And, the general answer to that question is that some item or behavior (artifact or otherwise) has a function insofar as it is held to some standard of performance. That a tokened ‘apple’ serves the function of representing apple reflects the fact that it is held to a standard of semantic performance. The physical facts that we should look to, then, are just those sorts of facts that ground any norm of performance.

If a token of a descriptive type is subject to a norm of semantic performance, then we need a sense of what could constitute correct/incorrect use of an intentional icon. To form a false representation is not to misuse an intentional icon. “Six apples are in the jar” can only be false if each of its composite terms are used correctly; I cannot have a false belief about the apples unless the tokened ‘apples’ means apples. Misrepresentation is a representational failure, but it is not a failure in respect to the composite intentional icons. Rather, paradigmatic cases of misuse in language are cases of misspeaking, e.g., saying “That is reverent” when intending to say “That is relevant” or calling a chicken a “cat”. The ‘reverent’ and ‘cat’ are misuses of an intentional icon, because the speaker is trying to get them to perform a task they do not serve. That is, what explains the pattern of ‘cat’ tokenings among English speakers is not the representing function of chicken. Given the prior suggestions, the stable pattern of use in respect to ‘cat’ tokens is explained by the fact that, when ‘cat’ tokens are not used to serve the function of representing cats, speakers are subject to modification in their use. Both inner episodic thought and thinking-out-loud involving misuse of intentional icons should be analogous to these sorts of misspeaking.

So, for example, we have a toad that exhibits compensatory predatory behavior. In that toad, there is a pattern of neurological events; call them ‘efft’s. Those neurological events play a contributory role in orienting the toad in the direction of prey and leading to the engagement of predatory behavior. Given the way the toad uses those ‘efft’s to generate predatory behavior, the contribution of 'effts', we might say, is to represent food. The hypothesis is that 'efft's have the semantic value of food, because the explanation of how the toad's cognitive engine makes use of 'efft's seems best explained by their contributing the representing function of food; alternatively, the best explanation for why there is a pattern 'efft's in our toad is that the toad uses them to serve the semantic role of food. If that is the best explanation of the pattern of 'efft's, then the toad needs to be sensitive to that use. It would be so sensitive if an appropriate normative regularity was in place, i.e., the pattern of consumption of 'efft' tokens by the toad was explained by the fact that 1) failing to use 'efft' correctly is subject to modification so as to use it correctly and 2) its being so subject to modification is in the virtue of that fact it is used incorrectly. Failures to use 'efft' correctly, however, are not to be found by pointing to occasions in which the toad tokens false representations. Again, if the toad has a false representation about food, then the tokened 'efft' better mean food. So, the toad's sensitivity or lack of it to the truth or falsity of its representations is not what will decide the content of its tokened 'efft'. Instead, we need something analogous to misapplying a term.

For example, let's say that our toad at times will use 'efft' to prompt mating behavior and, further, that it will orient its mating behavior in the direction of the perceived 'efft'. The consequence of this consumptive use is that our toad on these occasions will try and mate with a worm. Whatever is going on in the toad on such occasions the explanation is not likely to be that it representing falsely that there is food. Since toads are relatively simple instinctual critters, we can suppose that our toad is in no way sensitive to its successes or failures in respect to acquiring sustenance or successful mating. But this ontogenic insensitivity might be overcome by looking to the phylum. So, let's alter the target of explanation to why there is a pattern of use by toads of 'efft's. We can, of course, individuate two differing patterns of use. Since we have moved to the phylum, we should be able to assume that the relative frequency to engage in one or the other use reflects some genetic factors in our toads. Sometimes our toads use 'efft's in directing their predatory behavior and other times in directing mating behavior with worms. When they use 'efft's in directing predatory behavior, toads overall enjoy more energetic benefits than costs. It might be the case that, when toads engage in predatory behavior using 'efft's, toads fail to acquire sustenance most of the time. But, even if that is the case, let's suppose that the energetic costs of engaging in failed predatory behavior are outweighed by its occasional success. In contrast, using 'efft' for mating behavior comes with energetic costs, and those costs clearly do not garner reproductive benefits. Those toads that use 'efft's to engage in mating behavior are relatively less fit than those that use 'efft's primarily or exclusively to engage in predatory behavior. Even supposing that the energetic costs of a failed mating and a failed predatory behavior are the same, engaging in a mating behavior determined to fail compared to engaging predatory behavior with possible energetic benefit can suffice to make the use of 'efft' for mating reduce the relative fitness of those toads. Notice, then, that the reduced relative fitness is because the respective toads are not using 'efft's to engage in predatory behavior. Everything is in place for a normative pattern.

There is a pattern of use of 'efft's in toads with respect to predatory behavior. That pattern is inherently instable because of the possibility of genetic mutants that use 'efft's differently. However, the pattern of using 'efft's for the engagement of predatory behavior is as stable as it is, because exceptions to that pattern are subject to elimination in virtue of failing to use ‘efft’s for predatory behavior. By using ‘efft’ to engage in mating behavior, the toad is reducing its prospects for food and those reduced prospects explain its reduced relative fitness. So, if there is a selective regime in place, it is in virtue of failing to use ‘efft’s for predatory behavior that is causing the relevant toad lines to be eliminated from the population. As such, there is a standard in place governing the use of ‘efft’s. Toads that use ‘efft’s for predatory behavior meet that standard. Toads using ‘efft’s for mating behavior are failing to meet that standard of use; they are misusing ‘efft’ in way that is analogous to my misusing the term ‘cat’ when trying to order a steak by using the expression “I would like a cat.”

I intend the above only to be illustrative for how the norms governing intentional icons might work. The intent is just to show that we can specify the appropriate nonartifactual norms required. I do not intend the above to suggest that things will be as simple as with our toad. Specifying the normative patterns governing fictional entities, theoretical entities, or otherwise abstract thought all will be dramatically more complex and subtle. It is the in principle issue that I wanted to address here, and that is, I think, satisfied: 1) semantically-laden explanations of the cognitive operation of descriptive tokens are true when those tokens are bound by a nonartifactual norm of semantic performance; and, 2) there is such a norm when the pattern of use of a descriptive type is best explained by the fact that exceptions to that use are subject to modification in virtue of being exceptions.

My aim here has been to give some reason to think that both the gross individuation of mental states and more particularly semantics is essentially teleological. In both cases, the primary reason reflected the fact that the explanatory value in invoking mentality and semantics lay in the contribution mentality and semantics make to some further process. Further, I have aimed to show that the nonartifactual norms of performance developed in earlier chapters could ground such teleo-functions as genuine. However, in both cases, the above is only a first quick sketch for how things might go. That sketch, I hope, suffices to reject that there is any in principle reason to doubt the possibility of genuine nonartifactual teleology in respect to the mental generally and semantics more particularly. Further, my hope is that that sketch is thick enough to see how we might proceed in naturalizing mentality in a way that respects the normativity of inference, belief, semantics, and so on. Filling out that sketch is the hard work still to be done.

BIBLIOGRAPHY

Ache, B.W., 1982, “Chemoreception and thermoreception,” in *The Biology of Crustacean,* vol. 3., eds., Atwood, H. & Sandeman, D., New York: Academic Press, 369-98.

Adams, F., 1979, “A Goal-State Theory of Functional Attributions,” *Canadian Journal of Philosophy,* 9, 493-518.

Ahmadjian, V., 1993, *The Lichen Symbiosis*, New York: John Wiley and Sons.

Ariew, A., 2002, “Platonic and Aristotelian Roots,” in *Functions: New Essays in the Philosophy of Psychology and Biology,* eds., Ariew, A., Cummins, R., & Perlman, M., Oxford: Oxford University Press, 7-31.

Ayala, F., 1970, “Teleological Explanations in Evolutionary Biology,” *Philosophy of Science*¸ 37, 1-15.

Ayer, A., 1947, *Thinking and Meaning: Inaugural Lecture*, London: H.K. Lewis.

Baker, J. & Bailey, E., 1987, “Sources of Phenotypic Variation in the Separation Call of northern bobwhite (*Colinus virginianus*),” *Canadian Journal of Zoology*, 65, 1010-5.

Barluenga, M., Stölting, K., Salzburger, W., Muschick, M., and Meyer, A., 2006, “Sympatric speciation in Nicaraguan crater lake cichlid fish,” *Nature*, Feb. 9, 439, 719-23.

Barnett, S., 1958, “Experiments on ‘neophobia’ in wild and laboratory rats,” *British Journal of Psychology*, 49, 195-201.

Baum, D., 1998, “Individuality and Existence of Species Through Time,” *Systematic Biology*, 47, 641-53.

Bauer, A., Dwyer-Nield, L, & Malkinson, A., 2000, “High cyclooxygenase 1 (COX-1) and cyclooxygenase 2 (COX-2) in mouse lung tumors,” *Carcinogensis*, 4, 543-50.

Bauer, R., 1975, “Grooming behavior and morphology of the caridean shrimp *Pandalas danae* Stimpson (Decapoda: Nanatia: Pandaldae)” *Zoological Journal of the Linnean Society*, 56, 45-71

_____, 1977, “Antifouling adaptations of caridean shrimp (Crustacea: Decapoda: Caridea): Functional morphology and adaptive significance of antennular preening by third maxillipeds,” *Marine Biology*, 40, 261-76.

_____, 2004, *Remarkable Shrimps: adaptations and natural history of the Carideans*, University of Oklahoma Press.

Bechtel, W., 1984, "The evolution of our understanding of the cell: A study in the dynamics of scientific progress," *Studies in the History and Philosophy of Science*, 15, 309-56.

_____, 2006, *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*, Cambridge: Cambridge University Press.

_____, in press, "Biological Mechanisms: Organized to Maintain Autonomy"

Beckers, R., Holland, O., & Deneubourg, J., 1994, "From Local Actions to Global Tasks: Stigmergy and Collective Robotics," in *Artificial Life 4*, eds., Brooks, R., & Maes, P., Cambridge, MA: The MIT Press.

Beckner, M., 1959, *The Biological Way of Thought*, New York: Columbia University Press.

Bedau, M., & Packard, N., 2003, "Evolution of evolvability via adaptation of mutation rates," *Biosystems*, 69, 143-162.

Bergson, H., 1907, *Evolution Créative*, Paris: Alcan.

Björnhag, G., 1989, "Sufficient fermentation and rapid passage of digesta. A problem of adaptation in the hindgut." *Acta Veterinaria Scandinavica,* 86 (Suppl.), 204-11.

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

_____, 2002, "A Rebuttal on Functions," in *Functions: New Essays in the Philosophy of Psychology and Biology,* eds., Ariew, A., Cummins, R., & Perlman, M., Oxford: Oxford University Press, 63-112.

Bousso, R., & Polchinski, J., 2004, "The String Theory Landscape," *Scientific American*, 291.3: 78-88.

Brandon, R., 1984, "Grene on mechanism and reductionism: More than just a side issue," *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2: 345-53.

_____, 1990, *Adaptation and Environment*, Princeton, NJ: Princeton University Press.

Braithwaite, R., 1953, *Scientific Explanation*, Cambridge: University Press.

Brentano, F., 1874/1995, *Psychology from an Empirical Standpoint,* eds., Crane, T., & Wolff, J., New York, NY: Routledge.

_____, 1930/1966, *The True and the Evident*, ed. & trans., Chisholm, R., London: Routledge & Kegan.

Brittan, G., 1999, "The Secrets of Antelope," *Erkenntis*, 51: 59-77.

Brooks, R., 1997, "Intelligence without Representation," in *Mind Design II: Philosophy, Psychology, and Artificial Intelligence*, ed., Haugland, J., Cambridge, MA: The MIT Press, 395-420

_____, 1999, *Cambrian Intelligence*, Cambridge, MA: The MIT Press.

Bürger, R., and Schneider, K., 2006, "Intraspecific Competitive Divergence and Convergence under Assortative Mating," *The American Naturalist*, 167.2, 190-205.

Butts, R., 1990, "Teleology and the Scientific Method in Kant's *Critique of Judgment*," *Nous*, 24, 1-16.

Byers, J., 2002, "The Ungulate Mind," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds., Bekoff, M., Allen, C., & Burghardt, G., Cambridge, MA: The MIT Press.

Carnap, R., 1955, "Meaning and Synonymy in Natural Languages," *Philosophical Studies*, 4, 33-47.

Chaitin, G., 1975, "Randomness and Mathematical Proof," *Scientific American*, 232: 47-52.

Chiarello, A., 1998, "Diet of the Atlantic forest maned sloth *Bradypus torquatus*," *Journal of Zoology London*, 246, 11-9.

Chisholm, R., 1957, *Perceiving a Philosophical Study*, Ithaca, NY: Cornell University Press.

Clark, A., 1989, *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, MA: MIT Press.

_____, 1997, *Being There*, Cambridge, MA: MIT Press.

Clauss, M., 2004, "The potential interplay of posture, digestive anatomy, density of ingesta, and gravity in mammalian herbivores: why sloths do not rest upside down," *Mammal Review*, 34, 241-5.

Cleveland, L, 1923, "Symbiosis between Termites and Their Intestinal Protozoa," *Proceedings of the National Academy of Science*, 12, 424-8.

Crist, E., 2002, "The Inner Life of Earthworms: Darwin's Argument and Its Implications," in in *The Cognitive Animal: Empirical and Theoretical Perspectives on*

Animal Cognition, eds., Bekoff, M., Allen, C., & Burghardt, G., Cambridge, MA: The MIT Press, 3-8.

Cummins, R., 1975, “Functional Analysis,” *The Journal of Philosophy*, 72, pp. 741-60.

_____, 1983, *The Nature of Psychological Explanation*, Cambridge, MA: MIT Press.

_____, 2002, “Neo-Teleology,” in *Functions: New Essays in the Philosophy of Psychology and Biology,* eds., Ariew, A., Cummins, R., & Perlman, M., Oxford: Oxford University Press, 157-172.

Darwin, C., 1881/1985, *The Formation of Vegetable Mould, through the Action of Worms with Observation of Their Habits*, Chicago: University of Chicago Press.

Davies, P.S., 2001, *Norms of Nature: Naturalism and the Nature of Functions*, Cambridge, MA: The MIT Press.

Davidson, D., “Actions, Reasons, and Causes,” *The Journal of Philosophy*, 60.23: 685-700.

Dennett, D., 1978, *Brainstorms*, Cambridge, MA: MIT Press.

_____, 1987, *The Intentional Stance*, Cambridge, MA: MIT Press.

_____, 1991, “Real Patterns,” *The Journal of Philosophy*, 88, 27-51.

_____, 2005, “Author Meets Critics: Max Bennett & Peter Hacker, *Philosophical Foundations of Neuroscience*,” American Philosophical Association, Eastern Conference Division, New York, Dec 27-30, 2005.

Dretske, F., 1986, “Misrepresentation,” in *Belief: Form, Content and Function*, Bogdan, ed., Oxford: Clarendon Press.

_____, 1995, *Naturalizing the Mind*, Cambridge, Ma.: MIT Press.

_____, 1999, “Machines, Plants, and Animals: The Origins of Agency,” *Erkenntis*, 51: 19-31.

Driesch, H., 1909, *Philosophie des Organischen*, Leipzig: Quelle und Meyer.

Edelman, G., 1987, *Neural Darwinism: The Theory of Neuronal Group Selection*, New York: Basic Books.

_____,1992, *Bright Air, Brilliant Fire*, New York: Basic Books.

Farrar, J., 1976, "The Lichen as an Ecosystem: observation and experiment," in *Lichenology: Progress and Problems*, eds., Brown, D., Hawksworth, D., & Baily, R., London: Academic Press, 385-406.

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

_____, 1990, "A Theory of Content II: The Theory," in *Theory of Content and Other Essays*, Cambridge, MA: The MIT Press; reprinted in *Mind and Cognition: An Anthology,* 2nd edition, ed. Lycan, W., Oxford: Blackwell (1999), 230-49.

Fragaszy, D. & Perry, S., 2003, "Towards a Biology of Traditions," in *The Biology of Traditions: Models and Evidence*, eds., Fragaszy, D. & Perry, S., Cambridge: Cambridge University Press.

Frankena, W., 1950, "Obligation and Ability" in *Philosophical Analysis*, ed. Black, M., Ithaca: Cornell University Press, 157-75.

Galef, B., 1970, "Aggression and Timidity: Responses to Novelty in Feral Norway Rats," *Journal of Comparative and Physiological Psychology*, 75, 358-62.

_____, 1992, "The Question of Animal Culture," *Human Nature*, 3, 157-78.

______, 2003, " 'Traditional' foraging behaviors of brown and black rats (*Rattus norvegicus* and *Rattus rattus*)," in *The Biology of Traditions: Models and Evidence*, eds., Fragaszy, D. & Perry, S., Cambridge: Cambridge University Press.

Galef, B., & Allen, C., 1995, "A New Model System for Studying Animal Traditions," *Animal Behaviour,* 50, 705-17.

Galef, B, & Clark, M., 1971, "Parent-offspring Interactions Determine Time and Place of First Ingestion of Solid Food by Wild Rat Pups," *Psychonomic Science,* 25, 5-16.

Gallistal, C., 1990, *The Organization of Learning*, Cambridge, MA: The MIT Press.

Gandolfi, G., & Parisi, V., 1973, "Ethological Aspects of Predation by Rats, *Rattus norvegicus* (Berkenhout) on Bivalves, *Unio pictorum,* L., and *Cerastoderma lamarki* (Reeve), *Bullettino di Zoologia*, 40, 69-74.

Ganti, T., 2003, *The Principles of Life*, New York: Oxford.

Gates, D., 1962, *Energy Exchange in the Biosphere*, New York: Harper & Row.

Gauthier, D., 1963, *Practical Reasoning*, Oxford: Oxford University Press.

Ghiselin, M., 1969, *The Triumph of the Darwinian Method*, Berkeley: University of California Press.

_____, 1974, "A Radical Solution to the Species Problem," *Systematic Zoology*, 23: 536-44.

_____, 1997, *Metaphysics and the Origin of Species*, Albany, NY: State University of New York Press.

Gibbs, W., 2003, "The Unseen Genome: Beyond DNA," *Scientific American*, 289.6: 106-14.

Gibson, E., 1994, "Has Psychology a Future?" *Psychological Science*¸ 5.2 (March): 69-75.

Gibson, J., 1979, *The Ecological Approach to Visual Perception*, Boston, MA: Houghton-Mifflin.

Gillespie, J., 1991, "Mutation modification in a random environment," *Evolution*, 35, 468-76.

Godfrey-Smith, P., 1994, "A Modern History Theory of Functions," *Noûs*, 28, 344-62.

_____, 1996, *Complexity and the Function of Mind*, New York, NY: Cambridge University Press.

_____, 2002, "Environmental Complexity, Signal Detection, and the Evolution of Cognition," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds., Bekoff, M., Allen, C., & Burghardt, G., Cambridge, MA: The MIT Press, 135-42.

Goffart, M., 1971, *Function and Form in the Sloth*, Oxford: Pergamon Press.

Gould, S., 1977, *Ontogeny and Phylogeny*, Cambridge, MA: Belnap Press.

_____, 1983, "Worm for a century, and all seasons," in *Hen's Teeth and Horse's Toes*, New York: W.W. Norton, 120-133

Gould, S., and Lewontin, R., 1979, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London*, B 205: 581-98.

Graff, O., 1983, "Darwin on earthworms – The contemporary background and what critics thought," in *Earthworm Ecology: From Darwin to Vermiculture*, ed., Satchell, J., London: Chapman and Hall, 5-18.

Grene, M., 1971, "Reducibility: Another Side Issue?" in *Interpretations of Mind and Life*, Grene, M., ed., New York: Humanities Press (reprinted in *The Understanding of Nature*, Cohen, R., & Wartofsky, M., eds., Dordrecht: Reidel (1974, 53-73).

Griffen, D., 1984, *Animal Thinking*, Cambridge MA.: Harvard University Press.

Griffiths, 1993, "Functional Analysis and Proper Functions," *British Journal for Philosophy of Science*, 44, 409-22.

Gross, J., 1994, "Developmental Decisions in *Dictyostelium discoideum,*" *Microbiological Reviews*, 58, 330-51.

Hempel, C., 1959, "The Logic of Functional Analysis," *Symposium on Sociological Theory*, Gross, L., ed., Evanston, IL: Harper & Row.

_____, 1965, *Aspects of Scientific Explanation*, New York: Free Press.

Holsinger, K, and Fledman, M., 1991, "Modifiers of mutation rate: evolutionary optimum with compete selfing," *Proceedings of the National Academy of Science*, 80, 6732-4.

Hospers, J., 1958, "What Means This Freedom," in *Determinism and Freedom in the Age of Modern Science*, Hook, S., ed., New York, NY: Collier.

Hove, J., Koster, R., Forouhar, A., Acevedo-Bolton, G., Fraser, S., Gharib, M., 2003, "Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis," *Nature,* 6919, 172-7.

Hull, D., 1978, "A Matter of Individuality," *Philosophy of Science*, 45, 335-60.

Hume, D., *A Treatise of Human Nature*, Norton, D. & Norton, M., eds., Oxford: Oxford University Press (2000).

James, W., 1890/1950, *The Principles of Psychology*¸ vol. 1, New York: Dover Publications.

Jamieson, D., 2002, "Cognitive Ethology at the End of Neuroscience," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds., Bekoff, M., Allen, C., & Burghardt, G., Cambridge, MA: The MIT Press.

Jablonka, E., and Lamb, M., 1989, "The inheritance of acquired epigenetic variations," *Journal of Theoretical Biology,* 139, 69-83.

_____, 1998, "Epigenetic inheritance in evolution," *Journal of Evolutionary Biology*, 11, 159-83.

Kant, I., 1790/2000, *Critique of the Power of Judgment*, trans. Guyer, P., and Matthews, E., Cambridge: Cambridge University Press.

Kauffman, S., 2000, *Investigations*, New York, NY: Oxford University Press.

Kawecki, T., 1997, "Sympatric speciation by habitat specialization driven by deleterious mutations," *Evolution*, 51, 1751-63.

Kenney, A., 1963, *Action, Emotion, and Will,* London: Routledge.

Kluge, A., 1990, "Species as Historical Individuals," *Biology and Philosophy*, 5, 417-31.

Kirk, D., 1998, *Volvox: Molecular-Genetic Origins of Multicellularity and Cellular Differentiation*, Cambridge: Cambridge University Press.

_____, 2003, "Seeking the Ultimate and Proximate Causes of *Volvox* Multicellularity and Cellular Differentiation," *Integrative Comparative Biology*, 43, 247-53.

_____, 2005, "A Twelve-step Program for Evolving Multicellularity and a Division of Labor," *Bioessays*, 27, 299-310.

Kimura, 1960, "Optimal Mutation Rate and Degree of Dominance as Determined by the Principle of Minimum Genetic Load," *Journal of Genetics*, 57, 21-34.

Kitcher, P., 1993, "Function and Design," *Midwest Studies in Philosophy*, 18, 379-97.

Lariviere, S., 2002, "Vulpes zerda," *Mammalian Species*, 714: 1-5.

Lehman, H., 1965, "Functional Explanations in Biology," *Philosophy of Science*, 32, 1-20.

Liberman, U., and Feldman, M., 1986, "Modifiers of mutation rate: a general reduction principle." *Theoretical Population Biology*¸ 30, 125-42.

Lycan, W., 1981, "Form, Function, and Feel," *Journal of Philosophy*, 78, 24-49.

_____, 1994, *Modality and Meaning*, Boston: Kluwer.

MacIntyre, A., 1957, "Determinism," *Mind*, 66.261: 28-41.

Malcolm, N., 1968, "The Conceivability of Mechanism," *Philosophical Review*, 77: 45-72.

Matthen, M., 1988, "Biological Functions and Perceptual Content," *The Journal of Philosophy*, 85, 5-27.

Mayr, E., 1961, “Cause and Effect in Biology,” *Science,* 134, 1501-6.

_____, 1974, “Teleological and Teleonomic, A New Analysis,” *Boston Studies in the Philosophy of Science*, 14, 91-117.

Melden, A., 1961, *Free Action*, New York, NY: Humanities Press.

Merritt, D., 1985, “The two-toed Hoffman’s sloth,” in *The Evolution and Ecology of Armadillos, Sloths, and Vermilinguas*, ed., Montgomery, G., Washington D.C.: Smithsonian Institute Press, 151-62.

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

McLaughlin, P., 2001, *What Functions Explain: Functional Explanation and Self-Reproducing Systems*, Cambridge: Cambridge University Press.

Michod, R., 1996, “Cooperation and Conflict in the Evolution of Individuality (II): Conflict Mediation,” *Proceedings of the Royal Society*, B 263, 813-822.

______, 1997, “Cooperation and Conflict in the Evolution of Individuality (I): Multilevel Selection of the Organism,” *American Naturalist,* 149, 607-45.

_____, 1999, *Darwinian Dynamics, Evolutionary Transitions in Fitness, and Individuality*, Princeton NJ: Princeton University Press.

Michod, R., Nedelcu, A., & Roze, D., 2003, “Cooperation and Conflict in the Evolution of Individuality (IV): Conflict Mediation and Evolvability in *Volvox carteri*.” *Biosystems*¸ 69, 95-114.

Miller, S., 2002, “Taming the fierce roller: an ‘enhanced’ understanding of cellular differentiation in *Volvox*,” *Bioessays*, 24, 3-7.

Millikan, R., 1984, *Language, Truth and Other Biological Categories,* Cambridge, MA: MIT Press.

_____, 1989a, “Biosemantics,” *Journal of Philosophy*, 86, 281-97.

_____, 1989b, “In Defense of Proper Functions,” *Philosophy of Science,* 56, 288-302.

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

_____, 1999a, “Wings, Spoons, Pills, and Quills: A Pluralist Theory of Function,” *Journal of Philosophy*, 96.4, 191-206.

_____, 1999b, "Historical Kinds and the 'Special Sciences'," *Philosophical Studies*, 95, 45-65.

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

______, 2002, "Biofunctions: Two Paradigms," in *Functions: New Essays in the Philosophy of Psychology and Biology,* eds., Ariew, A., Cummins, R., & Perlman, M., Oxford: Oxford University Press, 113-143.

_____, 2004, *Varieties of Meaning: The 2002 Jean Nicod Lectures*, Cambridge, MA: MIT Press.

_____, 2005, *Language: A Biological Model*, Oxford: Oxford University Press.

Moran, N., 1992, "The Evolutionary Maintenance of Alternative Phenotypes," *American Naturalist*, 139, 971-89.

Morowitz, H., Heinz, B., & Deamer, D., 1988, "The Chemical Logic of a Minimum Protocell," *Origins of Life and Evolution of the Biosphere*, 18, 281-8.

Nagel, E., 1957, "A Formalization of Function," *Logic Without Metaphysics*, Glencoe, IL.: Free Press, 247-83.

_____, 1961, *The Structure of Science*, New York: Harcourt Brace Jovanovich.

_____, 1977a, "Goal-Directed Processes in Biology," *The Journal of Philosophy*, 74, 261-79.

_____. 1977b, "Functional Explanations in Biology", *The Journal of Philosophy*, 74, 280-301.

Nakamura, M., 2002, "Grooming-hand-clasp in Mahala M Group Chimpanzees: Implications for Culture in Social Behaviors," in *Behavioral Diversity in Chimpanzees and Bonobos*, eds., Boesch, C., Hohmann, G., & Marchant, L., Cambridge: Cambridge University Press.

Nakishima, K., Watanabe, H., Saitoh, H., Tokuda, G., Azuma, J., 2002, "Dual Cellulose-Digesting System of the Wood-Feeding Termite," *Insect Biochemistry & Molecular Biology*, 32, 777-84.

Nash, T., 1996, *Lichen Biology*, Cambridge: Cambridge University Press.

Neander, K., 1991a, "Functions as Selected Effects," *Philosophy of Science*, 168-84.

_____, 1991b, “The Teleological Notion of Function,” *Australasian Journal of Philosophy*, 454-68.

_____, 1995, “Misrepresenting and Malfunctioning,” *Philosophical Studies*, 79, 109-141.

Owings, D., 2002, “The Cognitive Defender: How Ground Squirrels Assess Their Predators,” in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds., Bekoff, M., Allen, C., & Burghardt, G., Cambridge, MA: The MIT Press, 19-26.

Papineau, D., 1987, *Reality and Representation*¸ Oxford: Blackwell.

_____, 1993, *Philosophical Naturalism*, Oxford: Blackwell.

Parker, G., and Partridge, L., 1998, “Sexual conflict and speciation,” *Philosophical Transactions of the Royal Society of London*, B 353, 261-74.

Perry, J., Staley, J., & Lory, S., 2002, *Microbial Life*, Sunderland, MA.: Sinauer Associates, Publishers.

Place, U.T., 1956, “Is Consciousness a Brain Process,” *British Journal of Psychology*, 47, 44-50.

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

Putnam, H., 1967, “Psychological Predicates,” in *Art, Mind, and Religion*, eds., Capitan, W. and Merrill, D., Pittsburgh: University of Pittsburgh Press.

Prescott, L., Harley, J., & Klein, D., 2002, *Microbiology*, 5th ed., Boston: McGraw Hill.

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

Prusiner, S., 1995, “The Prion Diseases,” *Scientific American*, 272: 48-56.

Reed, E., 1996, *Encountering the World: Toward an Ecological Psychology*, New York, NY: Oxford University Press.

Reichenbach, H., 1947, *Elements of Symbolic Logic*, New York: Macmillan.

Resnick, M., 1997, *Turtles, Termites, and Traffic Jams*, Cambridge, MA: The MIT Press.

Rosenblueth, A., Wiener, N., & Bigelow, J., 1943, “Behavior, Purpose, and Teleology,” *Philosophy of Science*, 10, 18-24.

Rowlands, M, 1997, "Teleological Semantics," *Mind*, 106, 279-303.

Ryan, M., & Wilczynski, F., 1991, "Evolution of Intraspecific Variation in Advertisement Calls," *Biological Journal of the Linnean Society*, 44, 249-71.

Ryle, G., 1949, *The Concept of Mind*, New York, NY: Barnes and Noble, Inc.

Scheffler, I., 1959, "Thoughts on Teleology," *British Journal for the Philosophy of Science*, 9, 265-84.

Schwendener, S., 1869, *Die Algentypen der Flechtengonidien*, Basel: Schultze.

Sellars, W., 1956, "Empiricism and Philosophy of Mind," *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, eds., Fiegl, H. & Scriven, M., Minneapolis, MI: University of Minnesota Press; reprinted in *Science, Perception, and Reality*, Atascadero, CA: Ridgeway Publishing Co. (1963).

_____, 1962, "Philosophy and the Scientific Image of Man," *Frontiers of Science and Philosophy*, ed. Colodny, R., Pittsburgh, PA: University of Pittsburgh Press; reprinted in *Science, Perception, and Reality*, Atascadero, CA: Ridgeway Publishing Co. (1963)

_____, 1963, "Some Reflections on Language Games," in *Science, Perception, and Reality*, Atascadero, CA: Ridgeway Publishing Co, 321-58. (This is a revised version of what originally appeared in *Philosophy of Science*, 21 (1954), 204-28.)

_____, 1967, "Notes on Intentionality," in *Philosophical Perspectives: Metaphysics and Epistemology*, Atascadero, CA: Ridgeway Publishing Co., 128-140.

_____, 1968, *Science and Metaphysics: Variations on Kantian Themes*, New York, NY: Humanities Press.

_____, 1969, "Language as Thought and as Communication," *Philosophy and Phenomenological Research*, 29, 506-27.

_____, 1974, "Meaning as Functional Classification," *Synthese*, 27, 417-37.

_____, 1979, *Naturalism and Ontology: The John Dewey Lectures for 1974*, Atascadero, CA: Ridgeway Publishing Co.

_____, 1981, "Mental Events," *Philosophical Studies*, 39, 325-45.

Shapiro, L, 2000, "Multiple Realizations," *Journal of Philosophy,* 47, 635-54.

Sheldon, J., 1992, *Wild Dogs: The Natural History of the Non-Domestic Canidae*, New York: NY: Academic Press.

Shettleworth, S., 2002, "Spatial Behavior, Food Storing, and the Modular Mind," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds., Bekoff, M., Allen, C., & Burghardt, G., Cambridge, MA: The MIT Press, 123-8.

Skarstein, F., Folstad, I., and Rønning, H., 2005, "Spawning colouration, parasites and habitat selection in *Salelius aplinus*: initiating speciation by sexual selection?" *Journal of Fish Biology*, 67, 969-80.

Smart, J., 1959, "Sensations and Brain Processes," *Philosophical Review*, 68, 141-56.

Smolin, L., 1997, *The Life of the Cosmos*, London: Weidenfeld & Nicolson.

Sober, E., 1984, *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, Cambridge, MA: MIT Press.

_____, 1994, "The Adaptive Advantage of Learning over *A Priori* Prejudice," in *From a Biological Point of View: Essays in Evolutionary Philosophy*, Cambridge: Cambridge University Press.

Sommerhoff, G., 1950, *Analytical Biology*, Oxford: Oxford University Press.

_____, 1959, "The Abstract Characteristics of Living Organisms," in *Systems Thinking*, ed., Emery, F.E., London: Harmondsworth.

Sorabji, R., 1964, "Function," *Philosophical Quarterly,* 14, 289-302.

Stanley, S., 1986, *Earth and Life Through Time*, New York: Freeman.

_____, 1987, *Extinction*, New York: Freeman.

Stich, S., 1978, "Autonomous Psychology and the Belief-Desire Thesis," *Monist*, 61, 573-91.

_____, 1983, *From Folk Psychology to Cognitive Science*, Cambridge, MA: Bradford Books.

Stix, G., 2006, "An Antibiotic Resistance Fighter," *Scientific American*, 294:4, 80-3.

Summers, A., 2005, "A Simple Heart," *Natural History*, 114, 22-3.

Taylor, R., 1966, *Action and Purpose*, Englewood Cliffs, NJ: Prentice-Hall.

Tehler, A., 1996, "Systematics, Phylogeny, and Classification," in *Lichen Biology*, ed., Nash, T., Cambridge: Cambridge University Press.

Underhill, G., 1904, "The Use and Abuse of Final Causes," *Mind*, 13, 220-241.

van Doorn, G., Noest, A., Hogeweg, , P., 1998, “Sympatric speciation and extinction driven by environmental sexual selection,” *Proceedings of the Royal Society of London,* B 265, 1915-19.

Verzijden, M., Lachlan, R., Servedio, M., 2005, “Female Mate-Choice Behavior and Sympatric Speciation,” *Evolution*, 59, 2097-108.

Vogel, S., 1977, “Current-induced flow through sponges *in situ*,” *Proceedings of the Natural Academy of Sciences*, 74, 2006-71.

_____, 1978, “Evidence for One-way Valves in the Water Flow System of Sponges,” *Journal of Experimental Biology*, 76, 747-49.

_____, 1981, “Behavior and the Physical World of an Animal,” *Perspectives in Ethology: Advantages of Diversity*, vol. 4, Batson, P. & Klopfer, P., eds., New York: Plenum Press, 179-97.

von Uexküll, J., 1934/1957, “A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds,” in *Instinctive Behavior: The Development of a Modern Concept*, Schiller, C., trans. and ed., New York, NY: International University Press, Inc. (1957).

Wächtershäuser, G., 1988, “Before Enzymes and Templates: Theory of Surface Metabolism,” *Microbiological Reviews,* 52, 452-84.

Waddington, C., 1968, *Towards a Theoretical Biology,* Edinburgh University Press.

Wertheimer, R., 1972, *The Significance of Sense*, Ithaca: Cornell University Press.

West, P., & Packer, C., 2002, “Sexual Selection, Temperature, and the Lion’s Mane,” *Science*, 5585: 1339-1343.

White, A., 1975, *Modal Thinking*, Ithaca: Cornell University Press.

Whiten, A., Goodall, J., McGrew, W., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C., Wrangham, R.., & Boesch, C., 1999, “Cultures in Chimpanzees,” *Nature*, 399, 682-5.

Whiten, A., 2005, “The second inheritance system of chimpanzees and humans,” *Nature*, 437, 52-55.

Williams, G., 1966, *Adaptation and Natural Selection*, Princeton: Princeton University Press.

Wooldridge, D., 1963, *The Machinery of the Brain*, New York, NY: McGraw Hill.

Wimsatt, W., 2002, "Functional Organization, Analogy, and Inference," in *Functions: New Essays in the Philosophy of Psychology and Biology,* eds., Ariew, A., Cummins, R., & Perlman, M., Oxford: Oxford University Press, 173-221.

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

Zhu, B., Henderson, G., & Laine, R., 2005, "Screening Method for Inhibitors Against Formosan Subterrean Termite β-Glucosidases In Vivo," *Journal for Economic Entomology*, 98, 41-6.