INTRODUCTION

Natural Teleology

David J. Buller

I. Introduction

Within the past decade a near-consensus has emerged among philosophers concerning how to understand teleological concepts in biology. This is not to say, of course, that there is complete agreement; but broad agreement about certain fundamental commitments can be discerned in the recent literature, and the essays reprinted in chapters 3–15 of this volume exemplify this consensus. Against the background of this core agreement, several issues stand out as the foci of recent debates, and the essays in this volume have also been selected to exemplify these debates. In this introduction, I want to identify both the core consensus regarding a theory of biological teleology and the dimensions of debate over the peripheral details of such a theory. But first some stage setting is in order.

II. The Philosophical Problem

Since all theories, philosophical or otherwise, are designed to solve certain problems, it is worth being clear about the problem the essays in this volume address before examining their proposed solutions. Broadly speaking, there are two sources of the problem of biological teleology, one in the philosophy of science and one in metaphysics. I will take these in turn.

1. Philosophy of Science. Philosophy of science developed in this century as an offshoot of epistemology concerned specifically

with investigating the nature of scientific explanation, the relation between scientific theory and evidence, the nature and viability of the ontological commitments of scientific theories, and the proper explication of theoretical concepts in science. When philosophers of science turned their attention to biology, one cluster of concepts stood out as particularly in need of explication—the concept of *function* and its synonyms *in order to*, *for the sake of*, etc. For biologists routinely employ the concept of function and its relatives in their descriptions of the organs and traits of organisms: the function of the heart is to pump blood, the function of the kidneys is to filter metabolic wastes from the blood, the function of the thymus is to manufacture lymphocytes, the function of cryptic coloration (as in chameleons) is to provide protection against predators, Thomson's gazelles stott (stiff-leggedly jump up and down) in order to communicate to predators that they have been noticed, black-headed gulls remove eggshells from their nests in order to protect their fledglings against predation, the function of the mimetic sex pheromone in some orchids is pollination (by tricking male thynnine wasps into landing on the petals before flying away with grains of the orchid's pollen stuck to its legs), and so on.

Each of these descriptions of function cites an effect of an organ or trait, but not every effect of a trait or organ corresponds to a function of it. The heart, notoriously, makes noise in addition to pumping blood. But while all biologists would agree that pumping blood is a function of the heart, none would take making noise to be. Similarly, when an orchid's pollen is stuck to a wasp's legs, the wasp is heavier and it burns more calories in flight; but no biologist would take the burning of calories in male thynnine wasps to be a function of the mimetic sex pheromone in orchids, although the pheromone does produce this effect. Thus, "the function of X is to Y" cannot simply mean "X produces the effect Y," since this would fail to distinguish effects it is the function of a trait or organ to produce from those it is not its function to produce.

This makes the biologist's concept of function particularly interesting to the philosopher of science; for it does not appear to be *wholly* explicable in terms of ordinary causation familiar from the physical sciences. David Hull (1974, p. 102) put the matter very nicely:

> Just as a physicist might say that heating a gas causes it to expand, a biologist might say that heating a mammal causes it to sweat. But a biologist might also say that a

> mammal sweats when heated in order to keep its temperature constant, while no physicist would say that a gas expands when heated in order to keep its temperature constant—even though that is exactly what happens.

What, then, are the theoretical commitments implicit in the biological concept of function that distinguish the case of the sweating mammal from that of the expanding gas? Why is constant temperature merely an effect of gas expansion while being the "function" of sweating in mammals? Explicating the biologist's concept of function in order to answer these questions is one of the problems for a philosopher of science interested in biology.

2. Metaphysics. An obvious way to think about the difference between the sweating mammal and the expanding gas is in terms of goals or purposes: sweating is in some sense "goal directed" or "purposive," occurring *in order to bring about* the cooling of the body, whereas expansion of a gas is not similarly directed toward the goal of maintaining temperature. This way of casting the issue views sweating as a *teleological* process and gas expansion as nonteleological. But, while we have a firm grasp of teleological processes in the behavioral arena (we know what it is for a person to act purposively or in a goal-directed manner), how can a process such as sweating be teleological (purposive or goal directed)? This question cuts to the heart of a long-standing philosophical problem. For ever since the scientific revolution, one of philosophy's foremost metaphysical problems has been whether, and if so how, teleology is possible in nature. This problem has its roots in the conflict between Aristotelian metaphysics, which dominated philosophical thought before the scientific revolution of the 16th and 17th centuries, and the "corpuscularian" or "mechanical philosophy" that accompanied the scientific revolution. To show how the metaphysical problem of teleology emerged, let me begin with a brief sketch of Aristotelian metaphysics.

In Aristotelian metaphysics, there are four types of causation: material, formal, efficient, and final causation. Of these, only efficient and final causation are relevant to the problem of teleology. Efficient causation is what we are all familiar with in our contemporary scientific world view: efficient causes are temporally prior to their effects and initiate the changes that result in them (they are, as Aristotle puts it in *Metaphysics* 983a30–31, "the source of the change"). If we consider the mimetic sex pheromone in some

orchids, the efficient causes of the release of the pheromone are the biochemical processes that trigger its release. From the perspective of Aristotelian metaphysics, however, efficient causation does not provide a complete explanation of why any given orchid releases a mimetic sex pheromone. A complete explanation of *why* a thing is requires citing its final cause, "that for the sake of which" it exists (as Aristotle repeatedly put it in the *Physics* and *Metaphysics*); and the release of the mimetic sex pheromone is "for the sake of" pollination, since it tricks a male thynnine wasp into thinking that in landing on the orchid it is landing on a female wasp that is ready to mate, and this results in the orchid's pollen being dispersed by the male wasp. A complete explanation of why orchids release the mimetic sex pheromone, according to Aristotelian metaphysics, involves the claim that they release the pheromone in order to disperse their pollen. It is thus final causation that is teleological, since the final cause of a thing is its purpose or goal, "that for the sake of which" it exists.

Aristotelian metaphysics did not restrict the concept of final causation to biological phenomena. Indeed, it was applied freely to physical phenomena as well; according to Aristotelian metaphysics, all motion and change in nature was teleological, occurring "for the sake of" some end. This was the most immediate source of problems for Aristotelian metaphysics with the rise of the scientific revolution. For the "corpuscularian" or "mechanical philosophy" sought to explain all physical phenomena as the product of particles of matter (called "corpuscles") in motion, and to explain the motion of a particle of matter as the product of the direct actions of other particles on it (Matthews, 1989). In addition, according to the mechanical philosophy, final causes do not do any genuine explanatory work over and above the work already performed by explanations of motion solely in terms of particles in motion acting on one another. The physical universe thus came to be seen as purely mechanical and all change within it as wholly explicable in terms of antecedent events. The image was that of a clock, whose spring unwinds and turns the gears that drive the entire mechanism inexorably forward in time. And once physical phenomena became conceived in this way, the door was open for philosophers such as Hobbes to extend the same mechanical conception of causation not only to biological phenomena, but to mental phenomena as well (Burtt, 1964). The mechanical philosophy thus viewed all causation as efficient causation working forward in time. In this world view, pollination is merely an *effect*, rather than a cause, of an orchid's

releasing its mimetic sex pheromone, since the pollination occurs *after* the release of the pheromone.

As a consequence of the rise of science and the mechanical philosophy, Aristotelian goal-directed final causation came to be seen as suffering from one or both of two principal difficulties. First, it appeared to put the cart before the horse—explaining a cause in terms of its effects—and thus to require "backward" causation. This conflicted with the mechanical philosophy's conception of all change as explicable in terms of the motions of particles and of all of a particle's motions as explicable in terms of the direct actions on it of other particles in motion. For the dispersal of pollen from an orchid cannot set in motion any actions of particles that would *initiate* the release of the mimetic pheromone, since it occurs *after* the release of the pheromone. Thus, pollination can in no way explain the pheromone's release. So goal-directed final causation seems physically impossible and, hence, not explanatory.

There is, however, one clear domain in which goal-directed causation seems both possible and explanatory. Intelligent beings *represent* to themselves the goals that they pursue in their deliberations and actions, and these representations of goals are among the efficient causes of their deliberations and actions. For example, we can explain my going to the freezer and retrieving the ice cream in terms of my desire to eat ice cream (and my belief that ice cream is in the freezer). In offering such an explanation, we are not claiming that the event of my eating ice cream caused the earlier event of my retrieving the ice cream from the freezer. Rather, we are claiming that *my desire* to eat ice cream was a cause of my retrieving the ice cream; and this desire was temporally prior to the event that it is invoked to explain (my retrieving the ice cream). Of course, what makes my desire the desire *to eat ice cream*, rather than some other desire, is the fact that it involves a representation of my eating ice cream, rather than a representation of something else. Such explanations of the deliberations and actions of intelligent beings, then, are paradigms of explanation in terms of goal-directed causation; but they do not actually require backward causation, since they are always formulated in terms of (antecedent) efficient causes that represent some goal.

If this form of goal-directed causation is extended to the full range of cases to which Aristotelian metaphysics applied the concept of final causation, however, a second problem arises. For this would require that sweat glands somehow represent the goal of lowering body temperature; and this, in turn, would require that sweat

glands be "intelligent" in some sense. Similarly, it would require that the thymus somehow "know" that it must manufacture lymphocytes and that the orchid somehow have "figured out" what it must do in order to deceive male thynnine wasps into dispersing its pollen. In short, treating all final causation as goal directed in the way that intelligent behavior is appears to require that intelligence pervade the physical universe, being present in the most unlikely plants and organs of organisms.

Thus, once the mechanical philosophy took hold and efficient causation was seen as sufficient to explain all change in nature, final causation appeared to be caught on the horns of a dilemma: either it involves the unacceptable postulation of backward causation or it involves a grossly implausible panpsychism. In short, there appeared to be no room at all for teleology, or purpose, within the scientific world view. And ever since the rise of this scientific world view, the metaphysical problem of teleology has been that of explaining whether, and if so how, there can be goal-directed processes in a universe governed solely by efficient causation.

In spite of the difficulties associated with teleology, however, biologists continued to use the teleological concept of function in describing the characteristics of organisms, finding the organization of organisms and the operation of their parts virtually incomprehensible in strictly non-teleological terms. As Darwin wrote in a letter to A. de Candolle, it is "difficult for any one who tries to make out the use of a structure to avoid the word purpose" (quoted in Ghiselin, 1997, p. 63). This poses the following problem, which derives from the two sources just discussed: How can the biological concept of function, which is *prima facie* infected with final causation, be analyzed so as to make it compatible with a scientific world view that countenances only efficient causation?

The problem is by no means easy. In the first place, the biological concept of function discriminates among the effects of an item: pumping blood is a function of the heart, but making noise is not; making noise is merely an "accidental" effect of the heart's pumping. So any solution to the problem of biological teleology must explain why the concept of function discriminates between functional and accidental effects in this way. Second, the biological concept of function does seem to imply that the effects that it is the function of an item to produce in some sense *explain* the existence of that thing: saying that the function of sweat is to maintain constant temperature implies that the need for constant body temperature in some way explains *why* mammals sweat. This actually

works together with the first point. The reason why pumping blood is a function of the heart, whereas making noise is not, is that its pumping blood in some sense explains *why* an organism has a heart, whereas its making noise does not. So any solution to the problem needs to account for how the biological concept of function can be explanatory in this way. But, third, in providing this account, the biological concept of function must be shown not to require the postulation of backward causation or panpsychism, or else it cannot be a proper scientific concept. In short, then, the problem of biological teleology is to explain how, for example, the claim that the function of kidneys is to remove metabolic wastes from the blood—that kidneys exist in order to do that—can in any way *explain why* animals have kidneys, when explanations must cite only efficient causes.

III. Recent Prehistory: The "State of the Art" in the 1960s

This was the problem that both Nagel (1961) and Hempel (1965) tried to solve within the framework of the *deductive-nomological model of explanation* (the classic formulation of which is to be found in Hempel & Oppenheim, 1948). According to this model, all explanations conform to the same general logical form: they all explain some phenomenon by deducing its existence or occurrence from a set of premises that include lawlike (nomological) regularities. When this logical model of explanation is conjoined with the assumption that all causation is efficient causation, the problem of analyzing the biologist's concept of function in statements such as "the function of *X* is to *Y*" takes the following form: How can the existence of *Xs* be deduced from lawlike statements that include the fact that *Ys* are the *effect* of *Xs*? For example, how can the presence of kidneys in humans be deduced from lawlike statements that include the fact that kidneys have the effect of removing metabolic wastes from the blood?

Both Hempel and Nagel started from the idea that a statement such as

(1) The function of the heart is to circulate the blood

is an elliptical, or shorthand, explanation of the existence of hearts. The task, then, is to make the explanation fully explicit and show how it conforms to the deductive-nomological model.

Doing so requires noting several things. First, (1) is implicitly relativized to organisms with hearts, in particular vertebrates; that is, (1) is a claim about the function of the heart in the kinds of organism that possess hearts. Second, the heart can perform this function in some organism only in virtue of its having a particular anatomical organization; it is only in the context of the structure of the circulatory system as a whole, for example, that hearts can function to circulate the blood. Third, to put it crudely, the heart only performs this function in live vertebrates, and vertebrates depend for life on particular chemical properties of their external environments. When we make these three things explicit, (1) becomes

(2) The function of the heart in vertebrates with a particular anatomical organization and in a particular environment is to enable them to circulate their blood.

But (2) is merely an instance of the following general schema, as Nagel pointed out:

(3) "The function of A in a system S with organization C is to enable S in the environment E to engage in process P" (Nagel, 1961, p. 403; cf. Hempel, 1965, p. 306).

This, then, was taken to be the fully explicit form of statements that describe the function of some trait of an organism.

This just leaves the problem of how to convert a statement with the form of (3) into a fully explicit deductive-nomological explanation of the existence of the entity A in the system S. Since Hempel and Nagel offered very similar solutions to this problem, I will focus on Nagel's solution and discuss Hempel's in relation to it. Nagel proposed that the informational content of (3) can be formulated as the following deductive argument, which, according to the deductive-nomological model of explanation, constitutes an explanation of the existence of A in system S (1961, p. 403):

(a) "Every system S with organization C and in environment E engages in process P,"

(b) "if S with organization C and in environment E does not have A, then S does not engage in P,"

(c) "hence, S with organization C must have A."

Here (a) provides a law to the effect that every *S* (e.g. vertebrate) does *P* (circulate blood), and (b) states that having *A* (a heart) is a *necessary condition* for doing *P*. From these two premises it deductively follows that *S* must have *A*. Consequently, if we make the substitutions for the variables as just noted, we appear to have an explanation of the existence of hearts in vertebrates that appeals to the fact that hearts have the *effect* of circulating blood. And Hempel and Nagel both discriminated functional effects from accidental effects by requiring that process *P* in the above schema be necessary for maintaining the system *S* in proper working order. So circulating blood is in turn a necessary condition for the survival of vertebrates, whereas the noise made by the circulation of the blood is not.

The problem with this explanation is that (b) is false. Cummins (chapter 2) discusses the difficulty at length, and I will not repeat his arguments here. For present purposes it is sufficient merely to note that artificial pumps could circulate the blood in vertebrates; so the heart is not necessary for circulating blood. Nagel was aware of this problem, and responded by saying that function statements in biology "are not explorations of merely logical possibilities, but deal with the actual functions of definite components in concretely given living systems" (1961, p. 404). That is, Nagel wanted to construe the "necessary" narrowly, as applying only to the available *biologically possible* options. But there are two problems with Nagel's response. First, prosthetic organs are not merely "logical possibilities," but have replaced, and performed the functions of, "definite components" in some actual "concretely given living systems." Second, if we restrict our examples to organs (hearts, kidneys, livers, and so on), it may appear plausible that they are the only available biological options for producing the effects that they have, and are in that sense necessary for those effects. But if we consider traits of organisms generally, the plausibility vanishes. For it would be highly implausible to think that cryptic coloration is the only way that chameleons could avoid predation, that eggshell removal is the only way that black-headed gulls can protect their fledglings from predation, or that stotting is the only way that Thomson's gazelles can communicate to cheetahs that they have been noticed. Surely, in each of these cases, there is a range of biologically possible ways of achieving the same effect, and in general the evolutionary process continually produces diverse solutions to adaptive problems. So construing "the function of *A* is to produce process *P*" as entailing that *A* is necessary for *P* seems far too strong.

This led Hempel (1965, pp. 310–312) to consider a variant of (b), which loosens the requirement that there be only one way of producing process *P*. Hempel's suggestion amounts to letting *A* be merely one element of a class of items that are jointly necessary for producing *P*. Thus, if *P* is the process of circulating blood, Hempel's proposal is to allow *A* to be one element of a class of items, *I*, that are necessary for circulating blood; so *I* would include the heart (*A*), artificial pumps, etc. This leads to the following revised version of (b):

(b′) if *S* with organization *C* and in environment *E* does not have one element of class *I* (of which *A* is a member), then *S* does not engage in *P*.

While this replaces (b) with a true premise, as Hempel pointed out (p. 312) the resulting schema fails to explain why *S* has *A*, since (c) is not deducible from (a) and (b′). The most that we could deduce from (a) and (b′) would be that *S* has *at least one* of the elements of the class *I*; we cannot deduce that the element of *I* that *S* has is *A*. To make this concrete, if all vertebrates circulate blood (as per (a)), and vertebrates can circulate blood only if they have one element of the class of items consisting of hearts, artificial pumps, and so on (as per (b′)), it only follows that a given vertebrate must have a heart *or* an artificial pump *or* one of the other items that is necessary for pumping blood. But this means that (a) and (b′) fail to *explain* why any given vertebrate has a heart; for, according to the deductive-nomological model of explanation, (a) and (b′) explain why some vertebrate has a heart only if they jointly entail that it has a heart rather than one of the other items that could pump blood. So, although (b′) provides a true premise for the explanation, it renders the deduction of (c) invalid. To make the deduction valid, (c) would have to be replaced with

(c′) hence, *S* with organization *C* must have one element of class *I* (of which *A* is a member).

But then we no longer have an explanation of why *S* has *A*; we have only an explanation of why *S* has *something* that produces process *P*.

This is a dilemma. Either a statement such as "the function of the heart in vertebrates is to circulate blood" gets analyzed as a deductively valid explanation with a false premise or it gets ana-

lyzed as a deductively valid argument with true premises that fails to explain why vertebrates have hearts. Strangely, Hempel grasped the dilemma by both horns and said that the only case in which we have a genuine functional explanation is one in which there actually is one and only one element of the class *I* (pp. 313–314). In that event, that element is indeed necessary for producing process *P* and Nagel's analysis holds. In every other case, a statement of the form "the function of *X* is to *Y*" fails to explain the existence of *X*s by citing the fact that they produce effects of type *Y*. But this is just to say that Hempel and Nagel failed to solve the problem of explicating the teleological content of the biological concept of function. For by their analyses, the functions of traits or organs rarely if ever explain the existence of those traits or organs. And this is just to say that, by their analyses, the concept of function performs no genuine explanatory work.

IV. Wright and Cummins

The analyses of Hempel and Nagel were typical of philosophical analyses of the concept of function until the mid-1970s, when two articles appeared that proved decisive in reshaping efforts to understand the concept of function. Those articles were Larry Wright's 1973 article "Functions" (chapter 1) and Robert Cummins' 1975 article "Functional Analysis" (chapter 2), and they are reprinted here not because they are part of the recent near-consensus, but because they have exerted such a strong influence on the work that has formed that near-consensus.

Wright and Cummins both reject the idea, common to Hempel and Nagel, that function statements are elliptical deductive-nomological explanations of the existence of a functional item. In addition, neither Wright nor Cummins is concerned solely with explicating the concept of function as it is used in biology; they both offer theories that are intended to apply to all instances of function, whether biological, mechanical, institutional, or artifactual. Apart from these two commonalities, however, their theories are radically different.

According to Wright (1973, p. 161), a statement of the form "the function of *X* is *Y*" means simply

(a) *X* is there because it does *Y*,

(b) *Y* is a consequence (or result) of *X*'s being there.

In this schema, (b) exhibits the fact that Y is an effect (not a cause) of X, and (a) exhibits the teleological explanation of the existence of X, since the "because" in (a) "is to be taken in its ordinary, conversational, causal-explanatory sense" (1973, p. 157). This is where Wright offers a novel twist. Instead of construing Y itself as a cause of X (which would reintroduce the problem of backward causation that appeared to plague Aristotelian metaphysics), Wright's analysis takes *the fact that Y is an effect of X* to be among the (antecedent) efficient causes of X, and thus provides an efficient causal explanation of the existence of X in terms of its producing Y. Wright calls this an "etiological" theory of teleology, since it analyzes statements that ascribe a function to X as explanations of the existence of X strictly in terms of its (antecedent) efficient causes—its etiology.

The causal explanation of X that substantiates (a) is something that Wright thinks can vary from context to context, so he doesn't build it into the analysis. If X is a fuel injection system, then the causal explanation of why it is in a car will be formulated in terms of the design of the car engine, and the explanation may make detailed reference to how alternatives to the fuel injection system may have been tried unsuccessfully or less successfully in the design of the engine. If X is the rock holding open my office door, then the explanation of why it is at the foot of my door will simply be in terms of my intention that my door be held open. And if X is the heart, then the causal explanation of why it is in a vertebrate will be formulated in terms of natural selection.

This last-mentioned biological case is the one that is most interesting for the purposes of this volume. Consider how natural selection provides an explanation of why humans, for example, have hearts. The heart is a complex organ and all complex traits are the product of accumulated modifications to antecedently existing structures. These modifications to existing structures occur randomly as a result of genetic mutation or recombination. When they occur, there is variation in a population of organisms (if there wasn't already) with respect to some trait. If one of the variants of the trait provides its possessor(s) with an advantage in the competition for survival and reproduction, then that variant will become better represented in the population in subsequent generations. When this occurs, that variant of the trait has increased the relative fitness of its possessor(s) and there has been "selection for" that variant (see Sober, 1984c, pp. 98–102). That variant can then provide the basis for further modification. Thus, humans have hearts

because hearts were the product of randomly generated modifications to preexisting structures that were *preserved* or *maintained* by natural selection due to their providing their possessors with a competitive edge (see Sober, 1984c, pp. 147–155). So natural selection explains the presence of a trait by explaining how it was preserved after being randomly generated.

Boorse (1976) constructed a counterexample to Wright's analysis that played on a similarity with explaining why a trait "is there" in terms of natural selection. Boorse (p. 72) presented the following scenario:

> Suppose that a scientist builds a laser which is connected by a rubber hose to a source of gaseous chlorine. After turning on the machine he notices a break in the hose, but before he can correct it he inhales the escaping gas and falls unconscious. . . . The release of the gas [*Y*] is a result of the break in the hose [*X*]; and the break is there—that is, as in natural selection, it continues to be there—because it releases the gas. If it did not do so, the scientist would correct it.

Here *Y* (the release of the gas) is clearly a consequence of *X*'s (the break's) being there; so condition (b) of Wright's analysis is satisfied. In addition, the fact that *X* produces *Y* helps to explain why *X* is there, since, if it were not for the release of the gas, the scientist would fix the hose and the break would not be there. This, Boorse argues, is like the case of natural selection in that the break in the hose originates as a result of some random event, and then is preserved (as a result of the scientist's inability to fix it). This means that condition (a) of Wright's analysis is also satisfied. But we surely wouldn't want to say that the function of the break in the hose is to release the gas. So Wright's analysis appears not to be fully successful.

Note that one thing that Wright's analysis has in common with Nagel's and Hempel's is a commitment to the idea that a function statement is an implicit explanation of the presence of the functional item. And one could see this commitment as the source of difficulties for all three analyses. In the case of Nagel and Hempel, the analyses failed to achieve their objective of providing an explanation of the existence of the functional item; and in the case of Wright, the analysis transforms an explanation of the existence of an item into a function of that item in some cases where it

shouldn't. Identifying this as the common problem of all three analyses could lead one to reject entirely the idea that a function statement is an implicit explanation of the presence of the functionally characterized item. And this is what Cummins does.

According to Cummins, function statements are implicit explanations of a unique sort; they do not explain the existence of the functional item, but rather its contribution to an activity or capacity of a system that contains that item, where that contribution emerges through a *functional analysis* of a capacity of that system. A functional analysis of a capacity C of some system S proceeds by analyzing C into the capacities of simpler components of S in such a way that C emerges as the "programmed manifestation" of the exercise of the capacities of those simpler components (where the latter capacities may themselves admit of functional analysis, until the analysis terminates in the capacities of nondecomposable structural components of S). Given this conception of functional analysis, the statement "The function of X is to Y" can be understood as stating that X is a component of some system S and X's doing Y features in a functional analysis of some capacity C of S.

To illustrate Cummins' analysis, consider respiration. To explain how the respiratory system (S) exhibits the capacity to exchange oxygen and carbon dioxide (C), we would specify the components of the respiratory system and the capacities of those components. From this it would emerge that the contraction of the diaphragm causes the expansion of the cavities containing the lungs, which in turn causes the lungs to expand and their internal pressure to drop. This drop in pressure causes outside air to fill the lungs and then oxygen and carbon dioxide are exchanged directly across the gas-permeable walls of the capillaries that cover the internal surface of the lungs. Thus, it is the function of diaphragm contraction (X) to produce the expansion of the cavities containing the lungs (Y), since the contracting diaphragm's producing that expansion features in a functional analysis of the capacity to exchange oxygen and carbon dioxide (C) of the respiratory system (S) that contains the diaphragm.

In sum, then, Cummins' analysis of the concept of function makes the function of an item merely its causal contribution to a complex process. While this certainly succeeds in avoiding appeals to anything other than efficient causation, it does so at the cost of emptying the concept of function of all its teleological content. In addition, Cummins' analysis seems unable to account for the ways

in which the concept of function is used in biology. Recall some of the obvious facts with which we began—for example, that biologists agree that pumping blood is the function of the heart, but that making noise is not. As Cummins admits (see section III.4), his analysis does not rule out saying that the function of the heart is to make noise. For we could take the mammalian circulatory system as S and its capacity to make "circulatory noise" as C, in which case a functional analysis would reveal that the heart contributes to C by making a thumping sound. But no biologist would reason in this way. Thus, if we are looking for the theoretical principles underlying the biological concept of function, which discriminate between the heart's pumping blood and its making noise, Cummins' analysis will not reveal them to us.

V. Millikan

In 1984 Millikan's *Language, Thought, and Other Biological Categories* (excerpts from which are reprinted as chapter 3) made significant headway on the problem of teleology. The details of Millikan's theory are complex and the reader will encounter them in chapter 3; so I will provide only a brief paraphrase here. Millikan's theory is intended to account for a very wide range of functions, from those of traits and organs of organisms to the functions of thoughts, words, artifacts, and cultural products; but it is clearly inspired by the theory of evolution by natural selection and intended to apply paradigmatically in the biological context. And, since the current focus is biological functions (and since her treatment of artifact functions is too complex to be discussed here and is not reprinted in chapter 3), I will focus only on Millikan's theory of biological functions and explain it by reference to a prominent way of describing the evolutionary process. While Millikan does not explain her theory in this way, my hope is that tying in Millikan's theory to broader issues in evolutionary biology will clarify how teleology can emerge from a natural process governed solely by efficient causation, and thereby clarify Millikan's solution to the philosophical problem of teleology.

As Hull (1989) has characterized it, the entities that function in the process of evolution by natural selection are *replicators* and *interactors*. A replicator is "an entity that passes on its structure largely intact in successive replications" and that replicates itself in accordance with causal laws of nature (p. 96). Genes, for example,

are replicators, since they replicate themselves by directly copying or reproducing their structure. An interactor, on the other hand, is "an entity that interacts as a cohesive whole with its environment in such a way that this interaction causes replication to be differential" (p. 96). (Dawkins (1989) uses the term "vehicle," instead of "interactor," to express roughly the same idea.) A replicator is also an interactor, since it interacts with its environment as a cohesive whole at the very least "to the extent necessary to replicate itself" (Hull, 1989, p. 96). But more paradigmatic instances of interactors are the organisms that are built by genes; for organisms interact with their environments as cohesive wholes. An organism's interaction with its environment, of course, largely consists of competition with other organisms for survival and reproduction, a competition in which different organisms meet with differing degrees of success due to differences in their characteristics. As a result of the differential success of organisms (interactors) in reproducing, there is differential replication of the genes (replicators) that built those organisms. Thus, Hull characterizes natural selection as "a process in which the differential extinction and proliferation of interactors cause the differential perpetuation of the relevant replicators" (p. 96). As a result of this process evolution occurs—that is, there are changes across generations in the relative frequencies of types of replicator and, consequently, changes in the relative frequencies of the characteristics of interactors that are developmentally constructed by those replicators.

Millikan's theory is designed to show how functions can emerge within such a process. According to Millikan, an item has a function only as a member of what she calls a "reproductively established family," where Millikan distinguishes between "first-order" and "higher-order" reproductively established families. First-order reproductively established families consist of replicators, which directly reproduce their structures when making copies of themselves in accordance with causal laws of nature. Thus, each temporal sequence of replicators that are related by descent (through copying) forms a first-order reproductively established family. The copies of the gene for blue eyes, for example, are members of a first-order reproductively established family. Replicators, however, do not just cause their own replication; they also cause the properties of the interactors that contain them (that is, differences in the properties of interactors are positively correlated with differences in the types of replicator that built them). Thus, as interactors reproduce, not only are the replicators they contain

reproduced, but the properties of interactors caused by those replicators are reproduced as well. In such cases, the replicators are *directly* reproduced through copying, whereas the properties of interactors are *indirectly* reproduced via the reproduction of the replicators that cause those properties. Such indirectly reproduced properties of interactors form higher-order reproductively established families, examples of which are traits such as eye color and blood type and organs such as livers, hearts, and lungs. Members of a higher-order reproductively established family are thus related by descent also, via the direct relations of descent of the members of the first-order reproductively established family that produce the members of the higher-order family. So, in both cases, we can say that the *ancestors* of a particular member of a reproductively established family are those temporally prior family members to which it is related by a continuous chain of (direct or indirect) reproduction.

According to Millikan, it is the function of a member X of a reproductively established family to do Y just in case ancestors of X did Y and their doing Y causally contributed to their family's having greater reproductive success than competing reproductively established families and, hence, causally contributed (eventually) to the production of X. In other words, the function of X is to do what its ancestors were "selected for" doing. This definition of "function" finds its initial application at the level of replicators, or first-order reproductively established families. For in their interactions with their environments some replicators do things that cause them to be more successful than others in replicating themselves. When there is such a history of competitive success within a family of replicators, we can pick some particular replicator X and see that its ancestors successfully replicated by doing Y; according to Millikan's theory, it is thus the *function* of X to do Y, since doing Y causally contributed to the (direct) reproduction of members of X's family. But some functions of a replicator will turn out to be *developmental*, whereby a replicator has the function of causally contributing to the production of an interactor property. In other words, members of some first-order family will have the function of producing members of some higher-order family. The members of that higher-order family of interactor properties may also contribute to the reproductive success of the interactors bearing those properties by having some particular effect. When they do, we can pick some particular interactor property X and see that its ancestors contributed to the reproductive success of the interactors bearing those ancestral properties

by doing Y; it is thus the *function* of X to do Y, since doing Y causally contributed to the (indirect) reproduction of member's of X's family. In this way, Millikan's theory of functions applies to members of all higher-order reproductively established families as well. In short, according to Millikan, it is the function of X to do Y just in case doing Y caused the proliferation of ancestors of X (through either direct or indirect reproduction).

It is worth noting how well this addresses the philosophical problem of teleology. First, it shows how "the function of X is to do Y" can causally explain the existence of X without invoking backward causation. For the fundamental idea of Millikan's theory is that it is not the fact that X itself produces Y that explains X's existence, but the fact that ancestors of X produced the effect Y and that Y was among the causes of the existence of X (via the processes of replication and/or development that eventually produced X). It is not a *current* pollination that caused this orchid to release its mimetic pheromone; rather, this orchid causally originated as a (higher-order) reproduction of ancestral orchids that released mimetic pheromones and one of those ancestral releases of mimetic pheromones succeeded in pollination that causally produced this particular orchid. *That* is *why* pollination is the function of this particular orchid's release of its mimetic pheromone.

Millikan's theory is thus clearly etiological, since it analyzes a statement that ascribes to X the function of producing Y as an explanation of the existence of X strictly in terms of X's antecedent (efficient) causes. But Millikan's etiological theory differs from Wright's. For, while Wright analyzes "the function of X is to do Y" as implying that X "is there" because *it* does Y, Millikan analyzes it as implying that X "is there" because *its ancestors* did Y; while Wright's analysis appears to focus on those instances of Y that are current effects of X, Millikan's analysis focuses on those instances of Y that were among *the causes* of X. So Millikan's etiological theory takes the causal explanation of why a functional item "is there" to refer strictly to the causal history of the item, not to what *it* causes.

Second, although clearly in the spirit of Wright's theory, Millikan's theory avoids Boorse's counterexample. For an item has the function of producing some effect only if that item originated as a copy or reproduction of earlier items that had the same effect. This avoids Boorse's counterexample, since the break in the hose did not originate as a copy of earlier breaks in hoses that had the effect of knocking scientists unconscious.

Third, Millikan's theory is able to successfully distinguish functional effects from accidental effects. For example, the reason that it is the function of my heart to pump blood, but not to make noise, is that the hearts of my ancestors contributed to their reproductive success—and thus to the reproduction of hearts—by pumping blood, not by making noise.

Fourth, Millikan's theory explains how teleology can emerge within a universe governed solely by efficient causation. For in the beginning the only interactors were probably replicators differentially replicating themselves in "the primeval soup" (see Dawkins, 1989, chap. 2), where their differential replication was due to differences in what they did in interacting with their environments. The first functions emerged, at this point in the universe, as functions of members of first-order reproductively established families. Eventually, however, some replicator interactions resulted in types of replicator bonding with one another to form more complex interactors consisting of "teams" of replicators. The properties of these complex teams of interactors biased interactions in their favor and, as a result, they began to enjoy greater differential success in replication. This eventually led to increased team size and the formation of protein walls for protection, and the first cells thus appeared. Multicellular interactors eventually followed and finally the complex organisms that we take to be paradigmatic interactors. And the emergence of these higher levels of complexity were accompanied by the emergence of the functions of members of higher-order reproductively established families. While this provides a plausible scenario about the origins of complex organisms and the functions of their traits and organs, nowhere does it make reference to anything but efficient causation and the differential perpetuation of replicators (which is a function of differential success in interaction with the environment).

VI. The Core Consensus and the Peripheral Disagreements

Millikan's approach to defining the function of an item in terms of its evolutionary, causal history succeeded in setting the agenda for subsequent discussions of the biological concept of function. Indeed, her approach so successfully set the agenda that the term "etiological theory" has come to refer in the literature to theories that define the function of an item in terms of its evolutionary history, in spite of the fact that Wright authored the (broadly) etio-

logical approach to understanding teleology. Goode and Griffiths (chapter 12), for example, reserve the term "etiological" for Millikanesque approaches and refer to Wright's theory as "proto-etiological." While this terminological convention does some violence to the history of philosophical work on functions, for the sake of expediency I will follow it, since the evolutionary approach of Millikan has exerted the strongest influence on recent work. In fact, of current theories, the (evolutionary) etiological theory has the largest number of adherents, with versions of it defended in this volume by Millikan (chapters 3 and 5), Neander (chapters 6 and 11), Griffiths (chapter 7), Godfrey-Smith (chapters 9 and 10), Goode and Griffiths (chapter 12), Allen and Bekoff (chapter 13), and Buller (chapter 15). Alternatives to the etiological theory are defended by Bigelow and Pargetter (chapter 4), Kitcher (chapter 8), and Walsh and Ariew (chapter 14).

Despite the disagreement, however, there is a common core of agreement that unites etiological theorists with the dissenters just mentioned, and this core of agreement represents as great a consensus as has been achieved in philosophy. For, in one way or another, all the authors agree that the biological concept of function is to be analyzed in terms of the theory of evolution by natural selection. That is, all agree that a trait or organ has a function in virtue of its role in a selection process—either in virtue of its role in a selection process that a lineage bearing that trait or organ actually has undergone, or in virtue of a selection process it is currently undergoing or is set to undergo. In this sense, there is consensus that the theory of evolution by natural selection can provide an analysis of the teleological concept of function strictly in terms of processes involving only efficient causation (although see Manning (1997) for an argument that this consensus viewpoint is mistaken). Agreement regarding this fundamental idea, however, still leaves a great deal of room for disagreement concerning the details of a theory of functions. In what remains of this introduction, I will sketch several foci of the debates that run through the essays that follow.

1. Looking Forward Versus Looking Back. Defining the function of a trait as its role in a selection process still leaves the following question unanswered: At what point in evolutionary history lies the selection process relevant to characterizing a trait's function? According to etiological theorists, the function of a trait is the *past* contribution that the trait made to the fitness of its bearers,