

# Chapter 28

## Function and Teleology

JUSTIN GARSON

### 1. Introduction

Function statements are used throughout the biological disciplines. For example, it is said that the function of the kidney is to extract waste products from the blood, the function of hemoglobin is the transportation of oxygen to tissue, and the function of myelin sheathing is to promote the efficient conduction of action potentials in the nervous system. In the case of many physical and mental disorders, it is believed that an inner part or process is *malfunctioning* or *dysfunctional* – such as the kidney in glomerulonephritis or myelin in multiple sclerosis – and knowledge of such dysfunctions guides medical research and intervention. Thus, functional language in biology has both theoretical and practical significance.

These examples draw attention to two interesting properties that function statements seem to possess. The first is that they are *explanatory*: to say that the function of myelin is to promote efficient nervous conduction is to say, roughly, *why myelin is there* or why many neural projections are sheathed in myelin. The second is that they are *normative*: the fact that the kidney, in the case of glomerulonephritis – a swelling of the glomeruli which filter the blood – can *fail* to perform its function implies that function statements do not necessarily *describe* what an entity actually does, but they set up a norm that specifies what that entity is *supposed* to do, or “what it is for.”

Explanations that purport to explain the existence, form, distribution, or location of an entity by referring to some future state that the entity tends to bring about are referred to as teleological. The term “teleological” derives from the Greek word *telos*, meaning “goal” or “end.” Hence function ascriptions are often thought to be a type of teleological explanation. Yet functional explanations seem problematic because they appear to violate the principle that temporally *posterior* events cannot figure into causal explanations for temporally *prior* events. The kidney must *already* be part of the organism in order to filter blood, just as neural projections must already be sheathed in myelin in order to efficiently conduct action potentials. How can a kidney’s capacity to filter blood explain why the kidney is there, unless the future is assumed to have some causal influence over the present? (This is often called the problem of “backwards causation.”)

The normative status of function statements is also puzzling. It is perfectly clear what one means by saying that an artifact, such as a camera, is *malfunctioning* – namely, that it is incapable of doing what the manufacturer made it for. But what could conceivably be the analog of a “manufacturer” in the biological realm, unless one assumes the existence of a supernatural creator – an assumption commonly deemed to have no place in legitimate scientific explanations? In what sense is the kidney *supposed to filter* the blood, rather than to support hard calcium formations along its inner wall, or to do nothing? Consequently, functional explanations are not only puzzling with respect to what they purport to explain, but they are also suspect of violating important tenets of the modern scientific worldview: the absence of final causes in nature and the illegitimacy of appealing to divine creation or intervention. Nonetheless, they are routinely appealed to throughout the biological disciplines. This suggests that they either ought to be eliminated from biology or analyzed in such a way that the appeal to final causes or supernatural beings is shown to be unnecessary.

One approach to the explication of function statements is simply to accept final causation as a distinct and irreducible type of causation. This is the solution that Aristotle is often thought to have provided. Aristotle’s view of causation (*aitia*, which can be translated as “cause” or “reason”) involves a rejection of the premise that future events cannot enter into explanations for the existence or form of a trait. His view is that the purpose, or *telos*, for which something exists cannot be eliminated from most biological explanations for the existence or form of a trait. (See his *Physics*, Book II.8 for several central arguments for this claim; also see *Parts of Animals*, Book I.1 for his defense of teleology in the context of biological explanation.)

Of course, to say that reference to future effects cannot be eliminated from an explanation is not to say that such explanations actually refer to a distinct type of causal pathway. Thus, one might interpret Aristotle liberally by suggesting that he was not really advocating the existence of final causes that somehow bring about their own realization, but advocating certain constraints on the nature of good *explanations* (translating *aitia* as “reason,” a feature of rational discourse, rather than “cause,” a mind-independent feature of the world). This latter reading is more generous, given that modern science has not accepted final causation as a distinct ontological relation. Consequently, supposing that functional language will not be eliminated from biology in the near future, any plausible account of “function” must *either* explain how it can be that the effect produced by a kind of entity can have causal relevance to the existence of the entity, *or* dissolve the misleading appearance that function ascriptions are causal explanations at all. *Etiological* approaches to function adopt the former route; *consequentialist* approaches the latter.

Intuitively, one might motivate either of the two main approaches to function by considering the following question: what distinguishes a *function* of an entity from a mere *effect* that it produces? To take a hackneyed, but simple, example, why is the function of the heart to pump blood rather than to make throbbing sounds? Two different answers present themselves as initially plausible:

- (i) according to the etiological view, what distinguishes the *function* of an entity from a mere *effect* is that the capacity of the entity to perform that function explains “why it is there” in that system. For example, it is the capacity of windshield wipers

to remove water from windshields that explains why they are on the windshield of a specific car; i.e., why the manufacturers placed them there. Similarly, one could argue that the fact that the heart has been selected for by natural selection because it pumped blood explains why, presently, creatures with hearts exist. Therefore, in conformity with the logic of teleological explanations, it is true to say that the heart's capacity to pump blood explains why hearts currently exist. However, the heart was not selected for because of the beating sounds that it makes, so there is no sense in which the heart "is there" because of its capacity to make such sounds;<sup>1</sup>

- (ii) according to the consequentialist view, the function of the heart is to beat, rather than to make noise, because the heart's beating typically contributes to some important activity of the system within which it is contained, and heart sounds do not. In this case, beating contributes to pumping blood and this in turn to the survival of the organism. This solution corresponds to the view that the *function* of an entity consists in a (special sort of) consequence that it produces, and has nothing to do with the cause or origin of the item itself.

The following is composed of two sections. Section 2 will describe the *etiological* (or "backwards-looking") approach, which rejects the premise that function statements refer exclusively to future events. It will enumerate the main conceptual challenges that philosophers have confronted it with, and some of the responses to those challenges. Section 3 will describe several contemporary variants of the consequentialist (or "forward-looking") approach to functions, which rejects the premise that function ascriptions are causal explanations for the form or existence of a trait.

## 2. Etiological Theories of Function

There are two main versions of the etiological approach: one which refers to the *reasons* that motivate a purposeful being to create a functional object ("representationalism"), and one that refers to the natural history of the functional entity, independently of the notion of representation. (The latter is typically referred to as "etiological," although "etiological," properly speaking, could refer to either view.) These views will be elaborated in turn.

### 2.1. Representationalist theories of function

The first version of the "backwards-looking" approach is standardly employed to explain the sense in which intelligent creatures act for the sake of the future: it is *not* the case

<sup>1</sup> There are, of course, exceptional cases in which it can be said that the heart's beating sounds explain why it is there. For example, if the beating sounds made by a person's heart alert a doctor to a life-threatening heart problem that is thereby remedied, then one can say that the heart sounds saved the person's life and therefore they partly explain why the person continues to exist, and hence why the heart continues to be there. Does that mean that that person's heart comes to have the function of making throbbing sounds? These sorts of cases will be described in greater detail below (see Section 2, under "The problem of overbreadth").

that the future effect of one's action (e.g., health as a consequence of exercise) causes the person to act; rather, it is the person's mental *representation* of the future effect, together with her other beliefs and desires, that cause her to act as she does. Thus an indirect reference to the future effect is preserved within the causal explanation for the purposeful action, and hence there is no violation of the normal temporal order of causation.

To the extent that, in order for a "representation" to exist, it must exist within, or have been created by, a mind, then representationalist theories are also *mentalistic* (Bedau, 1990). The assumption that functions are based on mental representations leads to two opposing views about how entities in the natural world come to have functions, the *theological* view and the *eliminativist* view (although the latter might just as appropriately be called the "analogical" view, for reasons that will be discussed below). According to the theological view – most notably advocated by Aquinas (1914 [1269–73]) – biological entities have purposes (e.g., functions) because God make them with those purposes in mind. This assumption is the basis for the fifth argument for the existence of God presented in his *Summa Theologica* (Question 2; Article 3). Roughly, his argument is that since mindless biological entities clearly have purposes, and something can only have a purpose if it has a mind or is controlled by something with a mind, then they must be controlled by something with a mind:

We see that things which lack intelligence, such as natural bodies, act for an end, and this is evident from their acting always, or nearly always, in the same way, so as to obtain the best result. Hence it is plain that not fortuitously, but designedly, do they achieve their end. Now whatever lacks intelligence cannot move towards an end, unless it be directed by some being endowed with knowledge and intelligence; as the arrow is shot to its mark by the archer. Therefore some intelligent being exists by whom all natural things are directed to their end; and this being we call God. (Ibid., p.27)

Paley's (1839 [1802]) famous design argument for the existence of God rests on a similar perplexity about how things that appear so well formed for a specific purpose could have been products of anything but intelligent design. Contemporary advocates of the theological view of functions include Plantinga (1993, Chapter 11) and Rea (2002, Chapter 5).

Proponents of the *eliminativist* view also believe that functions are based on prior representations, and therefore *if* anything in nature has a function it must have been created for that purpose by an intelligent being. But they argue that appeals to supernatural creation have no place in the context of scientific explanations. Therefore, to the extent that one accepts this stricture on scientific explanation, then one must also accept that biological entities do not "really" have functions (or refuse to countenance them in one's explanations) since they are not typically designed with purposes in mind.<sup>2</sup> Accepting this eliminativist position with respect to the existence of function does not, however, imply that scientists should never *ascribe* functions to biological

2 To say that natural entities are "not created with purposes in mind," excludes, of course, the effects of deliberate human manipulation, such as genetic engineering or artificial selection through breeding. Therefore, terminologically it is probably accurate to distinguish artificial functions and natural (rather than "biological") functions, where "natural" is intended in the sense of "not created or brought about by deliberate or conscious effort."

entities or that it is illegitimate or counterproductive to do so. They may legitimately do so, so long as they recognize that such usage is metaphorical (e.g., it involves examining biological forms "as if" they were created for a purpose) and that it performs a purely heuristic role in stimulating actual scientific theories.

Kant's *Critique of the Power of Judgement* of 1789 contains the classic statement of this eliminativist view. Although he expresses different views on natural ends (*Naturzweck*),<sup>3</sup> one view that he expresses is that biological purposiveness is based on a prior representation: "Here I understand by *absolute purposiveness* [*Zweckmäßigkeit*] of natural forms such an external shape as well as inner structure that are so constituted that their possibility must be grounded in an idea of them in our power of judgement" (2000 [1789], p.20; see Section VI of first introduction). In the case of natural ends, then, teleology presupposes the existence of a mind that can represent biological forms prior to creating them. Such a postulate, however, cannot enter into a causal explanation for the existence of such traits, since one of the *a priori* constraints on causal explanation is that both cause and effect must themselves be objects of the natural world. Causality cannot be a relation between the supernatural and the natural world, so long as one is operating within the perspective of natural science, since such a relation is not a possible object of experience: "But purposiveness in nature, as well as the concept of things as natural ends, places reason as cause into a relation with such things, as the ground of their possibility, in a way which we cannot know through any experience" (*ibid.*, p.35; see Section IX of first introduction). Therefore, a function ascription has the status of a heuristic device for scientific research, or what Kant refers to as a *regulative*, rather than constitutive, principle: it can guide the formation of scientific hypotheses or the discovery of new evidence but it does not enter into the content of those hypotheses or the evidential statements (*ibid.*, p.37; also see § 61).

A similar representationalist view, according to which the ascription of functions to the natural world rests upon an analogy to conscious design, is also adopted by the emergentist C. D. Broad (1925, p.82), and it finds more contemporary adherents in Woodfield (1976), Schaffner (1993, pp.403-4), Nissen (1997), and Ruse (1989, p.152). See Bedau (1990) for a critique of mentalistic views.

It was noted above that representationalist theories of function are almost always construed as *mentalistic* theories. Can there also be *non-mentalistic* representational theories of function, where representation is analyzed without appeal to minds? A possible such theory is associated with the distinction between "teleology" and "teleonomy." The term "teleonomy" was coined by the evolutionary biologist Pittendrigh (1958, p.394), to refer to systems that are in some sense "end-directed," but where this end-directedness does not rely on the problematic metaphysical assumptions associated with the word "teleology" – those of final causation or divine creation. However, he does not explicate his use of "end-directedness" or "goals." Mayr (1961, 1974), therefore, should primarily be credited with developing the concept of "teleonomy."

According to Mayr, a process or behavior is "teleonomic" if it is controlled by an internal "program" (1974, p.98). He defines a "program," in turn, as "coded or prear-

3 The following is a very partial account of Kant's view, and neglects his important phenomenological descriptions of the self-organization of living matter (e.g., §65), a phenomenon that he believes to warrant teleological explanation.

ranged information that controls a process (or behavior) leading it toward a given end" (ibid., p.102). Clearly, Mayr's analysis does not eliminate appeal to teleological concepts – such as "being led toward an end." Nonetheless, it does not seem implausible to suggest that the operative concept behind his formulation, like that of the mentalistic view, is that of "representation" – insofar as saying that one thing "carries information about" another thing seems tantamount to saying that the first thing *represents* the second. If this is true, then a teleonomic process might be equivalent to one that tends to develop along a specific trajectory, or into a specific form, by virtue of the fact that it is controlled, in part, by a non-mentalistic *representation* of that trajectory or form. Moreover, he clearly intends that segments of DNA that have been retained by natural selection, as well as neural structures that are shaped in some appropriate way by experience, qualify as containing "coded information." Hence, his analysis would require a naturalistic explication of "information" or "representation" that is appropriate for the biological context and that picks out the structures in question.

The feasibility of providing a naturalistic explication of biological information is defended by Maynard-Smith (2000), Sarkar (2000), and Sterelny (2000), as well as in the context of the "teleosemantic" account of information developed by, e.g., Stampe (1977), Enc (1982), and Millikan (1984) [SEE BIOLOGICAL INFORMATION]. However, most of the analyses depend centrally upon the concept of "function," and consequently cannot be used as part of an explication of "function" itself. The problem of defining a concept of "representation" that does not appeal to "function" is that representation, like function, is often assumed to be a *normative* concept. In other words, a representation can *misrepresent* something, just as a part of a system can *malfunction* (Millikan, 1984, p.17; Dretske, 1986). Hence it is sometimes suggested that the concept of function can be used to explicate the concept of representation, since it may be possible to explain the normative nature of representations by assuming that they have *functions*. A "misrepresentation," on this account, would be something like a sign that fails to perform its function. Moreover, since functions seem to be much more widespread in nature than representations (the heart has the function of pumping blood without being a representation of anything, whereas, plausibly, most representations have the function of guiding behavior), then defining representation in terms of function seems more likely to succeed than defining function in terms of representation.

## 2.2. *Non-representationalist theories of function*

Whereas representationalist views resolve the problem of backwards causation by seeking the origin of the functional entity in a prior mental representation, non-representationalist views seek to explain why such entities *currently* exist by appeal to entities of the same type that existed in the *past* and that, by virtue of producing the effect in question, were able to persist over time or to reproduce their kind. On this view, the function of an entity is that effect that entities of its kind produced in the past, which, in turn, contributed to the persistence or reproduction of that entity or type of entity. Thus, non-representationalist theories solve the problem of backwards causation by invoking a cyclical dimension: *X* did *Y* at time  $t_0$ , and as a consequence, *X* was able to continue to do *Y*, or *X*, by virtue of doing *Y*, was able to produce entities of the same type as *X* at time  $t_1$ . Such cyclical modes of production are sometimes referred to

as "consequence-etiologicals" (Wright, 1976, p.116), because one of the consequences that the functional item produces figures into an etiological account of why it continues to exist at a later time.

The most obvious example of a process that generates consequence-etiologicals is natural selection, since the reproduction of heritable traits that have higher relative fitness than alternate traits explains the maintenance of the former within a population of reproducing entities. Several biologists throughout the twentieth century drew attention to the connection between teleological statements and natural selection, and stated explicitly that the existence of natural selection can justify the use of teleology in biology.<sup>4</sup> Perhaps the earliest reference comes from the neuroscientist Charles Sherrington, in his *The Integrative Action of the Nervous System* (1906). In that work, Sherrington pauses to reflect on his oft-repeated use of teleological terms such as "purpose," and his considerations suggest strongly that he identifies the purpose of a reflex with what it was selected for:

That a reflex action should exhibit purpose is no longer considered evidence that a psychological process attaches to it; let alone that it represents any dictate of "choice" or "will." In light of the Darwinian theory every reflex *must* be purposive. We here trench upon a kind of teleology . . . The purpose of a reflex seems as legitimate and urgent an object for natural inquiry as the purpose of the colouring of an insect or a blossom. (Ibid., pp.235-6)

The ethologist Konrad Lorenz makes a similar remark in his 1963 book, *On Aggression*:

If we ask "What does a cat have sharp, curved claws for?" and answer simply "To catch mice with," this does not imply a profession of any mythical teleology, but the plain statement that catching mice is the function whose survival value, by the process of natural selection, has bred cats with this particular form of claw. Unless selection is at work, the question "What for?" cannot receive an answer with any real meaning. (Lorenz, 1966 [1963], pp.13-4; cited in Griffiths, 1993, p.412)

The evolutionary biologist George Williams also emphasizes this point: "The designation of something as the *means* or *mechanism* for a certain *goal* or *function* or *purpose* will imply that the machinery involved was fashioned by selection for the goal attributed to it" (1966, p.9).<sup>5</sup>

None of these accounts, however, state *why* explanations based on natural selection fit the pattern of teleological explanations—they simply express, as it were, the basic intuition that they do, without articulating a rationale. Perhaps the first attempt to explicitly justify this view is found in the work of the evolutionary biologist Ayala (1968, p.217; 1970, pp.40-1), who points out that in a selectionist explanation, an

4 Lennox (1993) argues that Darwin himself implicitly uses teleological terms such as "end" and "purpose" to describe the outcome of selection processes (ibid., p.415), though Darwin never explicitly states this fact about his usage.

5 It is ironic that the etiological theory was primarily developed by biologists, since one of the main arguments *against* the etiological analysis is that it does not correspond to actual biological usage! (See Section 3).

effect that an entity produces figures into an explanation of why that type of entity currently exists, and this, by definition, constitutes a teleological explanation. Wimsatt (1972) provides a comprehensive philosophical analysis of the logical structure of function statements and argues that insofar as function statements are construed as teleological explanations, selection processes are the only known and plausible way in which such statements can be justified: “[T]he operation of selection processes is not only *not* special to biology, but appears to be at the core of teleology and purposeful activity wherever they occur” (ibid., p.13).<sup>6</sup> More famously, Wright (1973, p.161; also see Wright, 1972) defines “function” in terms of these consequence-etiologicals and argues that natural selection can justify function statements (Wright, 1973, p.159).<sup>7</sup>

Several different theories of function stem from this basic insight, and much of the philosophical literature on functions consists in the attempt to ramify, extend, and qualify this viewpoint. Three major challenges to this etiological view, and some of the responses to these challenges, will be presented in order to elucidate the ways in which the position has been developed over time.

(i) *The problem of overbreadth.* The first problem can be understood as a response to Wright’s (1973) influential view, although in some form or another it continues to plague etiological theories. According to Wright’s explication:

The function of *X* is *Z* means

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

(b) *Z* is a consequence (or result) of *X*’s being there. (1973, p.161)

In the artifact context, *X*’s form can be explained by the fact that somebody recognized that form to have a certain capacity (*Z*), and produced it for that reason, thereby fulfilling the first premise. In the biological context, if *X* was selected for by virtue of one of its effects, *Z*, and this selection process partly accounts for its present existence, then it will be true to say that “*X* is there because it does *Z*,” thereby also satisfying the first premise. If *X*’s being there allows it to *continue* to do *Z*, then the second will be fulfilled as well. Clearly, the purpose of Wright’s fairly general analysis is to present the idea of a cyclical causal process, one that incorporates both natural and artifact functions.

But Wright’s general definition is also satisfied by processes that, intuitively, one would not want to ascribe functions to, such as the sort contrived by Boorse (1976) in his critique of Wright. Suppose, for example, that a hose in a laboratory springs a leak, and thereby emits a noxious chemical, and any scientist that attempts to seal the hose gets knocked unconscious by the chemical it emits. Thus it can be said that the leak in the hose contributes to its own persistence by knocking out anyone that comes close enough to fix it (ibid., p.72). But it seems counterintuitive to say that knocking out scientists is the function of the leak, or that the leak has any function at all. Similarly, obesity can contribute to a sedentary lifestyle, which in turn can reinforce obesity. Thus

6 However, he hesitates to build this insight into a *conceptual analysis* of “function,” since he comes up with counter-examples that purport to show that being selected for is, strictly speaking, neither necessary nor sufficient for having a teleological function (Wimsatt, 1972, pp.15–16).

7 Wright (1973), like Wimsatt (1972), does not define “function” explicitly in terms of selection, but claims that having been selected for, in fact, suffices for having a function.



it is possible to explain a person's current obesity in terms of one of the consequences his or her obesity produced in the past that contributes to its own persistence (*ibid.*, pp.75–6). Yet, like the hose example, it seems bizarre to suggest that the function of obesity is to contribute to a sedentary lifestyle.<sup>8</sup>

Boorse's counterexamples have been influential in shaping the development and refinement of etiological theories of function, since they have led many to accept that having been selected for by natural selection, rather than merely having contributed to the continuation of one's present state, is a necessary condition for having a function (see, e.g., Neander, 1983, p.103; Millikan, 1993, pp.34 – 6; Boorse himself (1976, p.76) raises this possibility but rejects it). This view will be referred to as the "selected effects" (SE) theory of function, and some version of it is probably the most widely held theory of function amongst philosophers (Neander, 1983, 1991; Millikan, 1984, 1989a, 1989b, 1993; Brandon, 1990; Griffiths, 1992, 1993; Godfrey-Smith, 1994; Mitchell, 1993, 1995; Allen & Bekoff, 1995a, 1995b). Obesity, though it secures its own persistence by contributing to a sedentary lifestyle, is in no sense *selected over* some other phenotypic trait *because* it contributes to a sedentary lifestyle. Similarly, the leak in the hose is not there because *it*, rather than something else, proved to be more effective in knocking out scientists. This also resolves the problem, noted earlier (fn. 2), of the function of heart sounds – since even if the heart's beating sounds help to protect the heart by alerting physicians to potential heart problems, the heart was not selected for because it makes these sounds. However, by introducing natural selection as a necessary condition on function ascriptions, Wright's theory loses some of its generality, and this is the basis for the second criticism.

(ii) *The problem of conceptual divergence.* The problem of conceptual divergence has two forms. First of all, it is not clear how the SE theory of function adequately explains the functions of artifacts, and hence it entails the existence of a conceptual divergence between artifact "functions" and biological "functions" that is not intuitively obvious. Certainly, *some* types of artifacts undergo a certain selection process, where, over a significant period of time, certain features of its form are replicated, others are modified, and still others are extinguished. Nonetheless, functions are typically ascribed to artifacts on their *first* appearance, and that is because the intention of the designer suffices to give an artifact its function. It does not seem that this can be reconciled with the SE view.

Some philosophers have attempted to lessen this discrepancy by suggesting that the process of *designing* an artifact is akin to natural selection, in that the designer typically *imagines* variations on a given form, and chooses to actualize only that one that is most suitable to his or her purposes. Hence a type of "virtual" selection process takes place (Wimsatt, 1972; Griffiths, 1993). For example, Wimsatt (1972, p.15) raises the possibility of "mental trial and error" in his attempt to assimilate artifact functions to his model of biological functions, and show that some concept of selection over a range of

8 Bedau (1992, p.786) uses the example of a stick floating down a stream that brushes against a rock and gets pinned there by the backwash it creates, and thus is responsible for perpetuating its current position, to make the same point. Clearly, such examples can be multiplied indefinitely.

alternatives underlies both.<sup>9</sup> Another response has been to deny that an accurate explication of the concept of “biological function” must also account for the functions of artifacts (e.g., Godfrey-Smith, 1993, p.347). Perhaps the intuition that there exists a unified concept of function merely reflects the persistence of the “dead metaphor” that biological forms are the product of design (Lewens, 2004, p.13).

Regardless of whether or not the SE theory can successfully assimilate artifact functions, its generality appears to be quite limited in a second way, namely, historically. Harvey, for example, discovered that the function of the heart is to circulate the blood, and he *believed that* he discovered its function: “it is absolutely necessary to conclude that the blood in the animal body is impelled in a circle, and is in a state of ceaseless movement; that this is the act or function which the heart performs by means of its pulse; and that it is the sole and only end of the movement and contraction of the heart” (Harvey, 1894 [1628], p.72). But he did not possess the theory of natural selection. Therefore, if the SE view is accurate, then Harvey meant something altogether different when he spoke of the function of the heart than what modern biologists mean (Frankfurt & Poole, 1966, p.71; Boorse, 1976, p.74; Nagel, 1977, p.284; Enc, 1979, p.346).

One response has been to argue that the SE theory is only intended to be accurate as a conceptual analysis of *modern biological usage* (Neander, 1991, p.176), regardless of whether it captures lay or historical usage. It has also been argued that the goal of explicating “function” is not to provide a *conceptual analysis* at all, but rather, a theoretical definition of “function” (Millikan, 1989b, p.293), in the same way that being H<sub>2</sub>O constitutes a theoretical definition of “water.” But since theoretical definitions are themselves often tantamount to conceptual analyses of modern scientific usage, the two responses are similar. Schwartz (2004) goes further by emphasizing the stipulative and constructive roles of philosophical definitions of “function,” arguing that such definitions constitute *explications* of biological usage, rather than conceptual analyses or theoretical definitions. According to Carnap (1950, see chapters 1 and 2), philosophical explication involves the replacement of a vague concept by a precise one, and hence it often entails making distinctions that did not previously exist in the scientific context in question. It has the character of a proposal, to be accepted or rejected on pragmatic grounds.

The attempt to justify the SE view by appealing to modern biological usage gives rise to a different problem, which is that modern biologists don’t always, or even typically, use “function” with *any* etiological import (Amundson & Lauder, 1994; Godfrey-Smith, 1994, p.351; Walsh, 1996, p.558; Schlosser, 1998, p.304; Wouters, 2003, p.658; Sarkar, 2005, p.18; Griffiths, 2005). Although, as noted above, biologists sometimes *do* use “function” more or less synonymously with “adaptation,” in many contexts “function” is tied more closely to the current survival value of a trait. For example, as Godfrey-Smith (1994, p.351) points out, according to an influential set of distinctions introduced by Tinbergen (1963), the field of behavioral ethology is largely concerned with four questions concerning behavior: its (proximate) causation, its survival value,

9 However, he also entertains the possibility that an omniscient being, if one exists, might never have to consider a range of alternatives before acting, and yet the actions would nonetheless be purposeful – and thus that it is *conceivable* that the actions of this being could be explained teleologically without being the product of a selection process!

its evolution, and its ontogeny (*ibid.*, p.411). In Tinbergen's usage, "survival value" is synonymous with "function," and explicitly separated from the question of evolution, and in particular, from the selective history of a behavior (*ibid.*, p.423). Mayr (1961), similarly, distinguishes "functional biology" and "evolutionary biology," arguing that the former is concerned with the realm of "proximate causes" and the latter, "ultimate causes" of an entity or process, whereas, according to SE, "function" describes only the realm of ultimate causes.

Even more broadly, "function" is often used to characterize the entire range of activities that a part of a system is capable of performing (e.g., the sense in which "function" is opposed to *structure*). For example, the evolutionary morphologists Bock and Von Walther (1965, p.274) define the function of an entity simply as "all physical and chemical properties arising from its form," provided that these properties are not relative to the environment, and Amundson and Lauder (1994) argue that this more liberal usage is standard in anatomy, comparative morphology, and physiology. This makes the use of function statements in those disciplines heavily dependent upon the interests of the investigator, since without at least imposing a pragmatic restriction on the appropriate use of function statements, virtually every structure in the natural world can be said to possess a "function." Given these multiple salient uses within biology, the most reasonable attitude to adopt seems to be a pluralistic one (e.g., Millikan, 1989a; Kitcher, 1993; Godfrey-Smith, 1994; Amundson & Lauder, 1994).

(iii) *The problem of vestiges*. A third criticism is that SE does not seem to allow for the possibility of *vestigial* traits, which are traits such as the human appendix which once possessed functions but have ceased to perform them for so long that they are said to be functionless (Boorse, 1976, p.76; Prior, 1985). The rudimentary ocular cyst of the cave-dwelling fish, *Phreatichthys andruzzii*, is not a dysfunctional eye, but a functionless vestige – even though at some point the organ had been selected for because of sight. But if the vestigial trait had *ever* been selected for, however distantly in the past, then its past contribution to the fitness of ancestral organisms figures into a complete explanation for its present persistence in the population. Therefore, without imposing any temporal restrictions on the explication of "function," it is not clear how that explication can capture the idea that a heritable trait, though it once possessed a function, no longer does, but has been retained because, e.g., the relevant mutations that would have allowed it to atrophy or be replaced never arose.

Another case which supports the need for introducing temporal restrictions on function ascriptions is the case of functional co-optation, in which a trait that initially spread within a population by selection for one of its consequences eventually came to be maintained by selection for something else, or in which a trait that was initially not selected for at all came to be selected for in a new environment. This distinction partly overlaps Gould and Vrba's (1982) well-known distinction between adaptation and exaptation, where a trait is an *adaptation* if it was "built by selection for its current role (*ibid.*, p.6)," and an *exaptation* if it was later "co-opted" for a useful role that it was not originally selected for. For example, plant species of the genus *Dalechampia* probably first used resin secretions as a defense against herbivores; later, they became used as a reward system for pollinators (Armbruster, 1997). Exaptations are ubiquitous in the biological realm and render problematic any simplistic attempt to infer the selective

history of a trait from its current contribution to fitness. SE must possess the resources to conceptualize such transitions appropriately.

Perhaps the most widely accepted etiological approach is that which identifies the function of a trait with the effect for which it was selected in the *recent evolutionary past* (Griffiths, 1992, 1993; Godfrey-Smith, 1994). But how should such a temporal unit be defined? Griffiths (1992) defends a version of SE according to which the trait in question must have contributed to its maintenance in a population during the last “evolutionary significant time period” for that trait, and he defines an evolutionarily significant time period for a trait, *T*, as that time period during which, given the mutation rate at the loci controlling *T*, and the population size, one would have expected some regression (atrophy) of *T* were it not making some contribution to fitness (ibid., p.128). Godfrey-Smith (1994), while introducing the expression “modern history theory of functions,” leaves the determination of such a unit implicit.

Two other important developments within the structure of etiological views are worth noting before describing consequentialist views. The first is a distinction between “function” and “design,” the importance of which is argued for in Allen and Bekoff (1995a, 1995b; also see Kitcher, 1993 and Buller, 2002, who elaborate notions of “design” in relation to which functions are identified). Unlike the concept of function, which can be used broadly to encompass whatever a trait was selected for, the concept of design, they claim, should only be applied to that subset of functions that partly explain the *structural modification of a trait* over time (1995a, p.615). They point out that what something is an “adaptation” for (in Gould and Vrba’s sense) is often what it is “designed” for, and that “exaptations” will often correspond to traits which merely have “functions” but were not designed, since they did not undergo any additional structural modification to perform the exapted function.<sup>10</sup>

A second distinction that is useful is that between the “strong” etiological theory and the “weak” etiological theory, which has been implicit in much of the literature but only articulated by Buller (1988, 2002). According to the strong etiological theory, a function of a trait is an effect that, in the past, the trait was selected for (hence it is identical to SE). According to the weak etiological theory, however, the function of a trait is an effect that, in the past, contributed to the fitness of its bearer and thereby contributed to its own reproduction, regardless of whether it was selected for—that is, regardless of whether the requisite variation existed upon which selection could act, or whether existing variation was correlated with differential reproduction. Another way of formulating the distinction is that the strong etiological theory emphasizes the contribution of a trait to *differential* reproduction; the weak etiological theory emphasizes reproduction as such. Both theories, clearly, only ascribe functions to heritable traits.<sup>11</sup>

A simple example drawn from Dover (2000, p.41) can help to clarify the distinction. Suppose that, in a small population, genetic drift carries an allele to fixation at  $t_0$ .

10 See Buller (2002), however, who argues that their distinction between “function” and “design” is unprincipled, because whether something is *designed for X*, or merely has the *function of performing X*, often depends upon purely conventional decisions about how selection pressures should be individuated.

11 Buller (2002, pp.230–3) points out that it is not uncommon for philosophers to vacillate between the two forms.

Although that allele has a phenotypic effect, it did not confer any fitness advantage on its possessors. Now suppose that, at  $t_1$ , the environment changes in such a way that possession of the allele is necessary for survival. Even though all of the individuals within the population have the allele – so there is no selection for it – they all would have perished at  $t_1$  had any of the alternate alleles gone to fixation at  $t_0$ . Thus, at  $t_2$ , it can be said that one of the reasons that the allele currently exists is because it produces the effect in question. Consequently, the scenario satisfies the pattern of teleological explanation. But since selection did not enter the scenario, the strong etiological theory does not bestow a function upon the trait, since at  $t_1$ , the requisite variation did not exist upon which selection could act, and at  $t_0$ , the differential reproduction of alleles was not correlated with differential fitness. Hence, the weak etiological theory is clearly more liberal with respect to the range of evolutionary mechanisms that it considers function bestowing, yet it still permits teleological explanation. Finally, since it only ascribes functions to heritable traits, it avoids the Boorse-style counterexamples described earlier.

### 3. Consequentialist Theories of Function

Despite the plurality of etiological theories, and despite the attempts to render etiological theories more consistent with modern biological usage, it is often pointed out that typically, when biologists seek to determine the function of an entity, they look to some subset of current dispositions or capacities of that entity rather than to the fossil record. This suggests that despite the modifications that can be imposed on the etiological theory to render it more compatible with biological usage, it does not adequately capture that usage. Thus, some argue, functions, whatever else they may be, must be thought of as current dispositions or consequences of traits, and hence function ascriptions cannot provide causal explanations for the current maintenance of a trait in a population. As noted above, consequentialist theories of function almost invariably conceive of the function of an entity as consisting in its contribution to something else, or its disposition to so contribute. Insofar as functions, in the biological context, are typically ascribed only to *parts* of systems (rather than to the system as a whole), then according to consequentialist theories the function of a trait is typically thought to consist in its contribution to some property or capacity of a more inclusive system – e.g., the contribution of a trait to the fitness of the organism. Hence, in the following, consequentialist theories will be classified according to the *sort* of systemic property or capacity which performance of the function contributes to bringing about or maintaining.<sup>12</sup> In the following, four types of contribution theories will be described:

12 It is not always the case that consequentialist theories define the function of an entity in terms of its contribution to something else. As noted above, according to one liberal biological conception of function, the function of a structure consists of the totality of effects it produces, independently of reference to the environment (Bock & von Wahlert, 1965, p.274). In this theory there is no sense in which a function contributes to anything else, much less a containing system. By the same token, it is not always the case that when a functional entity does contribute to a system, that system is its own inclusive system. This is most obviously true in the case of artifacts, which are typically not “part” of the person who uses them (see Wright, 1973, p.145).

interest-contribution theories; goal-contribution theories; good-contribution theories; and fitness-contribution theories.

### 3.1. *Interest-contribution theories*

The most general contribution theory is the interest-contribution theory, according to which the function of an entity consists, roughly, in its contribution to bringing about or maintaining some property of a system that is of interest to an investigator. The most well-known proponent of this theory is Cummins (1975; also see 2002) – so well known, in fact, that such functions are often simply referred to as “Cummins functions” (Millikan, 1989a; Godfrey-Smith, 1993), or even “C-functions” (Walsh & Ariew, 1996). However, as will be elaborated below, Cummins’ own view could be appropriately referred to as the “systemic capacity” view, because it restricts functions to the components of complex and hierarchically organized systems.

Cummins (1975) claims that most prior analyses of “function” were flawed because they overlooked the fact that functions refer primarily to a distinctive *style* of explanation (“functional analysis”), and only secondarily to a distinctive *object* of study (e.g., organismic fitness) (ibid., p.756). In keeping with this methodological approach, to ascribe a function to a *part* of a system is to ascribe a capacity to that part, and this capacity is picked out because it plays a salient role in an analytical account of a capacity of the system itself. In this sense, there is nothing mysterious about function ascriptions, since they do not imply that an effect of a trait explains that trait’s existence; rather, they merely show how a trait produces the effect in question. This analytical strategy constitutes a special style of scientific explanation, however, because it explains a complex capacity of a system by drawing attention to the simpler capacities of its subsystems and showing how they are organized in such a way as to yield the complex capacity. The more complex the capacity under investigation, the more complex the organization of the system, and the simpler the subsystem capacities, the more interesting such an explanation is. Nonetheless, the appropriateness of function statements is always relative to someone’s explanatory interest, even if such ascriptions are not particularly *interesting*.

Hempel (1965 [1959]) and Lehman (1965) appear to hold an early version of the interest-based view. According to Hempel, the function of a system *part* consists, roughly, in its contribution to fulfilling some condition which is necessary for the “adequate, or effective, or proper working order” of the system as a whole (ibid., p.306). Hempel, however, does not attempt a definition of “proper working order”; his view is that each scientific discipline that uses function statements, whether it be biology, psychology, or sociology, must operationalize the notion of “proper working order” in its own terms, and hence his concept of function is explicitly relativized to the explanatory and disciplinary context at hand (ibid., pp.321–2). Similar views that emphasize the explanatory or pragmatic context of function statements are held by Prior (1985), Amundson and Lauder (1994), Hardcastle (1999), Davies (2001), and Craver (2001).

Because of the fact that, according to these views, functions are only limited by the interests – epistemic or pragmatic – of the investigator, they are often accused of overbreadth. On the one hand, “functions” could be ascribed throughout the non-organic

world. For example, a particular arrangement of rocks can have the “function” of contributing the widening of a river delta downstream from it (Kitcher, 1993, p.390), and clouds can have the function of promoting vegetation growth (Millikan, 1989b, p.294). On the other hand, functions can be applied to entities that are clearly malfunctioning or maladaptive; as Cummins himself points out, the appendix keeps people vulnerable to appendicitis but it sounds strange to call this one of its functions (Cummins, 1975, p.752) – even though medical researchers are clearly interested in providing an analytical account of how this takes place! Yet these criticisms seem to misconstrue Cummins’ insistence on the methodological, rather than substantive, character of functional analysis. Certainly, if, on the systemic capacity theory, function ascriptions were primarily intended to perform the substantive role of delineating a special type of system, then the liberality objection would be well taken, since such ascriptions would be vacuous. But since functional analysis is held to mark a style of explanation, then the liberality objection does not hold – it would be tantamount to suggesting that “conceptual analysis” is too liberal because, in principle, it applies to any concept!

### 3.2. Goal-contribution theories

According to goal-contribution theories, the function of a part of a system consists in its contribution to a *goal* of that system. The notion of a “goal” or of a “goal-directed system” occupied a significant place in philosophical approaches to teleology from the 1940s through the early 1970s (Rosenblueth et al., 1943; Sommerhoff, 1950, 1969; Braithwaite, 1953; Nagel, 1953, 1961; Beckner, 1969; Manier, 1971). However, it largely fell out of favor among philosophers of biology in the early 1970s, partly owing to the predominance of evolutionary considerations within that tradition and partly owing to internal conceptual shortcomings (Wimsatt, 1972; Ruse, 1973; Hull, 1973). In short, a goal-directed system is one that exhibits a capacity to attain a specific value for some system variable, or to maintain the variable within a range of values, in the face of environmental perturbation, via the existence of compensatory activity operating amongst the system’s parts. The maintenance or attainment of a given value for the system variable is considered the *goal* of the system, and the specific contribution of a part of the system to that goal is considered to be the *function* of that part (Boorse, 1976, p.77; Nagel, 1977, p.297). Thus any system may have several goals; additionally, any sufficiently complex system can be analyzed as a hierarchy of goal-directed systems. Boorse (1976, 2002) advocates a goal-contribution theory and claims that individual survival and reproduction constitute the “apical goals” of the organism (2002, p.76); hence his general theory of function is largely coextensive with the fitness-contribution view when instantiated in the biological context.

Two paradigmatic cases of “natural” or “mechanical” purposiveness largely inspired this approach to teleology: homeostatic mechanisms drawn from physiology and servomechanisms that constitute the subject matter of cybernetics. As an example of the first type of mechanism, the percentage of water in the blood remains at around 90 percent throughout an individual’s lifetime. This is because if it drops far below this level, the muscles increase the rate at which they infuse the blood with water; if it rises far above this level, the kidneys increase the rate at which they extract water from the blood. In this manner, the constancy of the water level of the blood is not a static

phenomenon; it is actively maintained via compensatory mechanisms that operate throughout the body in the face of perturbation. Servomechanisms, such as heat-seeking missiles, exhibit a similar capacity to actively maintain a specific trajectory in the face of perturbation, and to adapt that trajectory to the moving position of the target. The oft-repeated slogan that goal-directed systems exhibit “plasticity” and “persistence” (e.g., Nagel, 1977, p.272; Enc & Adams, 1992, p.650) captures two central features of the concept of goal-directedness. On the one hand, such systems exhibit *plasticity* in that the same effect can be reached from a number of initial systemic configurations and by virtue of a number of different mechanisms or pathways. On the other hand, such systems *persist* in their course of action to the extent that they have the ability to attain or maintain a course of action in the face of environmental perturbation.<sup>13</sup>

Since negative feedback systems are capable of exhibiting self-regulation, the concept of goal-directedness has often been analyzed narrowly in terms of *negative feedback* (Rosenbleuth et al., 1943; Manier, 1971; Adams, 1979; Faber, 1984; but see Wimsatt, 1971, for criticism of the concept of “feedback”). However, theories of goal-directedness that emphasize the compensatory and self-regulatory activity of systems are not necessarily tied to negative feedback. Hull (1973) points out that a system can exhibit the plasticity required to be goal-directed without being guided by negative feedback. For example, if the kidney does not succeed in ridding the body of excess water, then sweating may do so, but the different responses are not clearly regulated by a single negative feedback system (ibid., pp.110–11). (Nagel, 1953, p.211, Sommerhoff, 1969, pp.198–9, and Schlosser, 1998, p.309 also point out limitations of the negative feedback model for analyzing goal-directedness.)

Recently, Schlosser (1998) adopted some of the basic insights from the goal-supporting theory while rejecting its association with negative feedback (ibid., p.309) – although, strictly speaking, his theory should not be conflated with a goal-contribution view. According to his view, if a state or property of a system has a function then there exists a set of circumstances under which it is necessary for its own “reproduction” – that is, its trans-generational reproduction or intra-generational persistence (ibid., p.326). However, in order to avoid the Boorse-type counterexamples described above, he stipulates that the system in question must be capable of *complex* self-reproduction—that is, the system must be capable of reproducing the state in different ways, depending upon the environmental circumstances (ibid., p.312). Hence his view incorporates the plasticity criterion associated with goal-supporting theories while leaving fairly open the mechanisms by which this plasticity is realized.

Two main problems afflict goal-contribution theories, the “problem of vacuousness” and the “problem of goal-failure.” The problem of vacuousness stems from the fact that the standard characterization of a goal-directed system as one that exhibits “plasticity

13 It is sometimes argued that the goal-supporting account does not allow one to determine a system *goal*, and consequently, that this goal must be arbitrarily stipulated (Wimsatt, 1972, pp.20–2; Schaffner, 1993, pp.367–8; Schlosser, 1998, p.327). However, the above examples show this claim to be inaccurate. In the homeostatic case, *that* maintaining the water content of the blood at around 90% qualifies as a “goal” of the system is a consequence of the definition of “goal” and a rudimentary understanding of physiology, and need not be arbitrarily stipulated.



and persistence" with respect to a given end is not sufficient for imposing a substantive distinction between different types of systems, for almost all systems can be described as seeking an equilibrium state which can be reached from different initial states and in different ways (Wimsatt, 1971; Woodfield, 1976; Nissen, 1980-1; Bedau, 1993). A pendulum swinging to a state of rest, a ball rolling from the top of a bowl to the bottom, and an elastic solid returning to its original condition after the imposition of tension would all represent goal-directed systems. Consequently, unless one specific mechanism, such as negative feedback, is included within the definition, it is difficult to exclude such counterexamples. Sommerhoff (1950, p.86), and Nagel (1977, p.273), attempt to exclude such systems by imposing an independence condition on the variables, which roughly states that all of the controlling variables must be independently manipulable.

The problem of goal-failure stems from the fact that most explications of goal-directedness have tacitly or explicitly assumed that the supposed goal-directed behavior is successful, and as a consequence it is not clear how to explain the intuition that a non-conscious entity can have a goal and yet fail to satisfy it (Scheffler, 1959; Beckner, 1959; Hull, 1973). Manier (1971, p.234) and Adams (1979, p.506) address this problem by arguing that what makes a negative feedback system "goal-directed" is not that it actually achieves its goal, but that it is governed by an internal representation of the goal-state. (This brings the goal-contribution theory closer to an etiological theory such as Mayr's (1961, 1974), as described above.) This, however, raises the additional onus of providing a naturalistic account of "representation" that does not itself appeal to function.

### 3.3. *Good-contribution theories*

The core idea behind good-contribution theories of function is that in order for an entity to possess a function, performance of that function must (usually or typically) have a *beneficiary*. It must be useful for, beneficial for, or otherwise represent a "good" for some agent or system. This type of teleology is fairly evident in the world of artifacts, because artifacts are produced for a purpose and hence for an end deemed useful or beneficial by someone. Consequently, the good-contribution view is closely associated with the mentalistic view described above. However, this doctrine is not identical with mentalism, because it is not incoherent to ascribe "interests" or "goods" to biological entities that cannot be said to possess the sort of mental life required by mentalism.

Canfield (1964), for example, defines the function of an entity simply as some useful contribution it makes to a system: "A function of *I* (in *S*) is to do *C* means *I* does *C* and that *C* is done is useful to *S*" (ibid., p.290). In the biological context, he argues, the "usefulness" of a trait can be identified with its making a contribution to the survival or reproductive capacity of its bearer (ibid., p.292). Sorabji (1964) also expounds a good-contribution theory, and he argues that Plato and Aristotle hold this view. Ayala (1970) amends his etiological analysis by incorporating the concept of "utility" into his account: a feature of a system is "teleological" if it possesses "utility for the system in which it exists and such utility explains the presence of the feature in the system" (ibid., p.45). Thus, although strictly speaking, Ayala's position is an etiological one, it also incorporates the concept of benefit. Bedau (e.g., 1991, 1992, 1993) is the most

prominent current defender of the good-contribution theory; also see McLaughlin (2001; especially chapter 8) for a recent defense of the view that any adequate theory of function must incorporate such a "welfare" provision.

Presumably, one of the main advantages of such a view is that it appears to bridge the divide between natural functions and artifact functions, for, whereas artifact functions are "useful" by virtue of conscious design, natural functions are "useful" by virtue of their fitness contribution. In other words, the same concept is instantiated differently depending on the context, and hence there is no deep conceptual divergence between the usages. Moreover, as Bedau (1992) points out, this solution would resolve some of the Boorse-style counterexamples described in relation to the etiological view – for example, the stick that is pinned to the rock because of the backwash it creates does not have the "function" of creating the backwash, and that is because being pinned to the rock is not "good for anything" (ibid., p.787).

However, a significant problem with the good-contribution view is that it does not allow functions to be distinguished from "fortuitous benefits" or "lucky accidents." Frankfurt and Poole (1965), for example, criticize Canfield (1964) because heart sounds sometimes *do* have good consequences for fitness by alerting a physician to a potential life-threatening ailment, yet it does not have this as a function. (Wright, 1973, pp.145–6 and Bedau, 1992, p.787 also raise this problem.) One solution to this would be to incorporate a statistical component: in order to have a function, the activity in question must usually, or typically, contribute to some good. But as Millikan (1984, p.29) famously points out, statistical normalcy is not a reliable guide to functionality, since the probability that a given sperm will actually fertilize an egg is extremely low, yet fertilization is without doubt the function of sperm. Most sperm are quite literally good for nothing. Finally, of course, accepting something like the good-contribution view would most likely spell the death of the project of "naturalizing teleology," since the ascription of function would be explicitly value-relative, and values are notoriously difficult to situate within the natural world.

Bedau (1992, p.794), like Ayala (1970), suggests the possibility of a theory of biological teleology that conjoins the etiological view and the good consequence view and that would ameliorate the problem of fortuitous benefits. According to this view, a trait would come to possess a function because its persistence is partly explained by its past contribution to a *beneficial consequence* (e.g., increased fitness). However, he does not go so far as to offer an unqualified endorsement of this view, since the *goodness* of the result (increased fitness) does not itself perform an essential explanatory role in the etiology of the trait, but is only, as it were, externally linked to that explanation (ibid., pp.801–2). McLaughlin (2001), however, develops a similar view according to which, in order to have a function, a trait must have produced a beneficial consequence that contributed to its own persistence or reproduction (ibid., p.168).

### 3.4. *Fitness-contribution theories*

The basic, unqualified idea behind fitness-contribution theories is that the function of a trait consists in its contribution to the fitness of the organism (or, more generally, to the fitness of the biological system of which it is a part). Thus, according to this view, the ascription of a function to a trait does not explain why that trait currently exists,

although ascription of a function to *ancestral* tokens of a trait can play a role in an explanation for the *current* persistence of that trait. Fitness-contribution views are proposed by Canfield (1964), Lehman (1965), Ruse (1971, 1973), Bechtel (1986), Bigelow and Pargetter (1987), Horan (1989), Walsh (1996), and Wouters (2003, 2005) (although, as pointed out above, Canfield (1964) accepts this view insofar as he defines "function" in terms of utility and believes that the fitness contribution made by a trait is "useful" to the organism). Sarkar (2005, p.18) presents a generalization of this view, according to which a part of a system must merely contribute to the persistence of its containing system in order to have a function, and not necessarily to the reproduction of that system. This would allow functions to be assigned to the parts of, e.g., sterile organisms.

One problem with this unqualified view is that, in principle, fitness assignments can vary wildly depending upon fluctuations in the current environment, but function assignments tend to be relatively stable. For example, one can create an abnormal, transient environment in which a trait that is usually maladaptive possesses survival value, but it seems counterintuitive to say that the trait comes to possess a new function in that environment. Moreover, even traits that are, on average, adaptive in a given environment can, in certain environments, become maladaptive. But it is not said that in such an environment the trait no longer has a function, but that it is unable to perform its function.

Such counterexamples suggest that such function ascriptions should be relativized to a "normal" or "average" environment, in order to exclude abnormal or transient ones. This recognition led Bigelow and Pargetter (1987) to propose that a trait has a function when it bestows a survival-enhancing propensity on the organism that possesses it, in that organism's natural habitat (*ibid.*, p.192). Thus, their definition of function introduces a counterfactual element – if the trait *were* in its natural habitat, then it would, *ceteris paribus*, contribute to the fitness of its bearer. Yet this introduces further problems. Obviously, the "natural habitat" for an organism is not necessarily the organism's *current* habitat. But if not, then what constitutes an organism's natural habitat? One candidate for the natural habitat of an organism is that habitat in which it has, historically, flourished (Millikan, 1989b, p.300; Mitchell, 1993, pp.258–9; Godfrey-Smith, 1994, p.352; Walsh, 1996, p.562). But then the propensity theory of functions is rendered perilously close to some version of the etiological theory, since its incorporation of a historical component violates the spirit of the "forward-looking" view they endorse. Walsh (1996; also see Walsh & Ariew, 1996) attempts to eliminate the problem of defining the organism's "natural habitat" by proposing a relational theory of function, according to which the function statement must be relativized to a specific "selective regime," which may have occurred in the past or the present. Hence, in his view, there are no functions *simpliciter*; in order to assign a function one must state precisely the nature of the environment within which the trait contributes to fitness.

A similar problem stems from the following consideration. In order to estimate the contribution of a trait to fitness, one must compare the average fitness of organisms that possess the trait with the average fitness of those that do not. But if no variation for that trait currently exists—such as the human kneecap—then it is not clear what to compare its performance with (Frankfurt & Poole, 1965, pp.71–2; Wimsatt, 1972,

pp.55–61; Millikan, 1989a; Godfrey Smith, 1994, p.352). One possibility would be to compare it with the variation that existed at an *earlier* time. But again, this brings the propensity theory closer in spirit to the etiological view.

Wouters (2003, 2005) proposes a version of the fitness-contribution view according to which, in order to have a function, a trait must confer a biological advantage upon its possessor, relative to some actual or counterfactual set of variants. This resolves the problem insofar as one must explicitly stipulate the range of variation in question. Moreover, he argues that this reflects standard practice within some fields of biology. In optimality models of adaptation, for example, the relevant range of alternatives (the “phenotype set”) is typically derived from biologically informed assumptions about what is physically, ecologically, or physiologically possible (Parker & Maynard Smith, 1990, p.27; also see Wouters, 2005, p.43). However, merely stipulating the range of variants in question seems to introduce an element of arbitrariness into function ascriptions. Relative to one hypothetical set of variants, a trait has a function; relative to another set, it does not. Clearly, something more substantive should be said about how this range of variation can be non-arbitrarily determined.

As noted above, the main advantage of contribution-based theories is that they are more consistent with the majority of biological usage. Moreover, given the difficulty of inferring the evolutionary history of a trait from its current activity, it makes the practice of ascribing functions much more amenable to empirical testing. However, these theories appear to deprive functions of two of the properties that have, historically, been associated with their use and that continue to be associated with them. The first is the notion that they are explanatory in the sense that they specify an efficient cause for the current existence of the trait. What this means is not that the fact that a trait *had a function* in the past explains its current existence, but a trait’s *having a function* explains its current existence. The second is that they are normative. On the etiological view, the distinction between functioning properly, malfunctioning, and inability to function due to an abnormal environment is rendered tolerably clear: because of the fact that function is a historical concept, something can have a function without being able to perform it. It is controversial whether these distinctions can be drawn clearly within consequentialist theories, though it has been argued that consequentialist views can sustain normative interpretations of function (Wimsatt, 1972, p.47; Walsh, 1996, p.568; Schlosser, 1998, p.327).

Such considerations reinforce the value of adopting a pluralistic and context-dependent approach to analyzing “function.” In other words, in order to evaluate the meaning of a particular usage of “function” in a biological context, one must first identify the particular explanatory or pragmatic context in which that usage is embedded. If, for example, the ascription is intended to support an inference about how a trait evolved, or, perhaps, to make a normative claim about how the trait ought to behave, then an etiological concept of function may be implied. Alternatively, if the ascription is intended to sketch a prediction about the future survival value of a trait, or simply a prediction about what sort of behavior one ought to expect the trait to produce under well-defined circumstances, then a fitness-contribution theory, or an interest-based view, may be sufficient. What is crucial, then, is that different concepts of function allow one to articulate precisely the ontological and epistemological commitments that are implied by a given usage, and to ensure either that those commitments are satisfied

in that context, or that the conditions under which the function ascription would be warranted can be explicitly stated.

### Acknowledgment

The author wishes to express his gratitude to David J. Buller, Paul Griffiths, Anya Plutynski, Sahotra Sarkar, Gerhard Schlosser, and Arno Wouters for comments and criticism on earlier drafts of this article.

### References

- Adams, F. (1979). A goal-state theory of function attributions. *Canadian Journal of Philosophy*, 9, 493–518.
- Allen, C., & Bekoff, M. (1995a). Biological function, adaptation, and natural design. *Philosophy of Science*, 62, 609–22.
- Allen, C., & Bekoff, M. (1995b). Function, natural design, and animal behavior: philosophical and ethological considerations. In N. S. Thompson (Ed.). *Perspectives in ethology* (Vol. 11): *Behavioral design* (pp. 1–46). New York: Plenum Press.
- Amundson, R., & Lauder, G. V. (1994). Function without purpose: the uses of causal role function in evolutionary biology. *Biology and Philosophy*, 9, 443–69.
- Aquinas, S. Thomas (1914 [1269–73]). *The "Summa Theologica" of St. Thomas Aquinas. Part I. QQ. I–XXVI*. London: Burns Oates & Washbourne.
- Armbruster, W. S. (1997). Exaptations link evolution of plant–herbivore and plant–pollinator interactions: a phylogenetic inquiry. *Ecology*, 78, 1661–72.
- Ayala, F. J. (1968). Biology as an autonomous science. *American Scientist*, 56, 207–21.
- Ayala, F. J. (1970). Teleological explanations in evolutionary biology. *Philosophy of Science*, 37, 1–15.
- Bechtel, W. (1986). Teleological function analyses and the hierarchical organization of nature. In N. Rescher (Ed.). *Current issues in teleology* (pp. 26–48). Lanham, MD: University Press of America.
- Beckner, M. (1959). *The biological way of thought*. New York: Columbia University Press.
- Beckner, M. (1969). Function and teleology. *Journal of the History of Biology*, 2, 151–64.
- Bedau, M. (1990). Against mentalism in teleology. *American Philosophical Quarterly*, 27, 61–70.
- Bedau, M. (1991). Can biological teleology be naturalized? *Journal of Philosophy*, 88, 647–55.
- Bedau, M. (1992). Where's the good in teleology? *Philosophy and Phenomenological Research*, 52, 781–805.
- Bedau, M. (1993). Naturalism and teleology. In S. J. Wagner & R. Warner (Eds). *Naturalism: a critical appraisal* (pp. 23–51). Notre Dame, IN: University of Notre Dame.
- Bigelow, J., & Pargetter, R. (1987). Functions. *Journal of Philosophy*, 84, 181–96.
- Bock, W., & von Wahlert, G. (1965). Adaptation and the form–function complex. *Evolution*, 19, 269–99.
- Boorse, C. (1976). Wright on functions. *Philosophical Review*, 85, 70–86.
- Boorse, C. (2002). A rebuttal on functions. In A. Ariew, R. Cummins, & M. Perlman (Eds). *Functions: new essays in the philosophy of psychology and biology* (pp. 63–112). Oxford: Oxford University Press.

- Braithwaite, R. B. (1953). *Scientific explanation*. Cambridge: Cambridge University Press.
- Brandon, R. N. (1990). *Adaptation and environment*. Princeton: Princeton University Press.
- Broad, C. D. (1925). *The mind and its place in nature*. New York: Harcourt, Brace, & Company.
- Buller, D. J. (1998). Etiological theories of function: a geographical survey. *Biology and Philosophy*, 13, 505–27.
- Buller, D. J. (2002). Function and design revisited. In A. Ariew, R. Cummins, & M. Perlman (Eds). *Functions: new essays in the philosophy of psychology and biology* (pp. 222–43). Oxford: Oxford University Press.
- Canfield, J. (1964). Teleological explanations in biology. *British Journal for the Philosophy of Science*, 14, 285–95.
- Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.
- Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68, 53–74.
- Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, 72, 741–65.
- Cummins, R. (2002). Neo-teleology. In A. Ariew, R. Cummins, & M. Perlman (Eds). *Functions: new essays in the philosophy of psychology and biology* (pp. 157–72). Oxford: Oxford University Press.
- Davies, P. S. (2001). *Norms of nature: naturalism and the nature of functions*. Cambridge, MA: MIT Press.
- Dover, G. (2000). *Dear Mr Darwin: letters on the evolution of life and human nature*. Berkeley: University of California Press.
- Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.). *Belief: form, content, and function* (pp.17–36). Oxford: Clarendon Press.
- Enc, B. (1982). Intentional states of mechanical devices. *Mind*, 91, 161–82.
- Enc, B., & Adams, F. (1992). Functions and goal-directedness. *Philosophy of Science*, 59, 635–54.
- Faber, R. J. (1984). Feedback, selection, and function: a reductionistic account of goal-orientation. In R. S. Cohen & M. W. Wartofsky (Eds). *Methodology, metaphysics and the history of science* (pp. 43–135). Dordrecht: D. Reidel.
- Frankfurt, H. G., & Poole, B. (1966). Functional analyses in biology. *British Journal for the Philosophy of Science*, 17, 69–72.
- Godfrey-Smith, P. (1993). Functions: consensus without unity. *Pacific Philosophical Quarterly*, 74, 196–208.
- Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous*, 28, 344–62.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation: a missing term in the science of form. *Paleobiology*, 8, 4–15.
- Griffiths, P. E. (1992). Adaptive explanation and the concept of a vestige. In P. Griffiths (Ed.). *Trees of life: essays in philosophy of biology* (pp. 111–31). Dordrecht: Kluwer.
- Griffiths, P. E. (1993). Functional analysis and proper function. *British Journal for the Philosophy of Science*, 44, 409–22.
- Griffiths, P. E. (2005). Function, homology, and character individuation. *Philosophy of Science*, forthcoming.
- Hardcastle, V. G. (1999). Understanding functions: a pragmatic approach. In V. G. Hardcastle (Ed.). *Where biology meets psychology: philosophical essays* (pp. 27–43). Cambridge, MA: MIT Press.
- Harvey, W. (1894 [1628]). *An anatomical dissertation upon the movement of the heart and blood in animals: being a statement of the discovery of the circulation of the blood*. Canterbury: G. Moreton.
- Hempel, C. G. (1965 [1959]). The logic of functional analysis. In C. G. Hempel (Ed.). *Aspects of scientific explanation* (pp. 297–330). New York: Free Press.

Ho

Hu

Ka

Ki

Le

Le

Le

Lc

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

M

- Horan, B. (1989). Functional explanations in sociobiology. *Philosophy and Biology*, 4, 131–58.
- Hull, D. (1973). *Philosophy of biological science*. Englewood Cliffs, NJ: Prentice-Hall.
- Kant, I. (2000 [1789]). *Critique of the power of judgement*. Cambridge: Cambridge University Press.
- Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy*, 18, 379–97.
- Lehman, H. (1965). Functional explanation in biology. *Philosophy of Science*, 32, 1–20.
- Lennox, J. G. (1993). Darwin was a teleologist. *Biology and Philosophy*, 8, 409–21.
- Lewens, T. (2004). *Organisms and artifacts: design in nature and elsewhere*. Cambridge, MA: MIT Press.
- Lorenz, K. (1966 [1963]). *On aggression*. New York: Harcourt, Brace & World.
- Manier, E. (1971). Functionalism and the negative feedback model in biology. *Boston Studies in the Philosophy of Science*, 8, 225–40.
- Maynard-Smith, J. (2000). The concept of information in biology. *Philosophy of Science*, 67, 177–94.
- Mayr, E. (1961). Cause and effect in biology. *Science*, 134, 1501–6.
- Mayr, E. (1974). Teleological and teleonomic: a new analysis. *Boston Studies in the Philosophy of Science*, 14, 91–117.
- McLaughlin, P. (2001). *What functions explain: functional explanation and self-reproducing systems*. Cambridge: Cambridge University Press.
- Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Millikan, R. G. (1989a). An ambiguity in the notion “function”. *Biology and Philosophy*, 4, 172–6.
- Millikan, R. G. (1989b). In defense of proper functions. *Philosophy of Science*, 56, 288–302.
- Millikan, R. G. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.
- Millikan, R. G. (2002). Biofunctions: two paradigms. In A. Ariew, R. Cummins, & M. Perlman (Eds). *Functions: new essays in the philosophy of psychology and biology* (pp. 113–43). Oxford: Oxford University Press.
- Mitchell, S. D. (1993). Dispositions or etiologies? A comment on Bigelow and Pargetter. *Journal of Philosophy*, 90, 249–59.
- Mitchell, S. D. (1995). Function, fitness, and disposition. *Biology and Philosophy*, 10, 39–54.
- Nagel, E. (1953). Teleological explanation and teleological systems. In S. Ratner (Ed.). *Vision and Action* (pp. 537–58). New Brunswick, NJ: Rutgers University Press.
- Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace and World.
- Nagel, E. (1977). Teleology revisited: goal-directed processes in biology and functional explanation in biology. *Journal of Philosophy*, 74, 261–301.
- Neander, K. (1983). *Abnormal psychobiology*. PhD Diss., La Trobe.
- Neander, K. (1991). Functions as selected effects: the conceptual analyst's defense. *Philosophy of Science*, 58, 168–84.
- Nissen, L. (1980–1). Nagel's self-regulation analysis of teleology. *The Philosophical Forum*, 12, 128–38.
- Nissen, L. (1997). *Teleological language in the life sciences*. Lanham, MD: Rowman and Littlefield.
- Paley, W. (1839 [1802]). *Natural theology*. New York: Harper.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348, 27–33.
- Pittendrigh, C. S. (1958). Adaptation, natural selection, and behavior. In A. Roe & G. G. Simpson (Eds). *Behavior and evolution* (pp. 390–416). New Haven: Yale University Press.
- Plantinga, A. (1993). *Warrant and proper function*. Oxford: Oxford University Press.

JUSTIN GARSON

- Prior, E. W. (1985). What is wrong with etiological accounts of biological function? *Pacific Philosophical Quarterly*, 66, 310–28.
- Rea, M. (2002). *World without design: the ontological consequences of naturalism*. Oxford: Clarendon Press.
- Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science*, 10, 18–24.
- Ruse, M. (1971). Functional statements in biology. *Philosophy of Science*, 38, 87–95.
- Ruse, M. (1973). *The philosophy of biology*. London: Hutchinson University Library.
- Ruse, M. (1989). *The Darwinian paradigm: essays on its history, philosophy and religious implications*. London: Routledge.
- Sarkar, S. (2000). Information in genetics and developmental biology: comments on Maynard-Smith. *Philosophy of Science*, 67, 208–13.
- Sarkar, S. (2005). *Molecular models of life: philosophical papers on molecular biology*. Cambridge, MA: MIT Press.
- Schaffner, K. F. (1993). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.
- Scheffler, I. (1959). Thoughts on teleology. *British Journal for the Philosophy of Science*, 9, 265–84.
- Schlosser, G. (1998). Self-reproduction and functionality: a systems-theoretical approach to teleological explanation. *Synthese*, 116, 303–54.
- Schwartz, P. (2004). An alternative to conceptual analysis in the function debate. *The Monist*, 87, 136–53.
- Sherrington, C. S. (1906). *The integrative action of the nervous system*. New Haven: Yale University Press.
- Sommerhoff, G. (1950). *Analytical biology*. London: Oxford University Press.
- Sommerhoff, G. (1969). The abstract characteristics of living systems. In F. E. Emery (Ed.), *Systems thinking* (pp. 147–202). Harmondsworth: Penguin.
- Sorabji, R. (1964). Function. *Philosophical Quarterly*, 14, 289–302.
- Stampe, D. (1977). Towards a causal theory of linguistic representation. *Midwest Studies in Philosophy*, 2, 42–63.
- Sterelny, K. (2000). The “genetic program” program: a commentary on Maynard Smith on information in biology. *Philosophy of Science*, 67, 195–201.
- Tinbergen, N. (1963). On the aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–29.
- Walsh, D. M. (1996). Fitness and function. *British Journal for the Philosophy of Science*, 47, 553–74.
- Walsh, D. M., & Ariew, A. (1996). A taxonomy of functions. *Canadian Journal of Philosophy*, 26, 493–514.
- Williams, G. C. (1966). *Adaptation and natural selection: a critique of some current evolutionary thought*. Princeton: Princeton University Press.
- Wimsatt, W. C. (1971). Some problems with the concept of feedback. *Boston Studies in the Philosophy of Science*, 8, 241–56.
- Wimsatt, W. C. (1972). Teleology and the logical structure of function statements. *Studies in the History and Philosophy of Science*, 3, 1–80.
- Woodfield, A. (1976). *Teleology*. Cambridge: Cambridge University Press.
- Wouters, A. (2003). Four notions of biological function. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 34, 633–68.
- Wouters, A. (2005). The functional perspective in organismic biology. In T. A. C. Reydon & L. Hemerik (Eds). *Current themes in theoretical biology* (pp. 33–69). Dordrecht: Springer.

Wt  
Wi  
Wi

Al  
Al  
Bl  
W



- Wright, L. (1972). A comment on Ruse's analysis of function statements. *Philosophy of Science*, 39, 512–14.
- Wright, L. (1973). Functions. *Philosophical Review*, 82, 139–68.
- Wright, L. (1976). *Teleological explanations: an etiological analysis of goals and functions*. Berkeley: University of California Press.

### Further Reading

- Allen, C., Bekoff, M., & Lauder, G. (Eds). (1998). *Nature's purposes*. Cambridge, MA: MIT Press.
- Ariew, A., Cummins, R., & Perlman, M. (2002). *Functions: new essays in the philosophy of psychology and biology*. Oxford: Oxford University Press.
- Buller, D. J. (Ed.). (1999). *Function, selection, and design*. Albany, NY: SUNY Press.
- Wouters, A. (2005). The function debate in philosophy. *Acta Biotheoretica*, 53, 123–51.