A MODAL THEORY OF FUNCTION*

The function of a trait token is usually defined in terms of some properties of other (past, present, or future) tokens of the same trait type. I argue that this strategy is problematic, as trait types are (at least partly) individuated by their functional properties, which would lead to circularity. In order to avoid this problem, I suggest a way to define the function of a trait token in terms of the properties of the very same trait token. To able to allow for the possibility of malfunctioning, some of these properties need to be modal ones: a function of a trait is to do F just in case its doing F would contribute to the inclusive fitness of the organism whose trait it is. Function attributions have modal force. Finally, I explore whether and how this theory of biological function could be modified to cover artifact function.

I. ARTIFACT FUNCTION AND BIOLOGICAL FUNCTION

The function of my corkserew is to open bottles. The function of my heart is to pump blood. These two function-attributions are of different kinds. My corkscrew is an artifact, whereas my heart is a biological organ. Artifact function seems to be the easier of the two kinds to analyze. The standard way of explaining artifact function is with reference to the notion of design. My corkscrew has the function to open wine bottles if and only if it was designed to open wine bottles. If we wonder what the function of an artifact may be, we should just ask the designer to get an answer.

This explanatory scheme will not work in the case of biological functions, as there is no designer who designed biological traits (or, in any case, there is no one we could ask). Thus, it seems that in spite of the fact that we talk about functions both in the artifact and in the biological case, these two kinds of function are very different indeed: one is fixed by design, whereas the other is not.

I focus on biological function in this paper, but at the end I will come back to the notion of artifact function and reevaluate the

standard analysis of artifact function in light of the argument pre-
sented in the case of biological function.

After considering some important desiderata every theory of func-
tion needs to satisfy (section II), I point out that the function of a trait
token is usually defined in terms of some properties of other (past,
present, future) tokens of the same trait type. I argue that this strategy
is problematic, as trait types are usually individuated (at least partly)
in terms of their functional properties, which would lead to circularity
(sections III–V). In order to avoid this problem, I suggest a way to
define the function of a trait token in terms of the modal properties
of the very same trait token (sections VI–VII). Finally, I explore whether
and how this theory of biological function could be modified to cover
artifact function (section VIII).

II. THREE DESIDERATA FOR A THEORY OF FUNCTION

There may be many more than three desiderata for a theory of bio-
logical (or artifact) function, but I will mention three of these, which
I take to be the most important ones and which apply in both the
biological and the artifact cases.

First, a trait can have two (or more) functions at a time. The func-
tion of my mouth is both to eat and to speak, for example. A theory of
function should be able to allow for the possibility that one trait has
two (or more) functions.

Second, function attributions can depend on the explanatory project
at hand. The function of my left eyelid is to blink, but its function is
also to keep my left eye moist. It depends on the explanatory project
at hand which function attribution we will opt for. Suppose that we are
concerned with the anatomy of the eyelid, regardless of its relation to
the eye. In this explanatory project, it will be irrelevant whether the
eye is kept moist or not: the function of the eyelid is to contract and
expand in a certain way: to blink. In some other explanatory projects,
however, where we analyze the moistness of the eye and we are not
concerned with the anatomy of the eyelid, the function of the eyelid
in this explanatory scheme will be to keep the eye moist. It should
not be a surprising claim that function attributions can depend on
the explanatory project. It has been argued that the function of a trait
explains why this trait is the way it is.[1] If, however, explanations are
considered to be responses to why-questions[2] and, therefore, themselves

---

[1] L. Wright, "Functions," *Philosophical Review*, LXXXII, 2 (1973): 139–68; Paul E.
Griffiths, "Functional Analysis and Proper Functions," *British Journal for the Philosophy
of Science*, XLIV, 3 (1993): 409–22.

[2] Bas C. van Fraassen, *The Scientific Image* (New York: Oxford, 1980).

depend on the explanatory project, then the function we attribute to a trait will also depend on the explanatory project.[3]

Third, any theory of function must be able to account for the phenomenon of malfunctioning. A trait can have a function but fail to perform this function. If my heart skips a beat, it still has the function to pump blood, but at that moment it fails to perform this function: it malfunctions.

III. THE ETIOLOGICAL THEORY OF BIOLOGICAL FUNCTION

The most widespread notion of biological function is the following: a trait of an organism has function F if and only if its performing F has contributed to the survival of the ancestors of this organism. This notion of function is usually referred to as 'etiological': what determines the function of a trait is its history. The function of the human heart is to pump blood because the fact that the heart pumped blood contributed to the survival of our ancestors.[4]

According to a widely accepted version of the etiological theory, the "modern history theory of functions," in order for a trait to have a function it must be the case that this trait has *recently* contributed to the survival of the organism's ancestors.[5] If a trait contributed to the survival of an organism's ancestors in the distant past but has not contributed since, it does not have a function. The human appendix, for example, has not contributed to our survival recently; thus, it does not have any function. To sum up, the etiological view of function asserts that the function of a trait is determined by its *recent* history.

Note that this theory of biological function restores the continuity between the explanation of biological and artifact functions. The function of my corkscrew is to open wine bottles because it has been designed to open wine bottles, whereas the function of my heart is to

---

[3] This feature of function attributions also explains why biological function is sometimes thought to be indeterminate (see, for example Peter Godfrey-Smith, "A Modern History Theory of Functions," *Noûs*, XXVIII, 3 (1994): 344–62, at p. 356). In the distant past, the human appendix had the function to decompose celluloid. Now it no longer has this function. At some point in our evolutionary history, the human appendix ceased to have this function. But it is difficult to see what could be the criterion for the exact time when it no longer had this function. If function attribution depends on the explanatory project, then, depending on with which explanatory project we are engaging, the human appendix may or may not have the function to decompose celluloid.

[4] Ruth G. Millikan, *Language, Thought and Other Biological Categories* (Cambridge: MIT, 1984); Karen Neander, "Functions as Selected Effects: The Conceptual Analyst's Defense," *Philosophy of Science*, LVIII, 2 (1991): 168–84; Neander, "The Teleological Notion of 'Function'," *Australasian Journal of Philosophy*, LXIX, 4 (1991): 454–68; Griffiths, *op. cit.*; Godfrey-Smith, *op. cit.*; Wright, *op. cit.*

[5] Godfrey-Smith, *op. cit.*

pump blood because it has been selected for pumping blood. In both cases, function is fixed by the past: past design or past selection.

Probably the most famous objection to the etiological view is based on the swampman thought experiment. A very direct consequence of the etiological definition of function is that what fixes the function of a trait is its past, not its present. Hence, if an organism that is molecule-for-molecule identical to me (the swampman) were created by chance, its organs would not have any functions, since it would lack the evolutionary history that would fix the function of these organs. Without engaging with the Byzantine swampman literature, I raise a more serious objection to the etiological theory in the next section and then generalize this objection to other theories of function in section v.

IV. A NEW OBJECTION: THE INDIVIDUATION OF TRAIT TYPES

The etiological definition of function presupposes that *trait types* can be individuated in an unproblematic manner. The trait whose function is to be defined and the traits that have been selected for in the past must be *of the same type*. But how can we individuate trait types? What makes hearts different from nonhearts?

I will argue that there is no coherent, noncircular way of individuating trait types that is available to the etiological theory of function.

The question, then, is how trait types are individuated. I will consider three options and point out that none of them is available to the etiological theory of function.

IV.A. Functional Criteria. The most widely accepted account of trait-type individuation holds that tokens of a certain trait type all have the same function. A token object belongs to trait type *T* if and only if it has certain functional properties: if it has the function to do *F.* Those entities are hearts that have the function of pumping blood. Those entities that do not have this function are not hearts.

As Karen Neander puts it: "Most biological categories are only definable in functional terms."[6] This account of individuating trait types is widely (though not universally) accepted as a general suggestion both in philosophy of biology and in philosophy in general. Tyler Burge, for example, writes that "to be a heart, an entity has to have the normal evolved function of pumping blood in a body's circulatory system."[7]

---

[6] Neander, "Functions as Selected Effects," *op. cit.*, p. 180; see also Morton Beckner, *The Biological Way of Thought* (New York: Columbia, 1959), p. 112, and Tim Lewens, *Organisms and Artifacts: Design in Nature and Elsewhere* (Cambridge: MIT, 2004), p. 99.

[7] Tyler Burge, "Individuation and Causation in Psychology," *Pacific Philosophical Quarterly*, DCCVII, 4 (1989): 303–22.

It is important to note, however, that the etiological theory of function cannot help itself to this way of individuating trait types when defining function without running into circularity.

As we have seen, the etiological definition of function presupposes an account of trait-type individuation. Now, if we want to avoid circularity, we cannot use the notion of function in order to explain trait-type individuation. When we are explaining function, the claim that x^* (the trait whose function we are explaining) is a token of type X (the traits that have been selected in the past) is part of the *explanans*. Hence, we cannot use the *explanandum* (function) to explain part of the *explanans* (why x^* is a token of type X).[8]

Thus, if we want to talk about trait types in the definition of function, we need some other way of individuating them.

IV.B. Morphological Criteria. A simpler suggestion is that we can use morphological criteria for individuating trait types. The proposal is that a token object belongs to trait type T if and only if it has certain morphological properties. An entity is a heart if, for example, it has a certain shape, size, and color, and it is not a heart otherwise.

One problem with this suggestion is that trait types need to range over different species, but the hearts of different species have very different morphological properties.[9]

But even if we only want to individuate a trait type within a certain species, we still cannot use morphological criteria. A malformed heart that does not have the morphological properties hearts have is a heart all the same (it is a malformed *heart*, after all). What keeps hearts and nonhearts apart cannot be a set of morphological properties.

A possible suggestion would be to say that hearts are those entities that play a certain causal role, and those entities that do not play this causal role are not hearts.[10] This proposal, however, would not work in the malformed heart case: a malformed heart does not have the causal role hearts have, but it is a heart nonetheless. What

---

[8] This problem is acknowledged by some of the defenders of the etiological theory of function. See Neander, "Types of Traits: Function, Structure, and Homology in the Classification of Traits," in Andre Ariew, Robert Cummins, and Mark Perlman, eds., *Functions: New Essays in the Philosophy of Biology and Psychology* (New York: Oxford, 2002), pp. 402–22, especially p. 403; see also Griffiths, *op. cit.*, and Paul Sheldon Davies, *Norms of Nature: Naturalism and the Nature of Functions* (Cambridge: MIT, 2001).

[9] Neander, "Functions as Selected Effects," at p. 180.

[10] I discuss this proposal here because its most natural rendering falls under the morphological account of trait-type individuation, but it is worth noting that it could also be interpreted as a version of the functional account—if we conceive of function in a way Cummins does in his "Functional Analysis," this JOURNAL, LXXII, 11 (November 1975): 741–64, and in his "Neo-Teleology," in Ariew, Cummins, and Perlman, eds., *op. cit.*, pp. 157–73.

matters is not what the heart does (or how it looks), but what it is supposed to do.

To sum up, the suggestion that morphological criteria could be found for individuating trait types does not work.[11] This leaves us with a third alternative.

IV.C. Homological Criteria. A third possible answer is the following. One could argue that what guarantees that two traits are tokens of the same type is that they are homologues: they have common descent; they are members of the same "reproductively established family."[12] To make this suggestion as plausible as possible, we should not confuse it with the view that homologous traits are 'coded by' the same gene, for the simple reason that no trait should be taken to be 'coded by' a gene. The way a trait turns out depends partly on the gene, but it also depends on the intra- and extra-cellular environment during the developmental process. Assuming that the genetic factor can be singled out from this complex causal network is biologically very implausible.[13] The most plausible homological accounts of trait-type individuation are not committed to this gene-centric view.[14] The suggestion is that whether a trait belongs to a homologically established trait type depends not only on what gene this trait was coded by, but also on the developmental process.

The homological account of trait-type individuation is vulnerable to a serious objection. Take the following example. The forelimbs of vertebrates, such as the wings of birds, and the forelegs of ancient amphibians are homologous: the wings of eagles belong to the same reproductively established family as the forelegs of some ancient amphibians. According to the suggestion for individuating trait types we are considering here, they must be of the same trait type.

But the wing of the eagle and the foreleg of an ancient amphibian are clearly not tokens of the same trait type. One of them is a wing, and the other is a foreleg. They belong to very different trait types indeed.

[11] The morphological account of trait-type individuation may be supplemented with a homological one (Ron Amundson and George V. Lauder, "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology," *Biology and Philosophy*, IX (1994): 443–69; Lauder, "Homology, Form, and Function," in Brian K. Hall, ed., *Homology: The Hierarchical Basis of Comparative Biology* (San Diego: Academic Press, 1994), pp. 151–96). As we will see, however, this homological account raises serious worries.

[12] See for example Amundson and Lauder, "Function without Purpose," *op. cit.*; Lewens, *op. cit.*, pp. 99–100; Millikan, *op. cit.* The term "reproductively established families" was introduced by Millikan (*ibid.*, p. 23ff).

[13] Griffiths and R. D. Gray, "Developmental Systems and Evolutionary Explanation," this JOURNAL, XCI, 6 (June 1994): 277–304, especially pp. 298ff.

[14] V. Louise Roth, "On Homology," *Biological Journal of the Linnean Society*, XXII (1984): 13–39, especially p. 17; Günther P. Wagner, "Homology and the Mechanisms of Development," in Hall, ed., *op. cit.*, pp. 273–99.

The defender of the homological way of individuating trait types may argue that these two traits do belong to the same trait type; both are forelimbs after all. A trait token can be typed in many different ways, and typing the eagle's wings as forelimbs is a perfectly valid way of doing so. The real problem is that this way of individuating trait types cannot individuate trait types narrowly enough: it will not even be able to differentiate between wings and forelegs. More importantly, such a broad way of typing traits does not help in our definition of function, as using this way of talking about trait types would attribute the function of crawling to the eagle's wings.

But perhaps such old members of a reproductively established family (for example, the traits of our ancient amphibian) just do not count. A possible suggestion would be to say that wings belong to the same trait type because they are all *recent* members of a reproductively established family. In other words, two token objects belong to the same trait type if and only if they are *recent* homologues: *recent* members of a reproductively established family.

The problem with this suggestion is that there is no noncircular way of cashing out what is meant by 'recent'. We would be happy to say that the eyes of the eagle and the eyes of the ancient amphibian are tokens of the same type. Then why can't we do the same with forelimbs? What is so different in the two cases that makes us sort the two token traits under the same type in the latter case but not in the former?

The only thing that differentiates the example of the eye from the example of the forelimb is that the selection pressure changed in the latter case but not in the former. Forelimbs have been selected for doing something *different* in the bird population and in the ancient amphibian population. Eyes, on the other hand, have been selected for doing *the same thing* in the bird population and in the ancient amphibian population.

Thus, if we want to make sense of the suggestion that two trait tokens belong to the same trait type if and only if they are *recent* homologues, then we will have difficulties defining what is meant by the term 'recent'. In defining the eagle's trait types, 'recent' includes the ancient amphibian population when we are analyzing the eye example, but it certainly does not include the ancient amphibian population when it comes to forelimbs. The bottom line is that what 'recent' amounts to depends on what the trait in question has been selected for; what 'recent' amounts to depends on the etiological function of the trait. The suggestion boils down to the claim that what makes a trait token a token of a certain trait type is that it is a homologue of trait tokens that were selected for doing the same thing as this token (or, in other words, that had the same etiological function as this token).

Hence, this way of individuating trait types collapses into the functional account of trait-type individuation. But we have seen above that the functional account of trait-type individuation cannot be used in the definition of function without running into circularity.

To sum up, the etiological theory of function cannot rely on any of the three ways of individuating trait types. Since the etiological notion of function requires an unproblematic way of individuating trait types, we need to dispose of this theory of function.[15]

V. BEYOND ETIOLOGY: A MORE GENERAL PROBLEM

If the argument I presented here is correct, then the etiological theory of function cannot stand, for it has to rely on an independent account of individuating trait types, and no such account is available for the etiological theory. So, the etiological theory should be disposed of, and we should look for some other theory of function. The problem is that *all* alternatives to the etiological theory of function rely on an independent account of individuating trait types.

The main alternative to the etiological theory of function has been the so-called propensity theory, which claims that it is not the past but the future of the organism that fixes the function of a trait.[16] The function of a trait is what *will (be likely to) contribute* to the survival of the organism. In other words, function is a propensity.

According to the propensity definition, a trait "has a (biological) function just when it confers a survival enhancing propensity on a creature that possesses it."[17] In other words, the function of a trait is its propensity to contribute to the fitness of the organism: "when we speak of the function of a character, therefore, we mean that the character generates propensities that are survival-enhancing in the creature's natural habitat."[18]

Several objections have been raised against this view.[19] Whether or not these objections are conclusive, it needs to be noted that the propensity definition of function also presupposes an unproblematic

---

[15] One could argue at this point that these three ways of individuating trait types are not exhaustive. More specifically, one could argue for some kind of hybrid account (see for example Neander, "Functions as Selected Effects," at p. 178, and "Types of Traits," especially pp. 403–04). It can be pointed out that these hybrid accounts of trait-type individuation would either collapse into a homological account or raise the same problems about circularity as the functional account of trait-type individuation.

[16] John Bigelow and Robert Pargetter, "Functions," this JOURNAL, LXXXIV, 4 (April 1987): 181–96.

[17] *Ibid.*, p. 192.

[18] *Ibid.*

[19] Godfrey-Smith, *op. cit.*, especially pp. 352–53; Neander, "The Teleological Notion of 'Function'"; Millikan, *White Queen Psychology and Other Tales for Alice* (Cambridge: MIT, 1993); Denis M. Walsh, "Fitness and Function," *British Journal for the Philosophy of Science*, XLVII, 4 (1996): 553–74.

account of how trait types are individuated: a trait token has the function to do *F* if and only if the fact that traits *of the same type* will do *F* will contribute to the organism's survival.[20]

There is a third theory of function that we should consider: the relational theory of function.[21] It has been argued that we cannot talk about the function of a trait in general, but only the function of a trait relative to a certain selective regime.[22] Here is Walsh's definition: "The/a function of a token of type *X* with respect to selective regime *R* is to *m* iff *X*'s doing *m* positively (and significantly) contributes to the average fitness of individuals possessing *X* with respect to *R*."[23] Again, this definition presupposes an independent way of individuating trait types.

In other words, the most important candidates for defining function presuppose an account of individuating trait types. Thus, we have a serious worry. We are left with no plausible theory of function.

VI. A MODAL THEORY OF FUNCTION

We have seen that every definition of function that talks about trait types faces the trait-type individuation objection. An obvious way to avoid this objection would be to define function without referring to trait types at all. If we accept a notion of function that does not rely on the prior individuation of trait types, then we obviously do not need to worry about trait-type individuation. If we could define function without appealing to trait-type individuation, then we could use this definition of function to individuate trait types without running into circularity.

Note, however, that if a definition of function does not rely on an account of trait-type individuation, then the function of a token trait must be determined entirely by the properties of *that very trait token* and not by the properties of other tokens of the trait type to which this token belongs. In that case, however, it is difficult to see how a trait can malfunction. When a trait malfunctions, it is supposed to do (that is, it has the function to do) *F*, but it does not do *F*. My heart malfunctions when it does not pump blood (though it is supposed to/has the function to do so). If we define the function of a trait

---

[20] As Godfrey-Smith pointed out, the propensity theory oscillates between talking about the propensity of a trait token (see Bigelow and Pargetter, *op. cit.*, p. 192) and taking about the propensity of a trait type (*ibid.*, pp. 194–95). He convincingly argues that the only plausible reading is the latter (Godfrey-Smith, *op. cit.*, at p. 360).

[21] Walsh, "Fitness and Function," *op. cit.*

[22] By selective regime Walsh means "the total set of abiological and biological (including social, developmental and physiological) factors in the environment of the trait which potentially affect the fitness of individuals with that trait" (*ibid.*, p. 564).

[23] *Ibid.*

token in terms of the properties of that trait token alone, then it is difficult to see how the function can be different from what the trait token actually does. In other words, it is difficult to see how such an account of function could explain malfunctioning.

One possible way to explain how a trait can malfunction is by attributing modal force to claims about function. Trait x may not perform F, but if it were to perform F, this would contribute to the survival of the organism with x. Thus, at first approximation, doing F is a function of x if and only if it is true that if x is doing F, then this *would* contribute to the survival of the organism with x.

Thus, the suggestion is that the tense of 'contribute' in the definition of function is not past tense as in the etiological account. It is not future tense either—this would be the suggestion of the propensity theory. And, finally, it is not present tense, which would be the way the relational theory defines it. According to my account of function, instead of 'contributed', 'will contribute', or 'contributes', we have to use 'would contribute'. Function attributions have modal force.[24]

Some further clarifications are needed about this general suggestion. First, the talk about contribution to the survival of an organism, which is a standard way of analyzing function, is vague. What really matters in natural selection is not the survival but the inclusive fitness of the organism. Further, if a trait's doing F contributes to the survival of an organism, the trait is doing F at time t, but what it contributes to, that is, the organism's survival, is at some other time, t^*. But many things can happen between t and t^*. Presumably some kind of appeal to *ceteris paribus* clauses could go around this problem, but to keep things simple, instead of talking about contribution to the survival of an organism, I will talk about contribution to the inclusive fitness of an organism. This will also allow me to define the function of a trait at time t in terms of some (modal) facts at time t.

Second, I define function with the help of a counterfactual. Any theory of counterfactuals could be used to fill in the details of this definition, but, for simplicity, I will use Lewis's theory.[25] Using a Lewisian account of counterfactuals, my definition of function would amount to the following. Performing F is a function of x if and only if some possible worlds where x is doing F and this contributes to the survival of organism O are closer to the actual world than any of those possible worlds where x is doing F but this does not contribute to O's survival.

---

[24] It has to be noted that at least some versions of the etiological notion of function could also be interpreted as carrying modal force—whether they do depends on how we interpret the concept of 'contribution' in the definition of function.

[25] David Lewis, *Counterfactuals* (Cambridge: Harvard, 1973).

Now, some of these possible worlds may be frightfully distant. It would contribute to any organisms' survival if their scratching their ear killed off any approaching predators. Still, this is not a function of scratching one's ear, because those possible worlds where scratching one's ear can kill off predators are very far away—the project of explaining the function of scratching one's ear should not take into consideration such distant possible worlds.

Thus, the set of possible worlds that we are considering when determining whether the counterfactual that defines function is true or not should be restricted to 'relatively close' possible worlds.

Performing F *is a function of organism* O's *trait* x *at time* t *if and only if some* '*relatively close*' *possible worlds where* x *is doing* F *at* t *and this contributes to* O's *inclusive fitness are closer to the actual world than any of those possible worlds where* x *is doing* F *at* t *but this does not contribute to* O's *inclusive fitness.*[26]

It is important to note that I only intended to define the function of a token trait. Sometimes we talk about the function of trait types: the function of hearts is to pump blood. I will not give a definition for the function of trait types, as this definition would depend on how we individuate trait types, which, as we have seen, is a very difficult question.[27]

If x is not doing (or even cannot do) F in the actual world, but in a 'relatively close' possible world it is doing F and its doing F contributes to the organism's inclusive fitness, then we can still attribute function F to x. This is exactly what happens if a trait is malfunctioning: if it fails to perform its function.

A *prima facie* worry about this modal account of function is that it proliferates functions: there is no limit to the various potential functions $F_1, F_2, ...F_n$ that are such that, if x were to do F_i, doing so would contribute to O's inclusive fitness. Note, however, that according to my definition, function attribution implies that some 'relatively close' possible worlds where x is doing F and this contributes to O's inclusive

---

[26] Note that this way of defining function individuates function quite narrowly. Grasping in general is not a function of my hand, because it may be the case that the closest possible world where my hand is grasping is one where its doing so does not contribute to my survival. If we want to attribute a function to my hand, we have to give a much more specific characterization: grasping food when I am hungry, for example. My hand, of course, has many functions: grasping food when I am hungry is only one of these. The fact that my definition gives a very specific characterization of the functions of a trait (and not a general specification, like grasping) is an explanatory advantage of my account.

[27] I also have misgivings about the overuse of trait types in evolutionary biology in general. See Bence Nanay, "Population Thinking as Trope Nominalism," *Synthese*, forthcoming.

fitness are closer to the actual world than any of those possible worlds where x is doing F but this does not contribute to O's inclusive fitness. In other words, if we find a 'relatively close' possible world where x is doing F and this contributes to O's inclusive fitness, this by no means guarantees that performing F is a function of x. What is also needed is that some of those worlds where x is doing F and this contributes to O's inclusive fitness are closer to the actual world than any of those possible worlds where x is doing F but this does not contribute to O's inclusive fitness.

Take the following example.[28] My left foot could serve as a paddle for swimming fast. This might improve my inclusive fitness—for example, if it made me sexually attractive (or famous). But if so, does this make paddling a function of my left foot according to the modal theory?

The answer is the following. If there is a 'relatively close' possible world where the F-ing of my left foot contributes to my inclusive fitness, this does not make F-ing a function of my left foot. What is required for my left foot to have a function F is that those worlds where it does F and this contributes to my inclusive fitness are closer to the actual world than the ones where it does F without contributing to my inclusive fitness. It is not enough for function attribution to find a possible world where doing F contributes to my inclusive fitness. We need to compare this world to those where it does not. By these standards, paddling would be disqualified from the elite circle of the functions of my left foot.

More slowly: the modal definition of function was this: some 'relatively close' possible worlds where x is doing F and this contributes to O's inclusive fitness are closer to the actual world than any of those possible worlds where x is doing F but this does not contribute to O's inclusive fitness. Let us apply this to paddling: some possible worlds where my left foot serves as a paddle for swimming fast and this contributes to my inclusive fitness are closer than any of those possible worlds where it serves as a paddle for swimming fast but this does not contribute to my inclusive fitness. However, there are lots of very nearby possible worlds where my left foot serves as a paddle for swimming fast without this making any difference to my inclusive fitness. In fact, most nearby possible worlds are extremely likely to be such worlds. There are some possible worlds, of course, where it does (say, where I am an Olympic swimmer). But, at least in my case,

---

it is extremely unlikely that these worlds would be closer to the actual world than the boring possible worlds where paddling does nothing for my inclusive fitness.

There may be some people for whom there are possible worlds where their feet serve as paddles for swimming fast and this does contribute to their inclusive fitness. Michael Phelps may be one of them. When we apply the modal theory in the case of his feet, we may have to attribute the function of paddling to them, at least in some explanatory projects. But I do not think that we should find this surprising. After the 2008 Summer Olympics, the media was full of commentaries about why he wins all the races, and the commentators literally talked about how his various body parts function to help him swim faster (the function of his palms, the function of his relatively short legs, and so on).

A last worry about paddling: how is it possible that it is not a function of my feet to paddle, but it is a function of Phelps's feet to paddle? This is an important worry because it highlights a crucial aspect of the modal theory of function. The modal theory of function attributes function to trait tokens, not trait types. The function(s) of a token trait is (are) defined in terms of modal facts about this very token trait. Hence, there is no guarantee that my feet and your feet will have the same functions.

Here is another important question about the modal theory of function. There are famous disputed cases of function attribution. Does the modal theory help us to resolve these? One such case is this. When there is a conflict between two male baboons, sometimes one of them picks up an infant (this phenomenon was first observed among Barbary macaques). What is the function of this behavior? There are (at least) two candidates. The first is that the baboon who picks up the infant is using the baby to protect himself from the other male, who does not want to risk hurting the baby because if he does the female baboons will start attacking him. This is called the 'agonistic buffering hypothesis'.[29] The alternative is that the function of this behavior is parental care. It has been observed that infants are most often picked up by a long-term resident male baboon, while the male he is having the conflict with is usually a recent immigrant. In other words, the holder could possibly be a father of the infant, and he is protecting it from the other, possibly infanticidal male.[30]

[29] J. M. Deag and J. H. Crook, "Social Behaviour and 'Agonistic Buffering' in the wild Barbary Macaque *Macaca sylvana* L.," *Folia Primatologica*, xv, 3/4 (1971): 183–200.
[30] Curt Busse and William J. Hamilton III, "Infant Carrying by Male Chacma Baboons," *Science*, ccxii, 4500 (June 12, 1981): 1281–83.

How should we decide whether the function of this behavior is self-defense or parental care? The first thing to notice is that according to the modal theory of function the behavior of picking up the infant has both functions: doing both would contribute to the organism's inclusive fitness. Thus, we should not ask which one is the function of this behavior: both are. Rather, the question should be framed in terms of which one of the two is the more relevant/important function of this behavior. And in order to answer this question we need to examine, unsurprisingly, the modal facts. Suppose that a male baboon A is picking up his son C during his fight with B. In order to decide what the (primary, most relevant) function of this behavior is, we need to consider the possible world where A is picking up D (not C), who is not his son. Does this behavior contribute to A's inclusive fitness in this possible world? If if does, then we have reason to believe that the primary function of this behavior is self-defense. But if it does not, then parental care seems to be a more relevant function. This example could be elaborated further, but this sketchy treatment should be enough to underline the importance of modal facts in resolving problematic cases of function attribution.[31]

I left open what counts as a 'relatively close' possible world in the above definition. The short answer is that what counts as a 'relatively close possible world' depends on the explanatory project. As we have seen, function attribution can depend on the explanatory project. One way of explaining this would be to say that different explanatory projects focus on different sets of possible worlds where x could be doing F. It needs to be spelled out, however, what this dependence on explanatory projects really means.

It depends on the explanatory project how we should analyze the function of my eyes in an environment where it is pitch dark. There is a possible world where everything is the same as in this one except that it is not pitch dark. If we count this possible world as 'relatively close', then my eye does have a function. If we are analyzing the function of my eye in my bedroom with the lights off, then it seems to be a good idea to include the possible world where it is not pitch dark. If we are analyzing a scenario where photons suddenly disappeared from the universe, then we probably should not include the possible world where it is not pitch dark.

In some explanatory projects, it is irrelevant what x would or could do if it had different intrinsic properties—we are interested in the

---

[31] Primatologists, of course, cannot do fieldwork in possible worlds. But they can infer the function of A's token behavior from observing other, similar instances of this behavior (taking for granted some unproblematic way of typing behavior).

function of x as it is. In these cases, the set of 'relatively close possible worlds' would amount to the set of possible worlds where the intrinsic properties of x are the same as in the actual one. Other things about these possible worlds and, most importantly, the environment x is in, can vary. A possible example for such explanatory projects would be to find the function of a seemingly functionless trait by extrapolating environments where this trait does contribute to the organism's inclusive fitness.

In some other explanatory projects, what x does or can do in environments different from the present one is irrelevant. In these cases, the set of 'relatively close possible worlds' means the set of possible worlds where the enviroument is the same as in the actual world. The function of x, then, is relative to the environment. Strictly speaking, in such explanatory projects we should talk about the function of x in environment E—just as the relational theory of function does. Examples where the same trait can do very different things that would contribute to the inclusive fitness of the organism in different environments are possible examples for explanatory projects of this kind.[32]

VII. OBJECTIONS

We need to make sure that the modal theory of function satisfies the three desiderata I enumerated in section II.

Any theory of function needs to be able to explain malfunctioning. As we have seen, a trait malfunctions if and only if it has a function but fails to perform it. This is perfectly possible in my account, since even if x is not doing F in the actual world, it may still be true that if x were performing F then this would contribute to the inclusive fitness of the organism that possesses x.

The other two desiderata are also satisfied. A trait can have two or more functions, as there may be many things the trait does that

---

[32] As we have seen, the relational theory of function talks about function relative to a selective regime, that is, to "the total set of abiological and biological (including social, developmental and physiological) factors in the environment of the trait which potentially affect the fitness of individuals with that trait" (Walsh, *op. cit.*, p. 564). Relativizing function to a selective regime may be thought of as being the same as relativizing it to an environment. The relational view is nevertheless different from this special case of my definition of function in two very important respects. First, as we have seen, the relational theory of function defines function in terms of the contribution of *trait types*, whereas my definition does not talk about trait types and defines function entirely in terms of the properties of the token trait. Second, the relational notion of function does not carry any modal force (see especially *ibid.*, section v.1), whereas my notion does.

would contribute to the organism's inclusive fitness. And, as we have seen, the attribution of functions depends on the explanatory project, since the explanatory project determines which nearby possible worlds we should take into consideration when assessing the function of a trait.

Let us see how this proposal can deal with the cases that are problematic for the etiological approach. If the swampman's heart pumped blood then this would contribute to the inclusive fitness of the swampman (this follows from the supposition that the swampman is molecule-by-molecule identical to a human being); hence, the swampman's heart has the function to pump blood, in spite of the fact that he lacks history.

My notion of function is obviously not vulnerable to the trait-type individuation objection, because it does not use trait types when defining function. It defines the function of a token trait entirely in terms of the properties of this token trait. To sum up, if we conceive of function the way I suggested, some of the worrying consequences of the etiological view disappear.

Finally, one could argue that this new theory of function is susceptible to new objections. More precisely, one may worry that this definition does not capture the notion of function, but rather the notion of usefulness.

My response is to bite the bullet: function may have a lot to do with usefulness. But it is important to distinguish usefulness from use. It would indeed be a worrying consequence of my view if it ended up assimilating function to use: to whatever the trait is being used for. But this is not the case. What a trait is being used for is determined by what goes on in the actual world. Function (and, arguably, usefulness), in contrast, depends on what goes on in nearby possible worlds. Function is a modal concept; use is not. As long as we clarify that usefulness is not the same as use and that it should be conceived of as a modal concept, it may not be such a bad idea to claim that function has a lot to do with usefulness.

The main consideration against thinking about function as usefulness is that the notion of function is generally taken to be tied to the notion of design, which is very different from usefulness.[33] As Philip

---

[33] It has been argued recently that if we conceive of function as usefulness we may avoid some undesirable consequences of conceiving of function as design (Wayne D. Christensen and Mark Bickhard, "The Process Dynamics of Normative Function," *Monist*, LXXXV, 1 (January 2002): 3–28; Richard Cameron, "How to Be a Realist about *Sui Generis* Teleology Yet Feel at Home in the 21st Century," *Monist*, LXXXVII, 1 (January 2004): 72–95).

Kitcher put it, "the function of S is what S is designed to do."[34] This seems to be a very widely accepted view.[35]

The main motivation for interpreting function as design comes from the artifact case: in the case of artifact function, design fixes function, so, if we want to maintain the continuity between biological and artifact function, we should expect something very similar in the case of biological function. If we manage to point out that even in the artifact case function has little to do with design, then the main motivation for this objection ceases to exist. This is exactly what I intend to do in the next section.

VIII. BACK TO ARTIFACT FUNCTIONS

I outlined a theory of biological function. However, if this theory is correct, then the explanation of biological function is very different from the explanation of artifact functions. Artifact function is fixed by design, whereas biological function is fixed by modal facts. Some would see this as a weakness of my account. One of the attractions of the etiological theory of function was that it could provide a theory of biological function that is continuous with the way we usually explain artifact function.[36]

My response is to say that instead of constructing a theory of biological function that would mirror the standard way of thinking about artifact function, we should reevaluate the standard understanding of artifact function. In short, my claim is that artifact function is not, or at least not always, fixed by design. It is important that this section is not intended to give a full account of artifact function, but rather to attempt to explore the possibility of modifying the modal theory of biological function in such a way that it would cover artifact functions.

---

[34] Philip Kitcher, "Function and Design," *Midwest Studies in Philosophy*, XVIII, 1 (September 1993): 379–97, at p. 380.

[35] See also Millikan, *Language, Thought and Other Biological Categories*, especially p. 17; and George C. Williams, *Adaptation and Natural Selection* (Princeton: University Press, 1966), especially p. 209. Even those who aim to reconsider the role the notion of design plays in the explanation of biological function (for example, Collin Allen and Marc Bekoff, "Biological Function, Adaptation, and Natural Design," *Philosophy of Science*, LXII, 4 (1995): 609–22; David J. Buller, "Function and Design Revisited," in Ariew, Cummins, and Perlman, eds., *op. cit.*, pp. 222–43) accept a weaker claim that if x is designed to do F, then the function of x is to do F.

[36] See Kitcher, *op. cit.*; Beth Preston, "Why Is a Wing Like a Spoon? A Pluralist Theory of Function," this JOURNAL, XCV, 5 (May 1998): 215–54. For a dissenting view, see Pieter E. Vermaas and Wybo Houkes, "Ascribing Functions to Technical Artefacts: A Challenge to Etiological Accounts of Functions," *British Journal for the Philosophy of Science*, LIV, 2 (2003): 261–89; and Houkes and Vermaas, "Actions versus Functions: A Plea for an Alternative Metaphysics of Artifacts," *Monist*, LXXXVII, 1 (January 2004): 52–71.

The slinky was not designed to be used as a toy that can 'walk' downstairs. It was designed to be a tension spring in a horsepower monitor for naval battleships. Nonetheless, its function now is to 'walk' downstairs. Similar examples include truck tires used for football practice and old chalkboards used as dinner tables in some trendy households.[37]

Thus, it is not true of artifacts in general that x has function F if and only if x was designed to do F. But then how can we explain artifact function?

My suggestion, not surprisingly, is that function attribution to artifacts also depends on modal facts about the token artifact. Thus, the function(s) of an artifact is fixed by what *would* contribute to the fulfillment of the goals of the agent who is using the artifact. This could be spelled out in the following way: artifact x has function F at time t if and only if some 'relatively close' possible worlds where x is doing F at t and this contributes to the fulfillment of the goals of the agent who is using the artifact are closer to the actual world than any of those possible worlds where x is doing F at t but this does not contribute to the fulfillment of the goals of the agent who is using the artifact.

The function of the slinky is to roll from one step to the other because some 'relatively close' possible worlds where it is rolling from one step to the other and this contributes to the fulfillment of the goals of the agent who is using it are closer to the actual world than any of those possible worlds where it is rolling from one step to the other but this does not contribute to the fulfillment of the goals of the agent who is using it. What it was designed for is irrelevant.

One may be slightly suspicious of the reliance on the notion of 'the fulfillment of the goals of the agent who is using the artifact', so I need to make some explanatory remarks about this notion. What if nobody is using the artifact at the moment? Would it follow that the artifact has no function? No. As we have seen, artifact function is defined by what *would* contribute to the fulfillment of the goals of the agent who is using the artifact. If nothing contributes to the fulfillment of the goals of the agent who is using the artifact in the actual world, say, because nobody is using the artifact, this does not mean that the artifact has no function. Whether it has a function depends not just on what happens in the actual world but also on what would happen if things were different. If some 'relatively close' possible

---

[37] Would it be possible to consider the person who puts these artifacts to new use as the designer? This would certainly be an option, but in this case design would not explain any properties of the artifact (except for what it is being used for); thus, we would lose the main motivation for comparing function to design.

worlds where someone is using the artifact and what it is doing con-
tributes to the fulfillment of the goals of this agent are closer to the
actual world than any of those possible worlds where someone is using
the artifact and what it is doing does not contribute to the fulfillment
of the goals of this agent, then it does have a function.

A possible worry about this way of thinking about artifact function
is the following. I could use my laptop as a doorstop, but does this
make it a function of the laptop to serve as a doorstop? If so, then
the account of artifact function I outlined here would make the con-
cept of x's function dangerously similar to what x is being used for.
And every given object could be used for thousands of things. Thus,
the danger is that if we accept the account I have been proposing,
every object will end up having thousands of functions.

It is important to point out that if I use an object as a doorstop in
the actual world then it does not follow under my definition that a
function of this object would be to be a doorstop. Again, the function
of artifacts is fixed by counterfactual facts. The function of an artifact
is not whatever it does that fulfills the goals of the agent who is using
it but what it does that *would* contribute to the fulfillment of the goals
of this agent.

Again, the definition of artifact function was the following: artifact x
has function F at time t if and only if some 'relatively close' possible
worlds where x is doing F at t and this contributes to the fulfillment of
the goals of the agent who is using the artifact are closer to the actual
world than any of those possible worlds where x is doing F at t but this
does not contribute to the fulfillment of the goals of the agent who
is using the artifact. How does this definition apply in the case of my
laptop? In the actual world, my laptop serves as a doorstop, and this
contributes to the fulfillment of my goals. But does my laptop's serving
as a doorstop contribute to the fulfillment of my goals in nearby pos-
sible words? In some, it does; in some it does not. We have no reason
to believe that some nearby possible worlds where it does are closer
than any possible worlds where it does not. Thus, we have no reason
to attribute the function of serving as a doorstop to the laptop. The
moral is that if a function of an artifact x is to do F, it is not enough
that I happen to use x for F-ing in the actual world (or that I could
do so). It is not even enough that if things were different, x would still
be used for F-ing. In order for an artifact x to have a function to do F,
it needs to be true that some 'relatively close' possible worlds where x is
doing F and this contributes to the fulfillment of the goals of the agent
who is using the artifact are closer to the actual world than any of those
possible worlds where x is doing F but this does not contribute to the
fulfillment of the goals of the agent who is using the artifact.

   The conclusion is that the symmetry between biological function
and artifact function could be restored if we accept a modal theory of
function: both the function of artifacts and the function of biological
traits are fixed by modal facts. I have only sketched, and not defended,
the possibilities of a modal theory of artifact function here. The aim of
this paper was to defend a modal theory of biological function.

IX. CONCLUSION

Finally, some readers may be skeptical about the modal theory of
function because of the appeal to possible worlds. In conclusion, I
find it important to emphasize that the modal theory of function does
not presuppose realism about possible worlds; nor does it presuppose
the Lewisian analysis of counterfactuals. I used the Lewisian frame-
work because it is the most widespread nowadays and because it
allowed me to make explicit some of the fine details of the modal
claim. But any other account of counterfactuals could be used to fill
in the details of the account. The main claim of the modal theory of
function is that function attributions have modal force. This claim
could be made with or without relying on possible worlds.

BENCE NANAY

University of Antwerp and University of Cambridge