Fagerberg, Harriet, and Garson, Justin. Forthcoming. British Journal for the Philosophy of Science
Penultimate draft - please see published version for corrections and citation purposes.

Title: Proper Functions are Proximal Functions

Abstract: This paper argues that proper functions are proximal functions. In other words, it rejects the notion that there are distal biological functions – strictly speaking, distal functions are not functions at all, but simply beneficial effects normally associated with a trait performing its function. Once we rule out distal functions, two further positions become available: dysfunctions are simply failures of proper function, and pathological conditions are dysfunctions. Although elegant and seemingly intuitive, this simple view has had surprisingly little uptake in the literature. Indeed, our position departs from that of almost every theorist who has engaged with the issue at any depth. We start by presenting three arguments for the position that proper functions are proximal: one from the specificity of functions, one from their relation to intervention, and one from their relation to pathology. We then consider two case studies evidencing the trouble that accepting distal functions causes for philosophical reflection on the nature of pathological conditions. Finally, we anticipate and respond to three objections: that there can be failure of function without dysfunction; that our account is unacceptably revisionary in respect of normal function-talk; and that our thesis over-generalises from a narrow set of cases.

1. The problem

The heart pumps; by pumping, it circulates blood; by circulating the blood, it brings nutrients to cells and eliminates waste; by doing that, it keeps us alive. Which of these activities is the heart's function? Are they all its functions, or just one? Is there some mind-independent "fact of the matter" about which of these activities is its function, or does it depend on the interests of the one doing the ascribing? And what, if anything, depends on what we conclude here?

A tempting answer is that this is a "pseudoproblem." All of these effects are equally the heart's function. Nature itself doesn't discriminate in this regard. People do. Sometimes people pick one or the other of these activities out when describing the heart's function because they're trying to explain slightly different things. The cardiologist, for example, might describe the heart's function as "pumping", while the pulmonologist describes it as "bringing nutrients to cells." Neither is more right than the other.

We call this attractive view the "pluralist contextualist" view. According to the pluralist contextualist point of view, the function of any particular trait, such as the heart, can be fruitfully thought of as a "concertina" of activities (see Papineau, 1998). Each one of these activities are legitimately functions of the heart. (That's the pluralist part.) Which one we happen to identify in our function ascriptions depends heavily on pragmatic and explanatory factors. (That's the contextualist part.) A cardiologist who studies heart disease might have an interest in the way that the heart can fail to pump, and so they may be inclined to describe the heart's function in more proximal terms. A pulmonologist who's interested in increasing the oxygen content of the blood might be inclined to describe its function in more distal terms. As Goode and Griffiths (1995) elegantly summarize the case for the pluralist contextualist view: "The apparent indeterminacy of etiological functions is a genuine indeterminacy, but a harmless one. Selection processes can be described at more or less abstract theoretical levels, all of which generate genuine, complementary descriptions of etiological function." (1995, p. 107).

From the pluralist contextualist vantage point, insisting that there is a deep, mind-independent fact of the matter as to "who's right" starts to look like narrow-mindedness; another example of philosophers muddying up clear thinking by trying to impose their own idiosyncratic metaphysical scruples. Nonetheless, we will argue that in our hypothetical scenario, the cardiologist is, in fact, right, and the pulmonologist wrong. The cardiologist's preferred function ascription mirrors the mind-independent facts better than the pulmonologist's. We also think that our conclusion matters, both for philosophy and for medicine.

In contrast to the pluralist contextualist view, we defend what we will call the proximal-function thesis: proper functions are proximal functions. That is, only the most proximal member of the "concertina" of activities produced by a trait is its actual function. The remainder are beneficial consequences or more distal selected effects of the trait performing its function. Put differently, we do not accept the existence of 'distal functions'.

How then are the "concertina" of activities related? Consider again the heart, and all the things the heart does: pumping, circulating blood, bringing nutrients to cells, keeping the organism alive. Each activity in this series is linked to the next via what Neander (1995) calls a "by" relation, or a means-end relation. (*By* pumping, the heart circulates blood; *by* circulating blood, it brings nutrients to cells, etcetera.) The most "proximal" member of this chain is, by definition, an activity that it performs but where it doesn't perform this activity *by* doing

something else. (The heart pumps, but it doesn't pump *by* doing something else.) As we move through the sequence of activities that make up our "concertina," we see that each activity requires the contributions of more and more parts. Put loosely, the heart can pump more or less "on its own;" but to circulate blood, it also needs veins and arteries and a sufficient supply of blood; to bring oxygen to cells, it requires all that *and* a working pair of lungs, and so on.[1]

With this hierarchical picture in mind, our view is that only the most proximal, or most "specific," activity of the trait is its function. Activities that occur further along in our "concertina" are proper functions of larger systems. The function of the heart is to beat. The function of the circulatory system is to circulate blood. To say that the function of the heart is to circulate blood is to commit a fallacy of division – attributing a property of a whole to a part thereof (see also Garson, 2019). By the same token, to attribute the function of pumping to the circulatory system is to commit a fallacy of composition. We think there may be an interesting psychological reason why we're more inclined to commit the first sort of fallacy when talking about functions than the second, which we will return to in section 4. 2.

One important implication of our view, both for philosophy and for medicine, is that dysfunctions are *just* failures of proper function. This might sound obvious and uncontroversial – indeed, many function theorists speak of dysfunction in precisely these terms – but as we shall see, it is a surprisingly unpopular view among those in the literature seeking to precisely define dysfunction. As we will suggest, skepticism as to this simple definition of dysfunction has tended to be motivated by the idea that a trait can fail to perform its function for two reasons: (1) because it is dysfunctional, and (2) because it is in an inappropriate environment. As we argue in section 4. 1, properly understood, (2) cannot obtain.

Although we build on the seminal works of Neander (1995), Garson (2019), Buller (1997) and others, our position sets us apart from almost everyone who has thought about this problem in a serious manner, including Millikan (1989a), Neander (1995), Goode and Griffiths (1995), Papineau (1998), Wakefield (2021) and Matthewson and Griffiths (2017), and Griffiths and Matthewson (2018; 2020). That is because all of the aforementioned thinkers took the existence of distal functions for granted. They each assert, quite explicitly, that when we attribute a

[1] Of course, one might point out that the heart pumps "by virtue of" having different components (atria, ventricles, valves) that have their own characteristic functions and that have various spatial, temporal, and causal relations to one another and to the rest of the body. But this isn't a situation in which the heart pumps *by* performing some other activity, but a situation in which the heart pumps because its *components* perform various activities.

"function" to a trait we are really describing a sequence of activities, and depending on one's explanatory interests, any particular activity in that sequence could legitimately be called its "function."[2] We deny this. We think that acknowledging the existence of distal functions contravenes some of our core, shared, assumptions about the nature of biological functions, and compromises clear thought about the relationship between dysfunction and disease.

In what follows, we will offer three arguments for the view that proper functions are proximal: one from the inherent specificity of proper functions, one from intervention, and one from the relation between biological dysfunctions and pathological conditions. We will then offer two case studies of how authors have erred in their treatment of this issue – Griffiths and Matthewson, followed by Wakefield – before considering three implications of our view.

Before we proceed, a few words on our methodological approach are in order. We take it that biological functions are things in the world, not things in our heads (e.g., Millikan 1984; 1989a). As such, our thesis is best understood as a descriptive theory of a class of real phenomena in the world. We are *not* attempting to describe the commonly held ideas or mental states which specialists or laymen 'have in mind' when applying terms such as function and dysfunction, and as such we are not beholden to usage at every twist and turn. Moreover, we are skeptical as to the utility in leaning heavily on intuitions, particularly in complicated, non-paradigm cases. Where, for example, 'function' comes apart from 'fitness' or 'dysfunction' comes apart 'bad', our intuitive sense of whether a trait is or isn't functioning as designed by evolution may easily lead us astray. Indeed, the task, as we see it, is precisely to provide a clear, instructive, and theoretically well-motivated rationale for identifying instances of function and dysfunction in complicated and unusual cases where intuitions may conflict.

Finally, we accept that a modest version of function pluralism is true. By "pluralism," we just mean that scientists use the term "function" in slightly different ways. We follow Neander (2017) in making a large-scale division between two core senses of function: *proper* functions and *minimal* functions. A proper function is the sort of function that admits of a function/accident distinction and a function/malfunction distinction. The selected effects theory is one theory of what proper functions are. It holds that a function of a trait is, roughly, whatever

[2] Neander (1995) pointed out that, in biomedical contexts, when we are trying to identify how a component of a system might fail to perform its function, we have a special interest in its most "specific" (i.e., proximal) function. But she was quite clear that any activity in the sequence of activities could constitute its function. We return to this point in Section 2.2.

it was recently selected for by natural selection or some comparable selection process (Neander 1983; Millikan 1984; Griffiths 1992; Godfrey-Smith 1993). We think the selected effects theory is the most appealing account of proper function, but our goal here is not to argue that point. The claim that functions are proximal should be a constraint on any theory of proper function.

2. Three Arguments

In what follows, we develop three main arguments for the thesis that proper functions are proximal functions. The first argument has a metaphysical character in that it points to a certain feature of reality that undergirds the ordinary practice of attributing functions to traits. The second reflects on some of the core methodological assumptions underlying the practice of fixing things that have broken. The third argument has a more "forward-looking" character: if biological functions are to play the role that many philosophers want them to play – namely, to support a naturalistic theory of disease, pathology, or disorder – then functions must be proximal. Admitting distal functions into one's ontology would subvert this purpose.

2.1 Functions are Distinctive Contributions

In the following section, we argue that proper functions are inherently specific – that is, that the constituent traits of organisms are supposed to do *different* things. We then argue that the only way to accommodate this feature of proper functions in a theory thereof is to adopt the view that proper functions are proximal.

In order for the practice of attributing functions to traits to make sense, to be coherent, reality must be a certain way. In other words, the practice of investigating functions, debating about functions, framing hypotheses about functions, and so on, depends on certain core features of the biological world. Some of those basic features of reality are so widely understood and appreciated among function theorists that they're rarely explicitly thematized as such. The thesis that proper functions are proximal, in our view, is supported by reflecting on some of these basic features of reality.

The feature of reality we wish to draw attention to is that functional systems (such as organisms) have different parts, and these parts are meant to do different things. In other words, all of the parts aren't meant to do the *same* thing. For example, for the sake of keeping us alive, the heart is meant to do one thing; the liver, a different thing; the lungs, a third thing. These

various doings are their *functions*. Put differently, a function of a trait is a distinctive contribution that a part makes, in tandem with the other parts, to keep the larger system going.[3]

This feature of reality (that different parts of a system are meant to do different things) has rarely been explicitly noted in the literature – perhaps because it appears scarcely worth pointing out. Mossio et al (2009), in defending their organizational account of functions, say, quite explicitly, that functions only belong to systems that exhibit "organizational differentiation" such that we can "distinguish between different contributions to self-maintenance made by the constitutive organization," and in which each component makes a "specific contribution to the conditions of existence of the whole organization." (p. 826, 2009). More simply put, an organizationally differentiated system is one that has different parts that do different sorts of things to keep the whole running.

This basic feature of reality – the specific nature of proper functions – is hierarchical and iterative. The organism, for example, consists of numerous systems (circulatory, digestive, respiratory…) that do different things to keep the organism "running." These systems, in turn, consist of various parts and processes (the heart, arteries, veins) that do different things that enable the system to make *its* distinctive contribution to the body. And each of these (e.g., the heart) has different parts – the valves, the ventricles, the atria – that do different things that enable the heart to make *its* distinctive contribution.

If what we've said is true – that different parts of the system are meant to do different things to keep the system up and running, and these doings are its functions, then functions *must* be proximal. We can show this via reductio ad absurdum. Suppose that there are distal functions (e.g., that the function of the heart is not only to pump, but to circulate blood, to bring nutrients to cells, and to keep us alive). Then all of the different parts of the body would have the same function. The function of the heart would be to keep us alive, as well as the liver, and the lungs. Distal functions are not distinctive contributions, but shared contributions. Therefore, functions are not distal.

[3] Our view is consistent with the fact that, in some cases, two or more bodily parts or processes have the exact same function. We have two kidneys, each of which has the same function - to filter blood and to serve as a backup in case the other fails. Here, the type/token distinction can be of help (e.g., Neander 1991, 174). Each of my two kidneys has the same function because they are two tokens of the same type.

2.2 Functions and Appropriate Interventions

The second argument stems from reflecting on the intimate relation between function attributions and interventions designed to fix systems. One core purpose for attributing functions to the parts of systems is to identify and intervene when the system isn't operating "as it should." Buller (1997) argues that proper functions are proximal functions by considering the connection between function attributions and intervention. He uses an example from the realm of artifacts. Suppose my car's engine stops running; I'm unhappy about this, and I want to get it running again. This requires distinguishing carefully among the parts of the engine that are "functioning properly" (i.e., doing what they're supposed to do) and those that are not "functioning properly" (not doing what they're supposed to do). As we shall argue, it turns out that this distinction *just* amounts to distinguishing between those parts that are, and those that are not, performing their proximal functions.

Go back to our engine. If we opt to describe the function of each component of the engine distally, then none of the components are "functioning properly" because all of them are failing to serve their distal function of making the car go. But for the purpose of intervention, it would be a foolish waste of time to replace the crankshaft when the solenoid timers are at fault. The ordinary practice of intervening appropriately on systems requires that we attend carefully to the proximal functions of a systems parts and processes.

In a sense, Neander (1995) was getting at the exact same point in her canonical discussion of function indeterminacy. However, because she was not focusing on the issue on intervention as tightly as Buller, she erred in certain ways. She used the following example. Suppose a female mammal is infertile because her fallopian tubes are blocked. If we specify functions distally, we would be forced to conclude that all of the components of the reproductive system are dysfunctional because none of them can perform their function correctly of aiding pregnancy. If we specify functions proximally, we would say that the reproductive system is dysfunctional, and, that the fallopian tubes are dysfunctional, but that none of the other parts of the reproductive system (such as the ovaries) are dysfunctional. She argued that not only was the latter (proximal) way of speaking more informative, but that it more accurately mirrored ordinary biological usage: "Not surprisingly, biologists do not maintain that the ovaries are malfunctioning because conception is impossible because the Fallopian tubes are blocked." (Neander, 1995, p. 120).

We agree in the main with Neander's treatment, but for one main difference. She thought, consistently with the pluralist contextualist view, that all of a trait's activities (e.g., beating, circulating blood, bringing nutrients to cells, etcetera) could equally count as its "functions." She observed, however, that this profusion of functions creates "conflicts in functional norms" since it opens the possibility that a given trait might perform one of its "functions" well but not another (e.g., perhaps the heart can beat but not circulate blood). In those cases, she thought, one should always avert to the trait's most proximal function (or as she put it, its most "specific" function). A trait has many functions, but only dysfunctions when it cannot perform its most specific function (also see Papineau 1998 for a similar treatment).

Neander's view – in which all of the effects of a trait are its "functions," but the most proximal effect is the privileged tiebreaker when these "functional norms" clash – strikes us as unnecessarily complicated. On our view, there is simply no conflict between functional norms, and hence no need for a rather *ad hoc* principle for remediating this conflict. In the case she describes where the fallopian tube is blocked, it is not the case that the ovary is functioning well at one level of description, but functioning badly at another level, leaving us confused on how to proceed. We simply reject that these "functional norms" conflict. In the case she describes, the ovaries are functioning perfectly well and the fallopian tubes are not.

2.3 Dysfunction and Pathology

Our final argument proceeds from the premise that there is a close relationship between dysfunction and pathology. This argument should appeal to the function theorist in general, but will carry special force for those who defend the thesis that pathological conditions (diseases, disorders) involve biological dysfunctions (see e.g. Wakefield, 1992; Boorse, 1977; Matthewson & Griffiths, 2017; Griffiths & Matthewson, 2018). The structure of the argument is as follows. Assume that pathological conditions are biological dysfunctions – at least necessarily, but perhaps sufficiently. In other words, let us suppose that the following premise is true:

> <u>Premise 1:</u> Pathological conditions are biological dysfunctions.

Assuming that this premise is true, and that dysfunctions are failures of proper function, what would proper functions have to be like? On which view of proper function does premise 1 hold

true? As we shall argue in what follows, in order for pathological conditions to be biological dysfunctions, proper functions would have to be proximal functions. If our thesis is false – that is, if proper functions can be distal functions and dysfunctions *failures* of distal function – then the plausibility of premise 1 is similarly under threat.

Let us first motivate the view that failures of proper function *only* map onto pathology given a theory of proper functions as proximal functions. Consider again the heart. The heart does many things – it pumps, it circulates blood, it contributes to fitness. In any given organism, any of these effects may fail. Which of these failures would imply that the heart is *pathological*? The organism can certainly fail to reproduce whilst the heart is perfectly healthy – for example, in the absence of available mates. Similarly, the heart might fail to circulate blood for reasons that have nothing to do with the heart, for example, in the event of an arterial clot blocking the passage of blood through the arteries. However, if the heart is *not pumping*, something has certainly gone awry with it. It would seem then, that failure of proximal function, at least in the case of the heart, maps unto failure of health.

Now suppose that we held, instead, that dysfunctions were failures of proper function, but that proper functions can be distal selected effects. What implications would such a view have for the dysfunction account?

Our view is that including "distal dysfunctions" among legitimate failures of biological function threatens to render the dysfunction account severely over-inclusive and thus implausibly revisionist. Consider again the heart. When an organism is failing to reproduce in the absence of appropriate mates it is true that one of the distal selected effects of the organism's heart is not being yielded – i.e. reproduction. But is there pathology? If so, *every* trait within the body of a non-reproducing organism would be pathological, regardless of the cause of the organism's failure to reproduce. The heart would be pathological, the lungs would be pathological, the liver would be pathological and so on. It would not matter that these traits are performing their proximal functions just fine, for there is a distal selected effect that is failing. This would yield some very revisionary attributions of pathology, and thus threatens to render the dysfunction account absurdly out of step with medical science and practice.

In sum, the view of function on which failures of proper function are pathological conditions, is the view of function on which proper functions are *proximal.* In other words, these two

philosophical commitments go hand in hand. This lends credence to our overall thesis in two ways.

Those that are committed to premise 1 – i.e., proponents of a dysfunction-account of disease – will want to accept our position on function, seeing as acceptance of premise 1 implies acceptance of our thesis, and since rejection of our thesis similarly implies the rejection of premise 1. Accordingly, those motivated to defend a dysfunction-account really ought to commit themselves to a theory of proper functions as proximal – on pain of rendering the dysfunction-account implausible.

Those theorists that are *not* committed to a dysfunction-account of pathology per se should nonetheless note the impressive explanatory power of our position, and the theoretical unity it yields. On our view, proper functions are proximal, dysfunctions are simply failures of proper function, and pathological conditions are (quite plausibly) biological dysfunctions. Our thesis thus helpfully illuminates the logical relations between three core ideas in this area – function, dysfunction, and disease. Accordingly, the fact that our view supports and is compatible with a dysfunction-based view of pathology, counts in its favor.

3. Two Case Studies

In the previous section we presented three novel arguments for a theory of proper functions as proximal functions. In the following section, we illustrate the value of our thesis by showing that accepting distal functions causes trouble for philosophical reflection on medicine and psychiatry. We will focus on two projects that we think err in endorsing (or appearing to endorse) the existence of distal functions: Wakefield (1992; 2021), and Matthewson and Griffiths (2017; 2018). Both Wakefield and Matthewson and Griffiths seek to define pathology (disease, disorder etcetera) in terms of dysfunction. We contend that the authors' acceptance of distal functions, and what we might call "distal *dysfunctions*," serves to undermine some of the core motivations of a naturalistic theory of disease; that is, to clarify the scope of medicine's authority and prevent its overreach.

3. 1. Griffiths & Matthewson

Griffiths and Matthewson have, over the course of a number of recent papers, advanced a theory of disorders as failures of naturally selected effects (2017; 2018; 2020). In this section, we argue

that Griffiths and Matthewson's prima facie convincing attempt at defining disease in terms of dysfunction falters due to their inclusion of "distal dysfunctions" and their pluralist contextualist take on function indeterminacy.

Griffiths and Matthewson argue that there is a need for an objective, naturalistic dysfunction criterion in defining disease. Such as criterion is needed to counter the threat of social relativism in disease attributions: "If an exclusively social account of disease is correct, then psychiatrists in the Soviet Union did not make a biological error when they diagnosed political dissidents as suffering from the disease of "sluggish schizophrenia." (p. 448, 2017). Accordingly, we need a robust biological dysfunction criterion in a theory of pathology. But what is biological dysfunction exactly?

Like us, Griffiths and Matthewson commit themselves to a theory of function on which functions are naturally selected effects of evolved traits. Also like us, they subscribe to a view of dysfunctions as mere failures of proper function (see Griffiths and Matthewson, 2020). However, unlike us, they fail to define proper functions as proximal. This opens the door to a severely over-inclusive notion of dysfunction and, thus, of disease. Moreover, their pluralist contextualist take on function indeterminacy, renders their account inappropriate as an objective naturalist criterion intended to serve as a check on relativism.

In their 2017 paper 'Four Ways of Going Wrong', Matthewson and Griffiths identify four types of dysfunction each of which, they suggest, could in principle constitute pathology.[4] Rather than provide a precise theoretical account of the distinction between these four 'ways of going wrong', or types of dysfunction, the authors distinguish them by way of illustrative examples. Among others, they ask us to consider the fate of an unfortunate glow worm fruitlessly searching for mate in a light polluted urban environment:

> Consider a European glow-worm (*Lampyris noctiluca*) living in urbanized areas, where there is a substantial amount of ambient light in the evening. Male glow-worms usually locate females by their light signature. However, it has been demonstrated that even quite dim light impairs their ability to discern female glow-worms signalling their availability…[S]omething seems to have gone wrong here: Male glow-worms are failing to locate prospective mates. (p. 453, 2017)

[4] The authors are rather equivocal on this point, and appear to recognize that some of them might not be pathological. We'll return to this interpretive issue shortly.

Matthewson and Griffiths are right that things aren't going quite to plan here. However, the failure is not one of proximal function. Instead, a more distal effect (the identification of suitable mates) which the glow-worm's eyes normally contributes to is not being yielded, due to an abnormal environment. This is a classic case of environmental or evolutionary mismatch. It is also, clearly, not a case of pathology. If we were to count such failures of distal selected effects as instances of pathology, then we would be forced to bite some substantial bullets. The point becomes obvious once we apply a similar line of reasoning to human biology. Suppose a distal selected effect of the amygdalae is physical escape from danger (via activation of the sympathetic nervous system). Suppose that you are strapped into a seat on an airplane, and so cannot simply run away from the scary movie you are currently enjoying. Are you pathological? Clearly not. And the reason why is really very simple – there is no failure of proximal function.

Assuming a pluralist contextualist take on function indeterminacy – on which we are free to privilege any effect in the casual chain, relative to our explanatory interests – even more trouble follows. In many cases there will simply be no fact of the matter about whether the trait in question is or isn't dysfunctional; precisely because, without a solution to the problem of function indeterminacy, functional norms will appear to conflict and thus to draw in different directions. Is the glow worm functioning as designed or not? It depends. If you think of the function of the glow worm in proximal terms, then it is functioning precisely as designed by natural selection. If you attend, instead, to more distal effects which the glow worm's visual perception normally contributes to in its selective environment, then there appears to be a genuine failure of proper function. After all, the glow worm is unable to identify potential mates in his environment – a (distal) selected effect. If, in accordance with the pluralist contextualist thesis, we are free to attend to any activity in the causal chain, as it suits our explanatory objectives, then it seems there is more than one correct answer to the question as it is posed in the above. The glow worm both is and isn't functioning as it should. You both do and do not have a pathology as you squirm in your airplane seat, terrified at the monster on screen but unable to run away.

An anonymous reviewer suggested that perhaps we are mischaracterizing Griffiths and Matthewson's view. In fact, the authors only think of 'mechanism failure' (as they term it) as dysfunction. The other three 'ways of going wrong' are examples of things going badly for the organism without any dysfunction occurring. We think there is good textual evidence that

Griffiths and Matthewson think of the 'four ways of going wrong' as varieties of dysfunction, for example: "Many people would describe this situation as one of dysfunction. However, because we need to distinguish between different kinds of dysfunction, we will use the more specific term mechanism failure" (2017, 453; emphasis ours). However, we agree that there is some ambiguity here. The authors are also somewhat equivocal as to how the 'ways of going wrong' are supposed to relate to disease, and appear to recognize that some such cases might not be pathological (see 2017, 461). However, given the paper's goal of elucidating the potential for a more inclusive dysfunction-based account of pathology, it is unclear to us why the authors would have included these scenarios unless they believed them to be in principle grounds for pathology. We are thus led to conclude that the authors' think failures of distal effects (such as the glow-worm's failure to identify mates) can constitute both dysfunction and pathology.[5]

In sum, Griffiths and Matthewson's take on biological proper functions renders their account of dysfunction unable to serve the very role they insist motivates a biological criterion for pathology in the first place. The inclusion of distal selected effects threatens to render their account severely over-inclusive, whilst the pluralist contextualist thesis offers no objective guide as to which 'function' – distal or proximal – we should attend to. In sum, if functions are to serve as a check on relativism, then functions have to be proximal.

3.2. Wakefield

Wakefield's well-known contribution to the philosophy of psychiatry is a philosophical definition of mental disorder known as the "harmful dysfunction" account (e.g., 1992). We are broadly sympathetic to this view of mental disorders but think that it is stymied by Wakefield's acceptance of "distal dysfunctions." We think this recognition subverts the core motivation of the harmful dysfunction analysis, namely, to prevent the "overmedicalization" of distressing or disturbing psychological phenomena.

Wakefield believes, as we do, that the selected effects theory is the correct empirical or scientific theory of what functions are.[6] As such, his harmful dysfunction analysis implies that

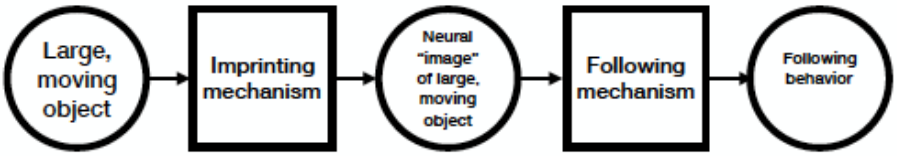
[5] Perhaps they do not – perhaps the cited passage is just a careless formulation, and they in fact have something else in mind. If so, then their account is at the very least ambiguous and in need of precisification. It is up to Griffiths and Matthewson to clarify their account. For now, we will proceed under the assumption that they think of dysfunctions as failures of (not necessarily proximal) function.

[6] For Wakefield this is a non-conceptual empirical claim and thus technically not part of his 'harmful dysfunction' conceptual analysis of medical disorder per se (see Wakefield, 2000) Wakefield's methodological commitments in

disorders are a subclass of failures of selected effects functions – i.e. harmful failures to function as designed by evolution.

So far, Wakefield's project strikes us as plausible and commendable. However, he departs from us in one key respect: Wakefield rejects our thesis that proper functions are always proximal functions, and rejects Neander's claim that a trait is dysfunctional if and only if it is unable to perform its most "specific" (or most proximal) function. In Wakefield's view, there is a whole class of cases in which a trait can be dysfunctional, not because it fails to perform a proximal function, but because it fails to perform a distal function despite performing a proximal one. The case he has in mind is "imprinting gone awry."

He asks us to consider the fate of an unfortunate gosling. In effect, for a few days after birth, the gosling has a mechanism in its brain which has the function of creating a neural "image" of the first large moving object it sees upon hatching. In most environments, the first large moving object it sees will be its own mother; so, in general, we can say that the imprinting device has the "distal function" of creating an image of the gosling's own mother. The image so created, in turn, is the input to a second mechanism, one that has the function of following around whatever that "image" refers to. (See figure 1).



We know, of course, that imprinting can and does go awry, as is the case of the goslings that imprinted upon, and subsequently followed around, Konrad Lorenz's boots. Now, here's the question that Wakefield poses: in a case of imprinting gone awry, is there any dysfunction inside the gosling? He imagines a situation where the first large moving object the gosling sees is a porcupine, and the imprinting mechanism induces in the gosling a disposition to follow about the

this regard are somewhat complex, and unpacking them in any detail would lead us too far afield (see Fagerberg, 2023 for a recent review).

porcupine. Clearly, this situation is maladaptive for the gosling – we can guess that its life expectancy will be rather short – but is there also a dysfunction? Wakefield thinks there is.

Wakefield acknowledges that if we only consider the *proximal* function of the imprinting mechanism, there has been no dysfunction (see Wakefield 2021, 394). The imprinting mechanism has, at the most proximal level of description, the function of creating a lasting neural image or representation of the first moving object it sees. In the porcupine case, it discharged that function with aplomb. (In fact, Wakefield acknowledges that the imprinting mechanism would have been dysfunctional if it *failed* to imprint on the porcupine.)

Still, he thinks, we should nevertheless say that it is a dysfunction. That's because, if you placed the gosling in its "normal environment," that is, the environment in which it was designed to function, it would still be at a serious disadvantage in life. Even if you placed the gosling right next to its mother, it wouldn't follow her around. Instead, it would seek the target of the neural image – a porcupine. The fact that the gosling still shows maladaptive behavior even in its normal environment is, for Wakefield, evidence that there is a dysfunction.

We think there are two major problems with the way that Wakefield has described this situation. The first is what we can call the "missing function-bearer" problem; the second is that we think his intuition that there is a dysfunction is being inappropriately influenced by features of the thought experiment that ought to be irrelevant to principled and theoretically well-motivated judgments about dysfunction.

We will start with the problem of the missing function bearer. Let us suppose, along with Wakefield, that something inside the gosling is dysfunctional. If so, which trait, precisely, is dysfunctional? As we can see from figure 1, there are several logically possible candidates for which trait that might be. It may be that the imprinting mechanism is dysfunctional; it may be that the neural image so imprinted is dysfunctional; it may be that the "following" mechanism is dysfunctional; it may be that the following *behavior* as such is dysfunctional. But we don't think any of these candidates are particularly plausible candidates for being the function bearer.[7] We'll focus, in what follows, on arguably the two most intuitively plausible candidates: the imprinting mechanism itself, and the neural image so produced.

[7] Wakefield himself seems to vacillate between various construals of what exactly is dysfunctional. Sometimes he suggests it's the imprinting mechanism itself; sometimes the neural image so imprinted (Wakefield, 2021, 395).

Let's start with the imprinting mechanism itself. Is the imprinting mechanism dysfunctional? Described proximally it is not dysfunctional. It discharged its proximal function perfectly when it produced a neural image of the first large moving object which happened to enter into the gosling's visual field. If we want to ascribe a 'distal function' to the mechanism, we could say that its distal function is to help the gosling follow around its mother (or follow around a protector, nutrient provider, and so on).

But we do not think this gives Wakefield the result that he wants. In Wakefield's view, the imprinting mechanism (that is, the device that takes a "snapshot" of the first large moving object it encounters and creates an image of that object) is only dysfunctional if it were the case that, were it in its normal environment, it wouldn't be able to perform its function. The question thus becomes: if the snapshot-taking mechanism was in its normal environment – the one with the mother around – would it be able to perform its function? And of course, it would. In the imprinting mechanism's normal environment (that is, one in which the gosling's mother was the first large moving object it saw upon hatching) the mechanism would have performed its function just fine. Thus, the imprinting mechanism itself is not a strong candidate for being dysfunctional – even granting Wakefield's inclusion of distal dysfunctions.

Perhaps a more suitable candidate, then, for a dysfunctional trait in imprinting gone awry is the neural image itself; a configuration of synapses that represents the porcupine to which the gosling was exposed upon hatching, and which has the function of orienting the gosling's movements – i.e., of causing it to follow the object of its attachment around. This image is retained and continues to guide the behavior of the gosling after the imprinting mechanism has discharged its function of imprinting, and so could theoretically serve as the bearer of Wakefield's distal dysfunction. But what exactly is the function of this enduring neural image?

This question turns out to be quite complicated. That's because, technically, the neural image so produced has what Millikan (1984) calls a *derived* proper function rather than a *direct* proper function. Her basic idea is this. Some traits have functions because they were selected for: the heart, the lungs, the liver, and so forth ("direct" proper functions). Some traits have functions because they were produced by a mechanism that was selected for producing things of that nature ("derived proper functions"). It's easiest to think about this concept in the case of artifacts. What's the function of a Polaroid camera? Well, its direct, proper function is to produce an image of whatever it happens to be pointed at. This is also what Millikan calls a "relational proper

function." Now, once the camera is pointed at a specific object, say, a sunset, then it comes to have a more specific function, namely the function of producing an image of that sunset. This is what she calls an "adapted proper function." The image so produced then comes to have a "derived proper function." Its function is derived, as it were, from the adapted proper function of the device. In this case, its function would be something like 'induce sunset-related thoughts'.

How should we describe the gosling case using Millikan's terminology? The imprinting mechanism has a relational proper function (produce neural images of the first moving object you see). If the first moving object is a porcupine, then that mechanism has the adapted proper function of producing an image of the porcupine. The neural image so produced would then have a derived proper function: to induce porcupine-related thoughts and behaviors (following porcupines, for example). As a consequence, the neural image is not *failing* to perform its proper function of inducing mother-following behaviors, but *succeeding* in performing its (derived) proper function of inducing porcupine-following behaviors.

Wakefield might retort that the neural image has, in addition to its proximal derived function, a further, distal derived function: to help the gosling follow around its mother. However, this would lead to an apparent absurdity. Should we say that the neural image in question has the proximal function of helping the gosling follow around the porcupine *and* the distal function of helping the gosling follow around its mother? If so, that would imply (following Neander's hierarchical conception of the "by" relationship between activities in the casual chain) that goslings sometimes follow around their mothers *by* following around porcupines, in the same way that the heart often circulates blood by pumping – which is false.

Let us be entirely clear here. We don't have any settled convictions on how one must describe the function of that particular neural image. We're open to alternative ways of understanding how the neural image might come to have a function and what function(s) it has. As we emphasized above, the issue doesn't depend on what our intuitions say, as, for us in these complicated non-paradigm cases, our intuitions are relatively silent and unlikely to be of much use. What is important is having a principled theoretical rationale for identifying functions in these unusual and complex cases. If one is inclined to speak of distal functions, and feels there to be some compelling rationale for doing so, we would simply insist that it is not obvious that the neural image so produced has the distal function of helping the gosling to follow its mother.

What, then, explains the strong intuition that there is something *wrong* with our unfortunate gosling? We shall close by offering an alternative explanation for Wakefield's intuition: we believe that in cases such as imprinting gone awry our intuitions are often led astray by extraneous facts, and so are bad guides.

For Wakefield, the evidence of a dysfunction is that the gosling who imprinted upon a porcupine would not follow around its mother even if it were returned to its normal environment, that is, one in which its mother is present. That means that whether or not the gosling is dysfunctional should *not* depend on what, exactly, the gosling imprints on, so long as it is not its mother (a porcupine, a pair of boots, a deer, a young child). But we suspect that Wakefield's intuition that something is dysfunctional is "contaminated," as it were, by irrelevant information about the scenario – namely, that if a gosling imprints on a porcupine, things are not going to "go well" for it. Suppose, instead, that the gosling imprinted on a benevolent human, who took it home, fed it, and cared for it. Let's suppose that gosling ends up having a longer, healthier life on that account. Would our intuition still tell us that there's a dysfunction? We think that regardless of what our intuition says, there should be parity between the two cases. If Wakefield's intuition tells him that there's only a dysfunction in the first case but not the second, then we would worry that those intuitions are distorted by irrelevant features of the example – precisely, whether things are likely to go well or badly for the gosling. This threatens to undermine the naturalistic distinction between objective cases of dysfunction, and judgments of whether it is 'good' or 'bad' that the organism is operating in this way.

4. Objections

In the previous two sections, we first offered three arguments for the position that proper functions are proximal functions, and then showed how prominent theorists have erred in failing to factor in this core feature of proper functions. In the following section, we anticipate three objections to our thesis.

4.1. No Failure of Function Without Dysfunction

We have said, throughout, that all that is required for dysfunction is for the trait to fail to perform its proximal function. This may have struck some readers as much too strong. Surely, a trait can fail to perform its function simply because it is in an abnormal environment? Are we not

obviously mistaken in refusing to allow for this possibility? Many in the literature have assumed that a trait can fail to perform its function, without being dysfunctional, when the organism is in a bad or unhelpful environment. For example, Millikan draws the following distinction in a recently published paper: "The confusion here is between (1) not performing or failing to perform a function and (2) malfunctioning. Many perfectly normal functional traits succeed in serving their functions only part of the time…because they need cooperation from the environment or from other parts of the animal's system to succeed." (Millikan, 2023, p. 3)

Our view is that theorists have been led to believe that a trait can fail to perform its function in the absence of dysfunction because they have 1) failed to recognize that proper functions are proximal functions, or 2) failed to appreciate the widespread significance of 'response functions' in biology (see Neander, 2017). Once we factor in these two principles, many putative examples of failure of function without dysfunction can be shown to, in fact, be cases of the trait performing its function precisely as designed.

Let's start with the first sorts of cases. Sometimes, one might be led to think that a trait can fail to perform its proper function without being dysfunctional, simply because one has misidentified the proper function of the trait. Instead of picking out the proper (proximal) function of the trait, one has instead picked out one of the more distal selected effects to which this proximal function normally contributes. One might then reason as follows: 'Blood-circulation is the function of the heart, but it is possible for the heart to fail to perform this function simply because it is in an abnormal environment – perhaps when there is a problem with one of the arteries. However, the heart would not be dysfunctional in this case. So it is possible for the heart to fail to perform its function without being dysfunctional.' Such cases are easily resolved: proper functions are proximal functions. So the heart *isn't* failing to perform its proper function in cases where an arterial problem disrupts circulation, provided the heart is still pumping.

A more complicated class of cases concerns those biological traits which have as their proximal function a response function and, in particular, a *complex* response function. Sometimes, a trait T has the proximal function of yielding some effect E in response to or in proportion to some signal or input S. In such cases, trait T has a *response function* (Neander, 2017). When a trait has a response function, we might mistakenly conclude that the trait is failing to perform its function without being dysfunctional, because we fail to realize that the trait's

function is a response function. That is, because we fail to detect that the proper function of the trait in question is to do E *only* when S occurs, we fail to recognize that cases of no S and no E are not cases of T *failing* to perform its function but in fact cases of T performing its response function just fine. If there is no S, then T shouldn't be E-ing – and in fact would be dysfunctional if it *were.*

For example, we might think, erroneously in our view, that the function of the uterus is 'gestating foetuses'. Thus, when a particular female organism is not pregnant, the uterus is not performing its function. However, her uterus is not dysfunctional. So isn't this a failure of function without dysfunction? This apparent counter-example is swiftly set aside once we realize that the uterus technically has a response function. It is a simplification to say that the function of the uterus is to gestate foetuses. In fact, the uterus is only supposed to gestate foetuses in response to a range of other biological processes occurring, beginning with fertilization. So, really, the uterus' proper proximal function is a response function – to gestate foetuses *only when* fertilization occurs. Once we realize that the uterus has a response function, it becomes immediately obvious that in the absence of fertilization, the uterus is *not* failing to perform its function at all. In fact, it is performing its proper function precisely as designed. No fertilization, no gestation.[8]

Our view is that response functions are much more ubiquitous in biology than realized by many contributors to this debate, and that this has led many reject our view prematurely. For example, in his otherwise persuasive 2019 book, *What Biological Functions Are and Why They Matter*, Garson writes:

> Sadly, being unable to perform something's most proximal function isn't quite enough for dysfunction. […] Consider an unplugged toaster. It cannot perform its function, no matter how proximally that function is described (say, to heat some coils when a lever is pressed), but it isn't dysfunctional. […] The mere inability to perform one's most proximal function cannot be enough. (Garson, p. 126, 2019)

[8] It's certainly possible for someone to have principled, theoretical objections against the idea of response functions. The point that we wish to make here is simply that it is not necessary to recognize a category of "function failure without dysfunction" in order to accommodate apparent counterexamples. Moreover, we appreciate that introducing the notion of a response function makes it far more difficult to determine whether a system is, or is not, functioning correctly – since such a determination requires a rich "model" of its inner and outer environment and the sorts of stimuli which it's designed to be sensitive to. We think this is a strength of our theory, and not a weakness.

Garson concludes that the toaster is in a state of failing to perform its proximal function due to an unhelpful environment. We would counter that, technically, the toaster has a response function as its proximal function. The function of the toaster is to heat coils *when* a lever is pressed and *when* the toaster is connected to the electricity supply. In the case of the toaster, these variables are (as we shall use the term) 'complex' – that is, there is more than one type of signal to which the toaster is designed to be sensitive. Our view is that many biological functions (sight, reproduction) are like this too.

On our view, if the toaster is *not* connected to the electricity supply – that is, if the toaster is not plugged in – then the toaster is not *supposed* to heat coils. As such, Garson's toaster is performing its response function precisely as designed. It would, on the other hand, be a failure of function if the toaster warmed up when it wasn't plugged in – for example, while stored away in a card-board box in the attic. In such an eventuality, we might exclaim, 'the toaster has malfunctioned and caused a fire in the attic!'

This illustrates a key feature of response functions. Response functions can fail in precisely two kinds of ways. (1) S obtains, and T does *not* do E. For example, the toaster is connected to the electricity supply, but the coils do *not* heat. (2) S does not obtain, but T *does* do E. The toaster is *not* plugged in, and yet somehow the coils heat up regardless. In our view, these are the only two ways in which a response function can fail to be performed, and they are both cases of legitimate dysfunction.

Factoring in response functions may have far-reaching implications for theorizing about the nature of beliefs. What, for example, is the function of the "belief-fixing mechanism" (Millikan 1989b), that is, the device that's designed to produce beliefs? Is it to produce true beliefs? Or is it to produce beliefs that best cohere with the perceptual evidence? One might think that, in the best of cases, the two activities stand in a by relation to one another: by producing beliefs that best cohere with the perceptual evidence, the system (when all goes well) produces true beliefs. But now consider the case of an auditory hallucination – perhaps one seems to hear a conversation taking place just behind a closed door but there is, in fact, nobody there, and subsequently forms the false belief that there are people behind the door. Is this a malfunctioning belief-forming mechanism, or a mechanism that is laboring exactly "as designed" under suboptimal perceptual conditions? We think the latter. This conclusion may have important implications for the treatment of delusions: don't presume, in the face of delusional beliefs, that

the belief-forming mechanism is faulty (see, e.g., Bortolotti 2022 for discussion). In our view, the difficult task of teasing apart the functioning and malfunctioning mechanisms within a complexly interdependent system such as the brain can only be attempted once we recognize that proper functions are proximal, and that many traits have as their proper, proximal function to do something in response to something else.

Now, of course, none of the above should be read as denying that bad environments and unfortunate circumstances can be *causes* of failures of function. Of course, they can. For example, in an environment with no food or water, we will quickly die of thirst or malnutrition. However, these are cases of the environment causing failures of proximal function and, thus, dysfunction. As such, these cases do not threaten a view of dysfunction as, simply, failures of proximal function.

4.2. Ordinary Function-Talk

Here is a potential drawback to accepting that proper functions are proximal functions. Sometimes, scientists *do* describe a trait's function in distal, rather than proximal, terms. Consider the motor neuron that causes a muscle contraction by releasing acetylcholine. In our view, the function of the motor neuron is to release acetylcholine when activated (this is another example of a response function). It is not to contract the muscle – a more distal selected effect. But we imagine that most physiologists would be comfortable saying, "the function of the motor neuron is to contract a muscle." Our view seems to entail that a large number of function ascriptions that scientists make are, strictly, false.

We suggest that in some contexts it's more useful to interpret such statements not as false, but elliptical. To be precise, taken literally that statement is false; but often enough, we can read the statement as shorthand for a somewhat longer, more cumbersome statement that is both true and informative. That longer, more cumbersome statement is that the function of the motor neuron is to release acetylcholine, and *one beneficial consequence* of it doing so is that it causes a muscle contraction. We could even say that the "function of the motor neuron is to release acetylcholine, because doing so causes a muscle contraction." We agree that in explaining why it is that the motor neuron has the function of releasing acetylcholine, one must cite the fact that doing so leads to a muscle contraction. We also think that, if someone lacks the background knowledge of physiology to appreciate why the motor neuron's releasing acetylcholine is beneficial, then it

would be appropriate to fill them in on the positive benefits, which include muscle contraction. So understood, such locutions are not only harmless, but quite informative. Still, we want to insist that there is a difference between our basic ontology of function and the pragmatic and epistemic considerations that guide function ascriptions.

The fact that it's often quite useful and informative to draw someone's attention to the beneficial distal consequences of a proximal function might explain why, as we noted in the introduction, it's far more common to commit the fallacy of division when talking about functions than the fallacy of composition. It's more common to say, for example, that the function of the acetylcholine receptor is to cause a muscle contraction than it is to say that the function of the muscular system is to release acetylcholine. The latter *sounds* wrong on its face; the former sounds right. One explanation is that the statement "the function of the acetylcholine receptor is to contract muscles" is plausibly read as elliptical for a statement about a useful, more distal effect of the acetylcholine receptor's performing its function.

4.3. Are We Overgeneralizing from a Limited Pool of Examples?

Many of the paradigm examples we've scrutinized (the function of the heart is to beat, not circulate blood) are drawn from biomedical contexts in which the proximal reading of function is already the most natural, intuitive or practically useful. One might suspect, then, that we've somehow "stacked the deck" in favor of our proximal reading of function by virtue of a carefully-curated list of examples. Isn't it possible that if we investigated other contexts, such as evolutionary contexts (the function of the zebra stripes is to deter biting flies) or ecological contexts (the function of a predator population is to limit prey populations), a distal reading is going to be more plausible?

Our thesis is that proper functions are proximal functions. Wherever proper functions are invoked, proximal functions are invoked. However, we've got to exercise caution here. As noted in Section 1, scientists sometimes deploy a different sense of "function," than the proper function sense. We call that, following Neander, the minimal sense of function. For example, philosophers of ecology are still quite divided regarding which concept, or concepts, of function is typically deployed in ecology. It wouldn't constitute a counterexample to our view to cite a specific function ascription in ecology ("the function of a predator population is to limit prey populations") where the proper function sense of "function" is not obviously on display. Still, it's

reasonable to suspect that paradigmatic evolutionary contexts in which proper functions are invoked ("the function of zebra stripes is to deter flies") cannot be easily accommodated within a proximal function framework. (We set aside, for the time being, what we noted in Section 4.1, namely that many function statements can be accommodated easily within a proximal function framework by noting that they are elliptical.)

That said, we have two arguments for why the plausibility of our proximal function thesis does not depend on a handful of carefully-selected examples. First, we began our paper with three arguments for the proximal function thesis; the validity of the first in no way depends on the specific context at hand, and the validity of the second (the appeal to intervention) plausibly is context-free as well. In other words, if those arguments are valid, then they're valid for any context in which proper functions appear. Our first argument appeals to the specificity of function: when one approaches any complex system for the purpose of assigning proper functions to it, one notes, first and foremost, how that system divides into different components and how the different components do different things to make the system "go." This is true regardless of what sort of system we're describing: an organism, an ecosystem, a social system, an artifact. Our second argument points to the link between function and intervention. It's true that when we think of intervention, we often think about biomedical sorts of interventions – fixing or replacing defective parts – which lends itself naturally to a proximal function reading. We conjecture, however, that there's an intimate conceptual connection between function ascriptions and possible interventions. For example, even in the context of ecosystems, we're generally preoccupied with mapping ecosystem functions because we are anticipating possible interventions in the case of ecosystem breakdown or collapse.

Our second response to the critic is simply to point out that we *do* analyze in detail, in Section 3.2, an initially plausible case of distal dysfunction without proximal dysfunction, namely, Wakefield's case of imprinting gone awry. There, we advanced two independent reasons to think this is not a plausible case of distal disorder without proximal dysfunction. Put differently, it's not as if we've ignored or sidestepped those cases that, on the face of things, would appear to be most problematic for our view; we've confronted them squarely and shown that they do not, in fact, pose a genuine challenge. For these reasons, we feel confident that the plausibility of our proximal function thesis does not tacitly depend on an overgeneralization from a biased or limited pool of examples.

5. Conclusion

In sum, we have argued that proper functions are proximal functions; that is, that distal biological functions properly understood do not exist. We think the non-existence of distal functions is supported not only by deeply entrenched ontological assumptions that have long animated the functions debate, but by considering how the notions of function and dysfunction are commonly deployed to help us understand health, disease, and mental disorder. We have argued that failure to recognize that functions are proximal threatens to undermine the coherence of otherwise valuable projects in the philosophy of medicine. We then showed how many cases one might be inclined to describe as "failure to perform a proximal function without dysfunction" are, in fact, not function failures at all but instead instances of a trait performing its proximal function just fine. We also argued that our thesis, which on the face of it appears to contravene some aspects of ordinary biological usage, can instead be understood as highlighting how function statements are sometimes elliptical. Finally, we defended our thesis against the charge that its apparent plausibility is only due to cherry-picking a set of cases where the context already implies a proximal-function reading.

References

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of science*, *44*(4), 542-573.

Buller, D. J. (1997). Individualism and evolutionary psychology (or: In defense of "narrow" functions). *Philosophy of Science*, *64*(1), 74-95.

Fagerberg, H. (2023). Medical Disorder Is Not a Black Box Essentialist Concept: Review of Defining Mental Disorder: Jerome Wakefield and His Critics, edited by Luc Faucher and Denis Forest. *Philosophy of Medicine*, *4*(1).

Garson, J. (2019). *What biological functions are and why they matter*. Cambridge University Press.

Godfrey-Smith, P. (1993). Functions: consensus without unity. *Pacific Philosophical Quarterly*, *74*(3)

Goode, R., & Griffiths, P. E. (1995). The misuse of Sober's selection for/selection of distinction. *Biology and Philosophy*, *10*, 99-108.

Griffiths, P. E., & Matthewson, J. (2018). Evolution, dysfunction, and disease: A reappraisal. *The British Journal for the Philosophy of Science*.

Griffiths, P. E., & Matthewson, J. (2020). Diseases are Not Adaptations and Neither are Their Causes: A Response to Ardern's" Dysfunction, Disease, and the Limits of Selection"(Biological Theory 13: 4–9, 2018). Biological Theory, 15(3), 136-142.

Matthewson, J., & Griffiths, P. E. (2017). Biological criteria of disease: Four ways of going wrong. *Journal of Medicine and Philosophy*, *42*(4), 447-466.

Millikan, R. G. (1984). Language, Thought, and Other Biological Categories. MIT Press.

Millikan, R. G. (1989a). In defense of proper functions. *Philosophy of science*, *56*(2), 288-302.

Millikan, R. G. (1989b). Biosemantics. *Journal of Philosophy*, *86*, 281-297.

Millikan, R. G. (2023). Teleosemantics and the frogs. *Mind & Language*. 1-9.

Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. *The British journal for the philosophy of science*. 60(4), 813-841.

Neander, K. (1983), "Abnormal Psychobiology". Ph.D. Dissertation, La Trobe University.

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense *Philosophy of Science*, *58*, 168-184..

Neander, K. (1995). Misrepresenting & malfunctioning. *Philosophical Studies, 79*, 109-141.

Neander, K. (2017). *A mark of the mental: In defense of informational teleosemantics*. MIT Press.

Papineau, D. (1998). Teleosemantics and indeterminacy. *Australasian Journal of Philosophy*, *76*(1), 1-14.

Wakefield, J. C. (1992). The concept of mental disorder: on the boundary between biological facts and social values. *American psychologist*, *47*(3), 373.

Wakefield, J. C. (2006). What makes a mental disorder mental?. *Philosophy, Psychiatry, & Psychology*, *13*(2), 123-131.

Wakefield, J. C. (2017). Addiction and the Concept of Disorder, part 2: Is every mental disorder a brain disorder?. *Neuroethics*, *10*, 55-67.

Wakefield, J. (2021). Biological Function Hierarchies and Indeterminacy of Dysfunction: Supplementary Reply to Justin Garson. In L. Faucher, D. Forest (Eds.), *Defining Mental Disorder: Jerome Wakefield and His Critics.* 385 – 396. MIT Press.