Philosophy of Science (2023), **90**, 686–703
doi:10.1017/psa.2023.30

ARTICLE

Dispositionalism and Dysfunction

Brandon A. Conley

Department of Philosophy and Religion, Clemson University, Clemson, SC, USA and Department of Philosophy, Washington and Lee University, Lexington, VA, USA
Email: bac248@cornell.edu

(Received 27 November 2021; revised 01 August 2022; accepted 01 February 2023; first published online 16 February 2023)

Abstract

My aim here is (1) to argue that the usual argument for thinking that dysfunction has no place in a dispositionalist approach to functions is deeply flawed and (2) to develop a positive account of the explanatory role dysfunction attributions play in dispositionalist-style functional analysis. I also argue that while my account undermines one common motivation for preferring an etiological over a dispositionalist approach, perhaps more interestingly, it also blurs the boundary between the two and opens a path to unifying them.

1. Introduction

Function attributions play an important explanatory role in the life, cognitive, medical, social, and engineering sciences; in the humanities; and in everyday discourse. They also appear to express that an item has a particular goal or purpose and is subject to a particular norm, namely, realizing its goal or purpose. An item that cannot perform its function—a heart that cannot pump blood, for example—is said to be dysfunctional. Because dominant conceptions of scientific explanation allow no role for the normative or teleological, functions have become a target for philosophical explication and an attractive tool for explicating other putatively normative but explanatory concepts. Most notably, the distinction between dysfunction and proper functioning has been used to explicate the nature of disease (Boorse 1977; Wakefield 1992) and how representations—mental, linguistic, and so on—can ever *mis*represent (Millikan 1984, 1989; Papineau 1987; Dretske 1988; Neander 1995, 2017; Maynard-Smith 2000; Shea 2018; Garson 2019).

Philosophical accounts of function are typically divided into one of two approaches: *etiological* approaches define something's function in terms of its history—usually, a history of design or natural selection—whereas *dispositionalist* approaches define something's function in terms of its contributions to the capacities and dispositions of an organized system in which it is embedded (Millikan 1989; Godfrey-Smith 1993; Amundson and Lauder 1994; Kingma 2020). There is near-universal consensus in the philosophical literature on functions that dispositionalist approaches cannot accommodate a notion of dysfunction, at least not without abandoning the usual way of understanding dysfunction as having a function but being unable to perform it, and without assigning dysfunction attributions little or no explanatory role. My aim here is (1) to argue that the usual argument for thinking that dysfunction has no place in a dispositionalist approach to functions is deeply flawed and (2) to develop a positive account of the explanatory role dysfunction attributions play in dispositionalist-style functional analysis. I also argue that while my account undermines one common motivation for preferring an etiological over a dispositionalist approach, perhaps more interestingly, it also blurs the boundary between the two and opens a path to unifying them.

2. Functional analysis

Robert Cummins (1975, 1983) argues that an item's function is its causal contribution to the workings of some system. We can explain a disposition or capacity of a system by decomposing it into simpler dispositions or capacities (subcapacities), organized in a particular way. This decomposition is a functional analysis, and the subcapacities are functions of the items that possess them. Cummins (1975, 762) formally defines what it is to possess a function as follows:

> *x* functions as a ϕ in [a system] *S* (or: the function of *x* in *S* is to ϕ) relative to an analytical account *A* of *S*'s capacity to ψ just in case x is capable of ϕ-ing in *S*, and *A* appropriately and adequately accounts for *S*'s capacity to ψ by, in part, appealing to the capacity of *x* to ϕ in *S*.

For example, the circulatory system (*S*) has the capacity (ψ) to transport materials like oxygen, waste, and hormones to different parts of the body. We can decompose this capacity (ψ) into several subcapacities, including pumping, directing flow, diffusing materials, and so on. The heart's capacity (ϕ) to pump blood contributes to the system's (*S*'s) capacity (ψ) to transport material. So, the function of the heart (*x*) in the circulatory system (*S*) is to pump blood (ϕ) because the heart is capable of pumping blood (ϕ-ing), and the standard physiological account (*A*) of how circulatory systems work adequately and appropriately accounts for the circulatory system's capacity (ψ) to transport material, in part, by appealing to the heart's (*x*'s) capacity (ϕ) to pump blood. I will say more about what makes an analysis "appropriate and adequate" in a later section.

Though Cummins frames his definition in terms of capacities, any dispositional property, including dispositions, tendencies, powers, proclivities, abilities, potentialities, and a host of others, can be explained by or cited in a functional analysis. I will move between talk of dispositions, capacities, and dispositional properties as context dictates, but nothing essential hangs on the subtle differences between them.

Dispositionalism is typically contrasted with etiological accounts of function that define function possession in terms of a historical property like having been designed or favored by natural selection. Not all accounts of functions fit neatly into the dispositionalist versus etiological dichotomy;[1] however, the distinction usefully tracks two

[1] Organizational accounts (Mossio, Saborido, and Moreno 2009) in particular are not easy to fit into this framework.

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

ways of thinking about the explanatory role of function attributions. While dispositionalists emphasize the role of function attributions in explanations of the workings of complex systems, etiological accounts are largely motivated by the observation that function attributions are often used to explain why the functionally characterized item exists (see esp. Wright 1973; Garson 2019). I am sympathetic to the idea that function attributions figure in the kind of "existence explanations" etiological accounts aim to capture, but the nature and cogency of these explanations are controversial (Davies 2001; Boorse 2002; Cummins 2002; Wouters 2005). In what follows, I will remain agnostic about existence explanations.

My primary aim is to undermine the view that dysfunction gives us a reason to think that Cummins's definition fails as a necessary condition for function possession and thus that we need a notion of function that does not contribute to the explanatory project Cummins describes. Cummins's definition has also been rejected by many as a sufficient condition for function possession because it is claimed to be too permissive—I will revisit this objection in more detail in a later section. However, restricting Cummins's definition by adding additional criteria for function possession, including etiological criteria, does not undermine Cummins's picture of function attributions' explanatory role. My primary aim here is to defend a dispositionalist picture of the explanatory role of function and dysfunction attributions, not to sort out the details of the extension of either concept. I thus use *dispositionalist* to refer broadly to any view that includes Cummins's definition as a necessary condition for function possession, including views that add additional criteria—the term is often reserved for Cummins's original view only.

For example, some have argued that functional analyses must pick out hierarchical mechanisms (Davies 2001; Craver 2001; Piccinini and Craver 2011; Craver and Darden 2013). Boorse (1977) has argued that we must define a cybernetic concept of a goal and that only contributions to goals, thus defined, count as genuine functions. Bigelow and Pargetter (1987) have argued that only contributions to survival and reproduction count as genuine functions. Insofar as these more restrictive views preserve Cummins's definition as a necessary condition and thus provide the same picture of functional explanation, they are dispositionalist views as I will use the term here.

3. A puzzle about dysfunction

According to the dispositionalist approach, a function is a capacity that contributes to some capacity of a system. Having a function implies having the corresponding capacity because an item's contribution to the system cannot be a capacity it does not have. Thus the first clause of Cummins's definition is that to possess the function of ϕ-ing, an item must "be capable of ϕ-ing." However, dysfunctional items are precisely those with a function they are not capable of performing. Ruth Millikan (1989, 294) has argued that

> the fact that we appeal to purposes and intentions when applying the term "function" results directly in ascriptions of functions to things that are not in fact capable of performing those functions; they neither function as nor have dispositions to function as anything in particular . . . . A diseased heart may not be capable of pumping, of functioning as a pump, although it is clearly its function, its biological purpose, to pump.

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

In other words, a dysfunctional item must have two characteristics: (1) it has the function of ϕ-ing and (2) it is incapable of ϕ-ing. Because being capable of ϕ-ing and incapable of ϕ-ing are contradictory states of affairs, Cummins's definition implies that something cannot satisfy both (1) and (2) at the same time, because his definition makes (1) imply the negation of (2). If Cummins has provided a necessary condition for what it is to have a function, there could be no such thing as being dysfunctional.

One obvious response is to note that many cases of dysfunction are a matter of performing a function less well rather than not at all. As we will see in the next section, a person with heart disease has a heart that pumps blood inefficiently rather than a heart that does not pump at all. This is right, and it will be important in my positive account of dysfunction, but this point alone does not circumvent the argument. We have only to be more precise about the capacity at issue. A heart's function is to pump blood, but more specifically, it is to pump at a particular efficiency. A failing heart is unable to pump at that efficiency.

It is difficult to overstate the extent to which this line of thought has shaped the current theoretical landscape of the function debate. It has been restated many times (e.g., Cummins 1975; Millikan 1989; Neander 1991; Davies 2001; Garson 2019), and it is not uncommon for the conclusion to be treated as an established result without bothering to restate the argument (e.g., Griffiths 1993; Walsh and Ariew 1996; Mossio, Saborido, and Moreno 2009; Kingma 2020). The continuing debate centers on what lesson we should glean from this failure.

It is relatively uncontroversial that function attributions do sometimes play an explanatory role like the one Cummins describes, and there have been two strategies to capture this while accounting for dysfunction. On one hand, fans of the dispositionalist approach (e.g., Davies 2001; Boorse 2002; Cummins 2002; Wouters 2005) have tended simply to deny that dysfunction attributions play an important explanatory role that the dispositionalist definition needs to capture. The usual strategy is to deny that dysfunctional items have the relevant function after all and to argue, instead, that we should think of dysfunction attributions as a way of indicating that something lacks a particular function we might care about. For example, Boorse (2002, 89) claims that a can opener that cannot open cans may have the typical or intended function of opening cans, but it does not have the function (without a qualifying adverb) of doing so. Davies (2001) argues that dysfunction attributions simply project our own desires and expectations onto our study systems and thus play no legitimate role in scientific explanations.

Fans of etiological accounts, on the other hand, typically argue that although Cummins has succeeded in explicating an important explanatory strategy in the sciences, he has failed to define function possession. His account, they say, is a precise description of what it is to *perform* a function. However, an item can accidentally perform a function it does not possess, and an item can have a function it is unable to perform. A laptop that slips out of one's bag might accidentally perform the function of a doorstop, and this may unfortunately lead to a malfunction whereby it is unable to perform many of the functions it actually possesses.

Note that these two strategies—(1) deny that dysfunctional items have the relevant function after all and (2) deny that Cummins has provided a necessary condition for function possession—are not logically bound to a particular view on existence explanations or the role of etiology in function attributions. One could adopt the usual

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

dispositionalist strategy but restrict function possession to items with a particular etiology (see, e.g., Kitcher 1993). On the other hand, it is not history per se that allows etiological accounts to avoid the puzzle about dysfunction facing dispositionalists but simply the denial that having the capacity to ϕ is a necessary condition for having the function of ϕ-ing. Nonetiological accounts can make the same move, though at the perhaps undesirable cost of introducing an asymmetry between the explanatory import of function attributions to dysfunctional versus properly functioning items. For example, Piccinini (2015, 109–10) has argued, like Boorse, that function possession is a matter of contributing to a goal but, unlike Boorse, argues that "the functions of malfunctioning tokens [of a given functional type] are grounded in part in the functions of well-functioning tokens." In Piccinini's view, function possession can consist in either satisfying a restricted version of Cummins's definition or, alternatively, standing in the right type–token relationship to other items that do. Because Piccinini allows that dysfunctional items possess functions without satisfying Cummins's definition, his account is not a dispositionalist account, but neither is it etiological.

According to both strategies I have outlined, we define a norm, a way that a given item might or might not contribute to the capacities of a larger system according to a Cummins-style functional analysis. Perhaps the norm is the way such items typically contribute to a system (Boorse 1977; Garson and Piccinini 2014), the way we would like or expect them to contribute to a system (Davies 2001), or the way counterparts contribute to the system in nearby possible worlds (Nanay 2010). Perhaps the normal contribution to a system is one that was intended by designers, human or divine (Plantinga 1993), or perhaps it is a contribution historically favored by natural selection (Millikan 1984; Griffiths 1993; Neander 1995; Buller 1998; Shea 2018) or some suitably similar kind of process (Dretske 1988; Garson 2019). Let us say an item has the Cummins-function of ϕ-ing if and only if it satisfies Cummins's definition for having the function of ϕ-ing. Behind the many opposing views in the literature about the nature of function and dysfunction, all sides agree that dysfunction consists in lacking a particular "normal" Cummins-function.

Disagreement about dysfunction concerns how to pick out the normal function, for example, using etiology or statistics, and whether we should take paradigmatic function attributions to refer to Cummins-functions or to normal functions. Either way, from the perspective of Cummins-style dispositionalist functional analysis, a dysfunction attribution indicates that a particular functional analysis simply does not apply. While a broken canary wing but not an ostrich wing has the normal function of facilitating flight, both of the extant strategies for capturing dysfunction say that this difference has no import for the explanatory project Cummins describes. All we can say is that neither wing contributes to a capacity to fly.

4. Making room for dispositionalist dysfunction

The argument outlined in the last section has been influential, accepted as obvious even by Cummins himself (Cummins 1975, note 13), but should we accept it? In this section, I will argue that we should not. It is possible for something to satisfy Cummins's definition of function possession and still be unable to perform its function. The problem with the argument to the contrary can be best appreciated by considering two clearly problematic arguments that rely on parallel reasoning.

Problematic argument 1: Congestive heart failure. Congestive heart failure is characterized by the inability of the heart to pump blood at the rate needed to sustain metabolizing tissues or the ability to do so only at an elevated filling pressure (Zipes et al. 2005). For simplicity, call the ratio of heart rate to filling pressure *efficiency*. A heart that can pump at the requisite efficiency, I will say, is capable of pumping efficiently. Thus a failing heart is one incapable of pumping efficiently. However, if a given failing heart is curable, it is possible to make it pump efficiently again. Because an item cannot be made to do something that it is not capable of doing, it must be that a curable heart is capable of pumping efficiently. Therefore a curable failing heart both is and is not capable of pumping efficiently, which is absurd, so the very idea of curing heart failure is incoherent.

Problematic argument 2: Carburetor troubles. Until the late 1980s, most car engines had carburetors, a part with the function of mixing air and fuel. If too much or too little fuel is in the mixture, a carburetor-equipped internal combustion engine will not run smoothly and may not run at all. A mechanic might identify a carburetor as dysfunctional if it is unable to mix air and fuel in a particular ratio. If it is possible to repair the carburetor, then that means the carburetor could be brought to mix air and fuel at the desired ratio, and that would imply that the carburetor is capable of mixing air and fuel at that ratio, because it is impossible for an item to do something that it is incapable of doing. Therefore a repairable dysfunctional carburetor both is and is not capable of mixing air and fuel at the desired ratio, which is absurd, so the very idea of a repairable dysfunctional carburetor is incoherent.

We would rightly scoff at these arguments. When a doctor says that a failing heart is curable, she means that it is capable of pumping efficiently given certain interventions like diet, exercise, medication, and surgery. However, when that same doctor diagnoses the heart as failing, the point is that the heart is incapable of pumping efficiently without those interventions. There is more than one capacity to pump efficiently at issue, and there is nothing absurd about the same heart having one and lacking the other. An exactly parallel story can be told about the mechanic and the carburetor. These examples point to a way not only to circumvent the argument of the last section but also to explicate how dysfunction attributions can be explanatorily useful.

Here is the general idea. The phrase "capable of ϕ-ing" underspecifies the capacity at issue because it leaves out the conditions under which the capacity would manifest, and the preceding examples show how this underspecification can lead to problems. The argument that dispositional conceptions of function render dysfunction attributions incoherent works only if we assume that the conditions under which an item must be capable of ϕ-ing to possess the Cummins-function of ϕ-ing are the same as the conditions under which it must be incapable of ϕ-ing to be dysfunctional. For an item's capacity to ϕ to be its contribution to the capacities of a system, that item must be capable of ϕ-ing. But being dysfunctional does not require that an item be incapable of ϕ-ing without qualification, only that it be incapable of ϕ-ing under some specified set of normal circumstances. Heart failure is classed as a dysfunction because hearts have the function of pumping efficiently, and a failing heart cannot.

However, this does not mean that a failing heart is incapable of ever, under any circumstances, pumping efficiently, only that it cannot do so without some special conditions, such as a regimen of diet, exercise, medication, and surgery.

Recognition that multiple capacities to ϕ are implicated in the context of a single functional explanation reveals where the explanatory difference between dysfunction and lack of function lies. That a failing heart is unable to pump efficiently under conditions in which a healthy heart could does not render the failing heart's capacity to pump irrelevant to the capacities of the circulatory system as a whole. Heart failure is a problem precisely because the capacity of the system as a whole to transport material through the body at sufficient efficiency to sustain metabolizing tissue depends on the heart's capacity to pump efficiently. The capacity of a circulatory system (*S*) to transport materials efficiently (ψ) is partially explained by the capacity of the heart (*x*) to pump efficiently (ϕ) for both failing and healthy hearts despite differences in the circumstances under which these capacities would manifest. The difference between lacking a function and being dysfunctional is whether the conditions under which an item is disposed to, or capable of, ϕ-ing makes a difference to the operations of the system as a whole. If the heart needs medication to pump at a given efficiency, then the system as a whole needs medication to move materials efficiently.

In short, function attributions express that there is an explanatory relation between the capacity of the functionally characterized item to ϕ, such as it is, and the capacity of a system to ψ, such as *it* is. Function attributions are neutral about which among a group of related capacities to ϕ the functionally characterized item has—that is, under which particular circumstances it would ϕ. Given the dependence of the system's capacity to ψ on the item's capacity to ϕ, it is useful to know under what conditions the item would ϕ, so it is useful to track which capacity to ϕ it has. Dysfunction attributions allow us to do exactly this.

I have offered what I take to be an intuitively appealing, albeit loose, characterization of how dysfunction attributions may not only make sense within a dispositionalist framework but actually contribute to the explanatory apparatus of dispositionalist functional analysis. Over the next four sections, I develop a model of how the explanatory strategy is supposed to work. Like any model, there is some sacrifice of nuance for the sake of precision. This move is justified because my aim is not to offer a full, extensionally adequate definition of function or dysfunction attributions but to show more precisely how the explanatory strategy I have roughly described can work. Once the model is on the table, I will address worries about the extension of the concepts of function and dysfunction.

5. Dispositional properties and conditionals

A functional analysis explains a disposition of a system by breaking down that disposition into subdispositions. To a first approximation, an item has a disposition (or capacity, or potentiality, or whatever) to ϕ just in case it would ϕ under a given set of circumstances. In the literature on the metaphysics and semantics of dispositions, it is controversial whether this simple formula is adequate, but the goal here is not to provide a metaphysical or semantic account of dispositions; rather, it is to show how functional analysis works, and for that purpose, the formula, and adaptations of some

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

associated terminology from that literature, will serve well (for a summary, see Choi and Fara 2021). I need not assume a conditional theory of dispositions, only that conditionals can be reliably associated with dispositions and thus serve as useful proxies for them.

Dispositions have three important structural features. A disposition has a *manifestation*, ϕ. The heart's disposition to pump blood is manifested in the heart's pumping blood, and solubility is manifested in the dissolution of the soluble substance. The disposition also has a set of *triggering conditions* under which the manifestation would occur. A heart pumps blood when there is blood present to be pumped, the brain is sending electrical impulses to the heart, it is at standard Earth temperature and pressure, and so on.

Finally, dispositions have a set of *basis conditions*, the conditions that must be satisfied for an item to possess the disposition and in virtue of which the triggering conditions give rise to the manifestation. Water-soluble materials have the disposition to dissolve in water. The triggering condition, being in water, leads to the manifestation, dissolving, because the soluble item has particular molecular properties like molecular polarity. Having a certain level of molecular polarity is thus a basis condition for being water-soluble. Possessing a particular contractile strength (especially relative to blood pressure) is a basis condition for a heart to have the capacity to pump efficiently given a particular set of triggering conditions.

Basis conditions need not be internal or intrinsic to an item. One basis condition for having the capacity to pass the salt is being located sufficiently close to the salt. Similarly, triggering conditions need not be external or extrinsic to an item. Most of us are disposed to seek food when we are hungry. Hunger is an internal triggering condition.

For every disposition, there is some corresponding (possibly very long and hard to generate) counterfactual conditional of the following form:

$$\text{Triggering conditions} \Rightarrow \text{Manifestation.}$$

The counterfactual relationship represented by the double arrow between the triggering conditions and manifestation can be explicated by spelling out the basis conditions, thereby generating a conditional of the following form:

$$(\text{Triggering conditions} \ \& \ \text{Basis conditions}) \rightarrow \text{Manifestation.}$$

The switch to a single arrow represents that once all of the basis conditions are included in the antecedent, the counterfactual conditional is in some important sense primitive. According to Cummins (1983, 2000), this happens when we reach regularities that can be subsumed under laws of nature. Some metaphysicians think dispositions must ultimately be explained by nondispositional, categorical properties, whereas others countenance primitive "bare dispositions" (Choi and Fara 2021). According to Craver and Piccinini (2011; see also Craver and Darden 2013), functional analyses provide a sketch of a hierarchical mechanism, and elsewhere, Craver has argued that specification of mechanisms "bottom out in lowest level mechanisms … [that are] accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group, or field" (Machamer, Darden,

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

and Craver 2000, 13). Whether the reasons are pragmatic or ontological, there is no infinite regress of analysis that might create problems for my picture of how functional analyses are explanatory or complicate the formal machinery.

6. How functional analyses explain

With the machinery developed in the last section, we can address an issue I set aside in section 2, namely, what makes a given functional analysis "appropriate and adequate." Cummins (1983) himself only tells us that the dispositions and capacities cited in an analysis must be organized in the right way and that this organization can be captured by a program or flowchart. Some philosophers have argued that the analysis must provide an outline or sketch of a mechanism in the sense developed in recent mechanistic philosophy of science (Davies 2001; Craver 2001; Piccinini and Craver 2011; Craver and Darden 2013). However, this claim is controversial (see, e.g., Weiskopf 2011), and I do not need to take a stand on this issue here. By using counterfactual conditionals of the form described in the last section as proxies for the dispositions that serve as both explanans and explananda in functional analyses, it is possible to provide a theoretically neutral account of what makes a functional analysis explanatory by focusing on the relationships that need to hold between these conditionals. I will focus on Cummins's claim that the relevant organization can be captured by a flowchart.

Corresponding to each disposition in a functional analysis, including both the disposition of the system we seek to explain and the subdispositions into which it is to be analyzed, is a conditional of the form "Triggering conditions ⇒ Manifestation." Listing these conditionals generates a partial flowchart-like structure with the conditional arrows linking triggering conditions to manifestations. Note two points.

First, the manifestation of many dispositions in an analysis will be identical to, cause, partially constitute, or otherwise explain the triggering conditions of other dispositions in the analysis. For example, the manifestation of the heart's capacity to pump blood is a triggering condition for the blood's capacity to move materials around the body. The manifestation of the blood's capacity to move materials is or causes the triggering conditions for the capillaries' capacity to diffuse material to and from cells.

Second, the triggering conditions and manifestations of some of the dispositions in an analysis are identical to, cause, partially constitute, or otherwise explain the triggering conditions or manifestation of the analyzed disposition. The presence of materials, for example, oxygen in the lungs, is a triggering condition for the capillaries to manifest their capacity to diffuse materials into the bloodstream, and it is also a triggering condition for the circulatory system as a whole to move materials around the body.

Given the list of conditionals representing dispositions in an analysis, adding arrows from listed conditions to others they explain, in the senses noted, completes the partial flowchart. Ideally, one can trace a path from the triggering conditions of the analyzed disposition to its manifestation. The arrows making up the path correspond to explanatory relations, so the flowchart specifies a string of explanatory connections in virtue of which the triggering conditions of the analyzed disposition

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

give rise to its manifestation. The flowchart thus constitutes a specification of the basis conditions of the disposition targeted by the analysis.

Note that the explanatory relations captured by the double arrows (⇒) connecting the triggering conditions to manifestations of subdispositions listed in the flow chart can themselves be explicated through functional analysis. The heart's capacity to pump blood can itself be functionally analyzed into the capacities of muscle cells, nerve fibers, and so on. The result can be embedded into the larger analysis of the circulatory system. In the limit, all of the basis conditions of a given system could be broken down until we have a maximally detailed analysis. Every counterfactual conditional would be of the primitive "(Triggering conditions & Basis conditions) → Manifestation" form. Though in practice, the details of the basis conditions for dispositions cited in a functional analysis may be unspecified, even unknown, they do form part of the explanatory chain linking the triggering conditions to the manifestation of the disposition targeted for analysis.

With this more detailed picture of how functional analyses work in hand, I now turn to arguing that if we have an adequate analysis of a system, we do not need a new analysis to account for a system in which some component process or part is dysfunctional. To make this case, I first need another piece of formal machinery, namely, a way of capturing what I referred to in section 4 as "related capacities."

7. Equivalence classes

Depending on the context, the same condition might be treated as a basis condition or as a triggering condition. Imagine two identical rockets rigged to launch when a corresponding red button is pressed; however, rocket 1 has fuel and rocket 2 has an empty tank; rocket 2 will be filled in the morning before launch. A supervisor inquires with one of the engineers about whether some recent maintenance is complete by asking whether rocket 2 is capable of launching. The engineer replies that it is—they have only to fill the tank and press the red button. Then, suddenly, a well-known supervillain arrives on the scene, declares that he has aimed a rocket at a nearby city, and proceeds to run for rocket 2's red button. "Thank goodness," the supervisor declares, "rocket 2 is incapable of launching because it has no fuel!" In the former case, the presence of fuel in the tank is treated as a triggering condition. In the latter, it is treated as a basis condition.

In one important sense, rocket 1 and rocket 2 have different capacities. After all, one but not the other would launch if its red button were pressed. However, there is another very important sense in which the two rockets have the same capacities: they are identically designed, and both require both fuel and the press of a red button to launch. On one way of looking at things, a particular triggering condition, the presence of fuel, for the same capacity has already been triggered for one rocket but not yet in the other. On another way of looking at things, rocket 2 lacks a capacity that rocket 1 possesses, namely, the capacity to launch when the red button is pressed, but it could gain that capacity by having its tank filled, thereby bringing it to satisfy the final basis condition required to possess that capacity.

The kind of relationship that holds between rocket 1's and rocket 2's capacities can be modeled using equivalence classes. For any two dispositions D_1 and D_2, D_1 is equivalent to (or in the same equivalence class as) D_2 if and only if the union of the sets of basis conditions and triggering conditions for D_1 is identical to the union of the sets of basis conditions and triggering conditions for D_2 and the manifestations of D_1 and D_2 are the same. Note that whether two capacities are in an equivalence class is relative to a list of conditions, so in modeling real cases, care must be taken to specify the right list.

Medication and surgery are not among the basis or triggering conditions for the capacity of a healthy heart to pump efficiently, but they may be for a failing heart. It thus seems the two capacities are not in an equivalence class. However, these interventions on a failing heart aim to restore a particular ratio of heart strength to blood pressure, which is a basis condition for the capacity of healthy hearts to pump efficiently. If we treat the ratio of heart strength to blood pressure as the triggering condition for failing hearts to pump efficiently and think of diet, medication, and surgery as ways of triggering it, but not themselves among the conditions on the list, we can generate a model that captures the relationship between the capacities of failing and healthy hearts. The capacity to pump efficiently that a healthy heart possesses, and a failing heart lacks, has as a basis condition that the ratio of heart strength to blood pressure be in a given range. Moving that condition from the set of basis conditions to the set of triggering conditions generates a capacity in the same equivalence class, one that a failing heart does have.

Scientists move fluidly between treating a given condition as a triggering or as a basis condition, as when a doctor moves from diagnosing heart failure to discussing treatment. In the diagnostic context, that the ratio of heart strength to blood pressure is within a specified range is treated as a basis condition, either satisfied or not, but once it is determined that this condition is not satisfied, the relevant question becomes how best to satisfy it. In the new context of treatment, the ratio of heart strength to blood pressure is treated as a triggering condition, as something to be manipulated. This shift alters which disposition within an equivalence class is contextually salient and renders it perfectly coherent to say in one context that an item lacks the capacity to ϕ and in another that it has the capacity to ϕ. However, because the two capacities to ϕ are in an important sense equivalent, there may also be contexts in which it makes sense to ambiguously refer to both as simply the capacity to ϕ. I will argue that this is typically what happens in a functional analysis. Put another way, functional analyses are really about equivalence classes of capacities.

8. Dysfunction in functional analysis

Imagine that we have a functional analysis that explains the capacity of a circulatory system (S) to move materials around efficiently, to which a heart (h) contributes by pumping at a given efficiency. What would it take to apply this analysis to another system (S^*) in which the heart (h^*) cannot pump at that efficiency?

The functional analysis of S's capacity to move material around the body efficiently shows that one basis condition of that capacity is h's capacity to pump efficiently; h's capacity forms part of a chain of explanatory relationships linking the conjunction of a list of triggering conditions, call it T_S, to S's moving materials around efficiently. Heart h^* lacks h's capacity to pump efficiently, but the argument of the last section established that h^* does have a different capacity to pump efficiently that is in an equivalence class with h's capacity to do so. We generated this capacity simply by

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

treating the relevant basis condition, in this case, a particular ratio between heart strength and blood pressure, as a triggering condition instead. If T_h is the conjunction of triggering conditions for h's capacity to pump efficiently and C is the condition that heart strength to blood pressure have a particular ratio, then h^* has the capacity to pump efficiently given T_h and C.

Now imagine that C comes to be satisfied for h^*; h^* thus acquires the capacity to pump that h already has. In virtue of this, S^* acquires the same capacity to move materials that S has because the basis condition S^* but not S failed to satisfy was precisely that the heart in the system have the capacity to pump efficiently. It follows that if h^* has the capacity to pump efficiently given T_h and C, then by virtue of that very fact, S^* has the capacity to move materials efficiently given T_S and C.

Put another way, our functional analysis shows that the reason S^* lacks the capacity S has to move materials efficiently is that the explanatory chain, the path in the flowchart, linking T_S to moving materials efficiently is broken by h^*'s inability to pump efficiently given T_h. Because h^* has the capacity to pump efficiently given T_h and C, bringing it about that C is satisfied would give h^* the capacity to pump efficiently given T_h, thereby completing the explanatory chain, that is, the path in the flowchart. But this is just to say that treating C as a triggering condition, rather than as a basis condition, shows how S^*'s capacity to move material is explained by h^*'s capacity to pump efficiently, such as they are, in exactly the same way h's corresponding capacity explains S's corresponding capacity. Importantly, generating this result required no changes to the functional analysis itself, only that a particular condition in the explanatory chain be labeled differently.

The lesson is that functional analyses do not explain just one disposition in an equivalence class; they explain the entire equivalence class. If this is so, function attributions should be expected not to fully specify triggering conditions, because otherwise, their explanatory scope would be artificially narrowed. To attribute to hearts the function of pumping efficiently is to say that a heart's capacity to pump blood efficiently, under whatever conditions it is capable of doing so, explains the containing circulatory system's capacity to move materials effectively under corresponding conditions.

Just as we can describe the location of a point in space using a reference point and a unit of distance, we can locate an item's capacity within an equivalence class relative to a privileged capacity. Dysfunction attributions do just that. We can model the way this works by defining an ordering on the equivalence class. Dispositions D_1 and D_2 in equivalence class $[D]$ are such that $D_1 < D_2$ if and only if the set of triggering conditions for D_1 is a superset of the set of triggering conditions for D_2. Let D^* be the privileged capacity in $[D]$. Then, an item is dysfunctional if it possesses the function of ϕ-ing and its greatest capacity to ϕ in $[D]$ is less than D^*. According to this model, a failing heart is dysfunctional because the triggering conditions for its capacity to ϕ include the set of normal circumstances—the triggering conditions for the privileged capacity—plus the condition that the ratio of heart strength to blood pressure be within a particular range, which is a basis condition for the privileged capacity. The intuitive idea behind this definition is that one must satisfy more conditions, or do more work, to get a dysfunctional item to perform its function. I will not take a position here on how to determine the privileged capacity. Several options were surveyed in section 3, including statistics, evolutionary history, designer intentions, human attitudes, and features of nearby possible worlds. And using different options in different contexts is a live option.

That a failing heart cannot pump efficiently under normal circumstances, and therefore lacks the privileged capacity, whatever it is, is interesting precisely because the heart has the Cummins-function of pumping efficiently. If it did not, it would make no difference to the larger system whether the heart could pump efficiently under a particular set of circumstances. Given that hearts, even dysfunctional hearts, do have the Cummins-function of pumping efficiently, it is useful to know the conditions under which specific hearts would pump efficiently. For example, this knowledge allows for manipulations of a whole system through manipulations of those conditions, in this case, through medication, exercise, diet, and surgery. Dysfunction attributions thus provide more fine-grained explanatorily relevant information about dispositions than function attributions alone.

9. The extension of "function" and "dysfunction"

Philosophers will be immediately tempted to consider various counterexamples, but doing so largely misses the point of this account. I do not attempt here to define the extension of function or dysfunction attributions. Rather, assuming a particular account of functional explanation, I argue that dysfunction attributions contribute to that project. We can toy with the boundaries of the concept without losing anything essential about the explanatory project I have outlined.

However, this does not mean that all worries about the extension of function or dysfunction attributions can be set aside. If the account I have offered excludes important classes of dysfunction attributions, then I may not have captured their explanatory role. This is why I have developed the model to be extremely flexible. By rigging the choice of privileged capacities, the list of conditions used to define equivalence classes, and other parameters, I believe it will be possible to incorporate just about any putative case of dysfunction the account might be thought to exclude. Where this fails, I think it likely that adding complexity to the model using more sophisticated formal tools would do the job. For example, we might use causal graphical models or information theory to capture the ways conditions correlate with one another, rather than assuming that they vary independently.

Also note that although I have used examples where the triggering conditions are realistically manipulable because I think such examples are the most clear and compelling, nothing in the model requires this. Dysfunction, on this model, is not relative to the actual potential for repair or reversal. Equivalence classes of dispositions still make sense if we cannot trigger some of the unsatisfied conditions in practice, and the explanatory relation between the analyzed disposition and subdispositions in the analysis can still be present.

It is also possible to capture the sense in which specific triggering conditions are treated as constitutive of a function. While we sometimes talk about the function of ϕ-ing without specifying any specific triggering conditions, we also talk about the function of ϕ-ing *when C*. The function of a smoke alarm is to sound when there is smoke present. This kind of attribution can be modeled by putting the entire conditional "if there is smoke present, then sound" into the consequent of the conditional, thereby treating the entire conditional as a manifestation of the disposition of

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

interest. This is formally equivalent to including the condition among the set of triggering conditions, because in general, $(C_1 \& C_2 \ldots \& C_n) \Rightarrow m$ is equivalent to $(C_1 \& C_2 \ldots \& C_{n-k}) \Rightarrow ((C_n \& C_{n-1} \ldots \& C_{n-k+1}) \Rightarrow m)$. Putting conditions into our specification of a disposition's manifestation changes nothing in the basic working of the model, though it does change a condition's relative salience and how it interacts with the specification of equivalence classes.

This extreme flexibility leads to another, opposite worry about the extension of function and dysfunction attributions. There is a risk of diluting the concepts to the point of triviality, and it appears that the price of the flexibility is that practically anything can count as dysfunctional. This issue arises because I have used Cummins's definition as a basis for my account, and it is commonly objected that Cummins's definition is overly permissive (Mossio, Saborido, and Moreno 2009; Garson 2016). For example, Millikan (1989) has argued that Cummins's account says rain clouds have the function of watering crops because the tendency for rain to fall contributes to the capacity of crops to grow. It appears to follow that my account implies that rain clouds that fail to drop rain on our fields might be dysfunctional, and this does not accord with ordinary usage or scientific practice.

It is worth emphasizing that I am defending Cummins's definition only as a necessary condition for function possession. I remain agnostic about whether his definition can also serve as a sufficient condition for function possession, and if not, what additional criteria would be needed. Depending on one's views on the permissiveness problem for Cummins, the implications for my account will be different. Following are a few options.

Option 1: Embrace permissiveness. One strategy is simply to accept a wildly permissive account of functions and a correspondingly permissive account of dysfunction. Weber (2017) defends a view along these lines according to which there are many more objective functions in the world than we typically countenance, but only some of them are scientifically interesting or useful to talk about. Counterintuitive function attributions, like the function of rain to water crops, that seem to follow from Cummins's definition are not mistaken, only unfamiliar and useless. If this strategy succeeds, then the same pragmatic reasoning applies to dysfunction attributions and the permissiveness of my account poses no special problem.

Option 2: Add restrictions. Although we can apply Cummins's definition to any dispositional property that we happen to be interested in analyzing, it is arguable that some systems have specific kinds of capacities or dispositions that are intrinsically interesting independent of our particular interest in them. I have already discussed Bigelow and Pargetter's (1987) view that only contributions to survival and reproduction count as genuine functions and Boorse's (1977) view that only contributions to a goal count as genuine functions. An item that cannot ϕ under some particular set of conditions only counts as dysfunctional if it has the function of ϕ-ing, so my account of dysfunction would inherit any restrictions on Cummins's account. If only contributions to dispositions that count as goals are genuinely functions, then only failures to contribute to goals under whatever the corresponding goal-contribution account determines to be the privileged set of circumstances will count as dysfunctional. My model of dysfunction's role in functional analysis is only as permissive as the definition of function fed into it.

Option 3: Embrace etiology. Though etiological accounts are typically contrasted with dispositionalist accounts, the boundary blurs once dysfunction is integrated into the dispositionalist framework. Etiological theorists have denied that Cummins's definition could be a necessary condition for function possession because it would exclude dysfunctional items from having the relevant function. With the usual argument against dispositionalist dysfunction undermined, etiological accounts emerge as fully compatible with dispositionalism. We can simply treat a history of design or natural selection as a restriction on Cummins's definition without excluding dysfunctional items as genuine function bearers. My account of dysfunction would inherit this restriction just as it would on option 2, above. The model of dysfunction, again, is no more permissive than the definition of function fed into it, and that includes an etiological definition.

A second kind of permissiveness worry is that my account cannot differentiate failure to perform a privileged capacity due to dysfunction from failure to do so for some other reason. For example, a stomach may fail to perform its digestive function due to poisoning or due to a normal response to strenuous exercise, but only the former is dysfunctional (Kingma 2010, 2016; Piccinini 2020). Here it is relevant again that dysfunction attributions in the model are relative to a list of conditions, and I have not placed any particular restrictions on those conditions. The model can capture the difference between exercise and poison by careful choice of the conditions included in the sets used to define equivalence classes and the associated ordering. Among the triggering conditions of the privileged capacity of the stomach to digest we include that the stomach's owner is not currently undergoing strenuous exercise. A stomach that ceases digestion due to healthy strenuous exercise is not dysfunctional because it is simply not in the appropriate conditions to perform the function.

The response is not ad hoc because the parameters of the model are inputs, not outputs, to my account. How the various parameters, including the list of conditions that define equivalence classes and the privileged capacity, are or are not constrained will depend on the details of the definition of functions fed into it. All that is required for the model to apply is that we assume that Cummins's definition is a necessary condition. If we assume that Cummins has also provided a sufficient condition, then determination of the parameters will be merely pragmatic. Well-motivated restrictions on Cummins's definition will introduce new assumptions about the explanatory role of functions and the nature of things that possess them, which will in turn constrain the parameters of the model.

It is also worth noting that my account allows for a principled kind of pluralism. It may be that different projects or domains of inquiry restrict the general dispositionalist formula in different ways. On this kind of view, the parameters of the model are not determined merely by pragmatic concerns, but neither is there a unique set of sufficient conditions for function possession applicable to all explanatory uses of function attributions.

https://doi.org/10.1017/psa.2023.30 Published online by Cambridge University Press

One reader has objected that my account would objectionably countenance cases of negative causation because dysfunction consists in an item not doing something, yet I have claimed that dysfunction can contribute to a positive explanation of a system's capacities, as opposed to simply explaining why it lacks certain capacities. However, my account implies that being dysfunctional is not merely a negative property; being dysfunctional consists, in part, in positively possessing a disposition or capacity in a particular class of related capacities, and that capacity must partially explain the capacities of the whole system. Perhaps negative causation would have to be implicated in particular examples, but this is difficult to assess without concrete putative cases.

If there are some examples in which we need negative causation to make sense of dysfunction, I am inclined to say that this constitutes an argument for negative causation rather than an argument against my account, because dysfunction appears to play a role in positive explanations just as this account says it can. A patient experiences fatigue when climbing stairs because she has a failing (i.e., dysfunctional) heart. The car did not start because the ignition malfunctioned. Cancer can be caused by a dysfunction in the gene-copying machinery of the cell. Biologists define animal signals as behaviors with the function of carrying information, and the failure of some signals to perform their information-carrying function is central to work in animal communication because the central question is how signals can evolve to be reliable. The most straightforward description of knockout experiments in neuroscience and genetics is that they rely on making causal inferences about a given gene or brain region by experimentally probing the effects of rendering it dysfunctional. There are also functions that arguably presuppose a coherent notion of dysfunction. A backup system's function is to engage when the primary system fails to perform its function. Attributions of functions of error detection, repair, or redundancy all presuppose that one item's function can be to deal with dysfunction elsewhere in the system. If a dysfunction has effects, then it must be causally relevant. If that relevance can be cashed out only in terms of negative causation, I doubt, given its flexibility, that my account will be the theoretical assumption that forces our hand. I do not think worries about negative causation cast serious doubt on my account. That the account can capture the sense in which dysfunction attributions contribute to positive explanations, however, is a mark in its favor.

Acknowledgments. I dedicate this article to the memory of Richard N. Boyd, who passed away in February 2020, after having advised the dissertation work and early drafts on which this article is based. I thank Derk Pereboom, W. Starr, Harold Hodes, Shane Glackin, and Antoine Dussault for helpful comments on earlier drafts. Special thanks go to Eric Epstein for insights about formulating some of the formal machinery of the article.

References

Amundson, Ron, and George Lauder. 1994. "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology." *Biology and Philosophy* 9:443–69.

Bigelow, John, and Robert Pargetter. 1987. "Functions." *Journal of Philosophy* 84 (4):181–96.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44 (4):542–73.

Boorse, Christopher. 2002. "A Rebuttal on Functions." In *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by Andrew Ariew, Robert Cummins, and Mark Perlman, 63–112. Oxford: Oxford University Press.

Buller, David J. 1998. "Etiological Theories of Function: A Geographical Survey." *Biology and Philosophy* 13:505–27.

Choi, Sungho, and Michael Fara. 2021. "Dispositions." In *The Stanford Encyclopedia of Philosophy*, Spring 2021 ed., edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2021/entries/dispositions/.

Craver, Carl F. 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science* 68 (1):53–74.

Craver, Carl F., and Lindley Darden. 2013. *In Search of Mechanisms: Discovery across the Life Sciences*. Chicago: University of Chicago Press.

Craver, Carl and Gualtiero Piccinini. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183:283–311.

Cummins, Robert. 1975. "Functional Analysis." *Journal of Philosophy* 72 (20):741–65.

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

Cummins, Robert. 2000. "'How Does It Work' versus 'What Are the Laws?': Two Conceptions of Psychological Explanation." In *Explanation and Cognition*, edited by F. Keil and Robert A. Wilson, 117–45. Cambridge, MA: MIT Press.

Cummins, Robert. 2002. "Neo-teleology." In *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by Andrew Ariew, Robert Cummins, and Mark Perlman, 147–72. Oxford: Oxford University Press.

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

Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Garson, Justin. 2016. *A Critical Overview of Biological Functions*. Berlin: Springer.

Garson, Justin. 2019. *What Biological Functions Are and Why They Matter*. New York: Cambridge University Press.

Garson, Justin, and Gaultiero Piccinini. 2014. "Functions Must Be Performed at Appropriate Rates in Appropriate Situations." *British Journal for the Philosophy of Science* 65 (1):1–20.

Godfrey-Smith, Peter. 1993. "Functions: Consensus without Unity." *Pacific Philosophical Quarterly* 74 (3):196–208.

Griffiths, Paul E. 1993. "Functional Analysis and Proper Functions." *British Journal for the Philosophy of Science* 44 (3):409–22.

Kingma, Elselijn. 2010. "Paracetamol, Poison, and Polio: Why Boorse's Account of Function Fails to Distinguish Health and Disease." *British Journal for the Philosophy of Science* 61 (2):241–64.

Kingma, Elselijn. 2016. "Situation-Specific Disease and Dispositional Function." *British Journal for the Philosophy of Science* 67 (2):391–404.

Kingma, Elselijn. 2020. "Function and Health at the Interface of Biology and Technology." *Noûs* 54 (1):182–203.

Kitcher, Philip. 1993. "Function and Design." *Midwest Studies in Philosophy* 18:379–97.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1):1–25.

Maynard-Smith, John. 2000. "The Concept of Information in Biology." *Philosophy of Science* 67 (2):177–94.

Millikan, Ruth G. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Millikan, Ruth G. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56 (2):288–302.

Mossio, Matteo, Cristian Saborido, and Alvaro Moreno. 2009. "An Organizational Account of Biological Functions." *British Journal for the Philosophy of Science* 60 (4):814–31.

Nanay, Bence. 2010. "A Modal Theory of Function." *Journal of Philosophy* 107 (8):412–31.

Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* 58 (2):168–84.

Neander, Karen. 1995. "Misrepresenting and Malfunctioning." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 79:109–41.

Neander, Karen. 2017. *A Mark of the Mental: A Defense of Informational Teleosemantics*. Cambridge, MA: MIT Press.

Papineau, David. 1987. *Reality and Representation*. Oxford: Basil Blackwell.

Piccinini, Gualtiero. 2015. *Physical Computation: A Mechanistic Account*. Oxford: Oxford University Press.

Piccinini, Gualtiero. 2020. *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford: Oxford University Press.

Piccinini, Gualtiero, and Carl F. Craver. 2011. "Integrating Psychology and Neuroscience: Functional Analyses as Mechanisms Sketches." *Synthese* 183:283–311.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.

Shea, Nicholas. 2018. *Representation in Cognitive Science*. Oxford: Oxford University Press.

Wakefield, Jerome C. 1992. "The Concept of Mental Disorder." *American Psychologist* 47 (3):373–88.

Walsh, Denis M. and André Ariew. 1996. "A Taxonomy of Functions." *Canadian Journal of Philosophy* 26(4):493–514.

Weber, Marcel. 2017. "How Objective Are Biological Functions?" *Synthese* 194:4741–55.

Weiskopf, Daniel. 2011. "Models and Mechanisms in Psychological Explanation." *Synthese* 183:313–38.

Wouters, Arno. 2005. "The Function Debate in Philosophy." *Acta Biotheoretica* 53:123–51.

Wright, Larry. 1973. "Functions." *Philosophical Review* 82 (2):139–68.

Zipes, Douglas P., Peter Libby, Robert O. Bonow, and Eugene Braunwald. 2005. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*, 7th ed., vol. 1. Philadelphia: Elsevier.

Cite this article: Conley, Brandon A. 2023. "Dispositionalism and Dysfunction." *Philosophy of Science* 90 (3): 686–703. https://doi.org/10.1017/psa.2023.30