Defining Mental Disorder

Jerome Wakefield and His Critics

The title-level DOI for this work is:

doi:10.7551/mitpress/9949.001.0001

14 Can Causal Role Functions Yield Objective Judgments of Medical Dysfunction and Replace the Harmful Dysfunction Analysis's Evolutionary Component? Reply to Dominic Murphy

Jerome Wakefield

I thank Dominic Murphy for his contribution to this volume and for his philosophically sophisticated writings over the years, including his illuminating *Psychiatry in the Scientific Image* (2006). In his present chapter, Murphy challenges the evolutionary component of my harmful dysfunction analysis (HDA) of medical, including mental, disorder. The HDA claims that "disorder" refers to "harmful dysfunction," where dysfunction is the failure of some feature to perform a natural function for which it is biologically designed by evolutionary processes and harm is judged in accordance with social values (First and Wakefield 2010, 2013; Spitzer 1997, 1999; Wakefield 1992a, 1992b, 1993, 1995, 1997a, 1997b, 1997c, 1997d, 1998, 1999a, 1999b, 1999c, 2000a, 2000b, 2001, 2006, 2007, 2009, 2011a, 2014, 2016a, 2016b; Wakefield and First 2003, 2012).

Murphy has long been a critic of my claim that the sense of "dysfunction" that is most relevant to understanding psychiatric attributions of disorder and nondisorder is failure of natural (or biologically designed) function, which cashes out in evolutionary terms. He frames his argument as a contest between the HDA's "selected effects" (SE) approach to "function" and Robert Cummins's "systemic capacity" or "functional analysis" or "causal role" (CR) understanding of "function" that Murphy embraces. My blurry recollection is that this dispute has been going on ever since our first (and I think only) face-to-face meeting many years ago when he was a graduate student and I a guest speaker in Stephen Stich's seminar at Rutgers. I have addressed this issue elsewhere (e.g., Wakefield 2000b, 2001), but I am happy to have this opportunity to tackle this central disagreement between us from the fresh perspective of his chapter, especially knowing that many others in our field share Murphy's preference for the causal role approach.

Murphy finds the CR function approach so persuasive that he suggests that my choice of an evolutionary approach must be due either to my missing nuances or not considering other options in the function literature. In fact, it was immersion in the function literature that persuaded me to reject Cummins's CR-function account and construct my own "black-box essentialist" evolutionary version of the nature of natural or biological function concepts. A central impetus for my formulating the HDA was my

rejection of Boorse's (1976) critique of Larry Wright's (1973) seminal work on the etiological approach to function when applied to the specific case of biological functions. I concluded that Boorse's critique was flawed and left untouched a correct core idea in Wright's analysis when applied specifically to biological or natural functions and that this core idea is essential to the foundations of medicine. Another formative influence was a graduate seminar I sat in on at Berkeley with Charles Taylor—if memory serves, with Mark Bedau and Robert Wachbroit also attending, both of whom later published on functions—in which issues concerning function raised by Taylor's book *The Explanation of Behavior* (1964) were discussed. An additional provocation was the controversy about the evolutionary and function-related issues in regard to female orgasmic dysfunction raised by Don Symonds's (1979) book *The Evolution of Human Sexuality* (later examined in depth by Elizabeth Lloyd [2005]) as well as by research on orgasmic dysfunction that I was pursuing at the time with my colleague, Nadine Payn (Payn and Wakefield 1982) (for comments on my early papers on sexuality, see Demazeux's chapter in this volume).

So, if I have gotten "dysfunction" in its diagnostic sense wrong—and I don't think I have—it is not because I have missed or failed to consider various options regarding how to construe biological functions. It is because I think there are powerful arguments that other views are inadequate to provide a foundation for psychiatric and medical diagnosis. However, Murphy's pushing me on this point has usefully prodded me to address this topic in greater depth and thoroughness than I have before.

Murphy tends to emphasize a sharp divide between folk concepts and scientific concepts, and this has been the basis for another criticism by him, that in my criticisms of false-positive psychiatric diagnostic criteria, I give priority to folk psychology over science. Although I remain unclear about the precise basis for this claim, I am happy to see that in the present chapter, he adopts an alternative interpretation: "A more charitable reading would be that Wakefield thinks that the underlying logic of psychiatric practice accords with the HDA but that sometimes psychiatrists depart from the logic of their own projects." That is correct; my critiques of psychiatric overdiagnosis are based on inconsistencies between psychiatry's classification of certain conditions as disorders and what I take to be psychiatry's own HDA concept of disorder. There is in my view no in-principle tension between common sense and science and certainly no dominance of folk theories over science. On my view, the intuitive concept of disorder and the scientific concept are basically the same with science's essentialist theory added, so there is no in-principle conflict at that level. And, common sense succumbs to science on the facts that determine whether specific conditions satisfy the conceptual and essentialist criteria and thus warrant disorder attribution. In any event, Murphy's more nuanced view of my position allows me to put aside this previous misunderstanding and focus on what is now the central issue between us, the nature of function and dysfunction in the medically relevant sense that pertains to disorder attributions.

The Ubiquity of CR Functions and the Reality of SE Functions

It is important to be crystal clear at the outset that I completely accept with Murphy that non-SE functional language is of course an integral part of biology's vocabulary. It is extremely common in biology to distinguish structure from function (which is a parallel task to distinguishing anatomy from physiology), with "function" most often simply meaning causal action. Indeed, an about-to-be-issued new journal in physiology sponsored by the American Physiological Association will be titled *Function* and almost certainly won't be limited to biologically designed causal roles. Throughout the contemporary biological sciences, and historically from William Harvey to Bock and von Wahlert (1965; see below) and on to Murphy's example of Hubel and Wiesel, "function" has been used by researchers to specify the "causal action" or "causal role" of a feature within a larger system being studied.

It is true, as SE theorists have often pointed out, that the language of function as causal action is often weaker in syntactic structure than the characteristic SE-function assertion such as "the natural (or biological) function of X is to Y" and may be more like "X functions to Y," "X functions as a Y," "X functionally interacts with Y," "the functional effect of X is Y," "X divides into functional units of type Y," and so on. For example, if one examines Hubel and Wiesel's (1959, 1962) classic work on the workings of the visual perceptual system of the cat cited by Murphy as an example of CR functions, it is true that they are primarily trying to understand the details of causal interactions and are not explicitly concerned with background assumptions about the biological design of the visual system, although such assumptions guide the kinds of hypotheses they formulate. Consequently, they almost never use "function" in the strong sense of "the function of" of an anatomical feature. Rather, consistent with their focus on causal role, their uses of "function" generally occur in adjectival and adverbial forms such as "functional architecture," "functionally different," "functional cell types," "functionally separated," "functional description," "functional subdivisions," "functionally disrupt," "functional role," "functional connexions," and "functional units." Only once do they step back and look at the big picture of the meaning of their detailed causal-action discoveries for overall biological roles in perceptual discrimination, and there they use "function" in a noun form that is at least ambiguous between CR and SE function: "Our findings in the striate cortex would suggest two further possible functions … to determine the form, size and orientation of the most effective stimuli, and secondly, … perception of movement" (1959, 588). So, Murphy's example supports his point that "function" and its cognates are used in CR-type ways. One need not go back to Hubel and Wiesel to nail down this point; surveying recent biological literature quickly proves the point. Indeed, I would argue that most theoretical terms have several or many meanings and can be used fluidly with different "semantic markers" (in Putnam's terminology; see my reply to Lemoine in this volume) in a multiplicity of ways.

Locutional nuances are not always a reliable indicator of meaning, and even the strong locution, "the function of X," is ambiguous and can be used in a CR-function context as a variant of "the functioning of X" even though it generally refers to SE functions. "Function" sometimes is used this CR way when describing pathological causal actions, as in the article title, "Retinoid X Receptors: X-ploring Their (Patho) physiological Functions" (Szanto et al. 2004). Similarly, in an editorial commenting on Engle et al.'s (2019) discovery that glycan carbohydrate antigen CA19–9, a known biomarker for pancreatic cancer, actually plays a pathogenic causal role in the disease, Halbrook and Crawford (2019) say that the Engle et al. study "ascribes a critical function to the most commonly used biomarker of the disease" (1132), that "there has been little understanding of CA19–9 function in pancreatic pathophysiology" (1132), and "the discovery of a new function of CA19–9 is exciting" (1133). Sometimes, CR and SE uses of "function" appear side by side. Thus, the same article that refers to "erectile dysfunction" with a clear implication of objective failure of natural function also refers to the effects of genetic variation on "sexual function" in the CR sense that encompasses causation of disorder: "Because the variants associated with erectile dysfunction are not associated with differences in BMI, our findings suggest a mechanism that is specific to sexual function" (Jorgenson et al. 2018, 11018), and the same article uses "function" to report the identification of the biologically designed SE-function role of a gene: "Finally, through in silico and in vitro functional investigations, we linked our risk locus to gene function. As evolutionary conservation is a strong marker of functional genomic sequences, we focused our follow-up analyses on … the only SNP located in an evolutionarily conserved region" (Jorgenson et al. 2018, 11019).

Although the focus on CR analysis often without explicit reference to biological design is ubiquitous in biological research, the depth of this separation can easily be overstated. Design considerations are almost always at least implicitly or potentially in the CR-analytic background in studies of normal functioning. Papers in functional anatomy almost always assume that salient anatomical structures have adaptive biological roles (even if the study examines a pathogenic causal role), and CR-type analyses are generally guided by implicit biological-design hunches. Analysis of functional mechanics is generally accompanied or followed by hypotheses about design function that guide further research. For example, consider these excerpts from article abstracts:

> In 1678, Stefano Lorenzini first described a network of organs of unknown function in the torpedo ray-the ampullae of Lorenzini (AoL). An individual ampulla consists of a pore on the skin that is open to the environment, a canal containing a jelly and leading to an alveolus with a series of electrosensing cells. The role of the AoL remained a mystery for almost 300 years until research demonstrated that skates, sharks, and rays detect very weak electric fields produced by a potential prey. The AoL jelly likely contributes to this electrosensing function, yet the exact details of this contribution remain unclear. … We hope that the observed high

> proton conductivity of the AoL jelly may contribute to future studies of the AoL function. (Josberger et al. 2016, 1)
>
> We report on the discovery of a remarkable defensive specialization in stonefishes that was identified during a phylogenetic study of scorpionfishes. … The lachrymal saber, involves modifications to the circumorbitals, maxilla, *adductor mandibulae*, and associated tendons. At its core, the lachrymal saber is an elongation of an anterior spine … that stonefishes are capable of rotating from the standard ventral position to a locked lateral position. … that we hypothesize reduces predation on stonefishes. (Smith, Everman, and Richardson 2018, 94)
>
> Upon continued submersion in water, the glabrous skin on human hands and feet forms wrinkles. The formation of these wrinkles is known to be an active process, controlled by the autonomic nervous system. Such an active control suggests that these wrinkles may have an important function, but this function has not been clear. In this study, we show that submerged objects are handled more quickly with wrinkled fingers than with unwrinkled fingers, whereas wrinkles make no difference to manipulating dry objects. These findings support the hypothesis that water-induced finger wrinkles … may be an adaptation for handling objects in wet conditions. (Kareklas, Nettle, and Smulders 2013, 1)

These cases illustrate that analysis of how an organism CR-functions is almost always in service of understanding in the long run how the organism SE-functions. So, I agree with Murphy that reference to CR functions (in some broad sense) is a regular and perhaps predominant occurrence in biological science, because mostly researchers focus on how things work, which is both a practical concern and a foundation for attributing SE functions. I also assume that CR functions exist alongside SE functions and the investigation of CR functions generally takes place within the broader assumption of biological design, which waits in the background for occasional but pivotal illumination. I proceed on this basis to examine whether there are grounds for Murphy's position that medical and psychiatric diagnosis of disorder can be based on CR functions alone.

One Doubt about CR Functions

Before proceeding, I want to express one possible doubt about CR functions. This is not a doubt about the existence of CR functions as causal roles, because such causal roles obviously do exist. Rather, it is a doubt about whether CR functions really are a second, *independent* and scientifically interesting sense of "function" beyond SE functions that adds to the ontology of science. Such usage of "function" might be understandable instead as a common synecdochical usage that simply captures the causal-role aspect of SE functions and is parasitic on the primary SE usage. Consider an analogy: "reason," which inherently refers to a psychological cause of action, is commonly used for nonpsychological causation as well. There are no literal reasons (i.e., belief-desire pairs) that cause headaches or earthquakes or hearts to beat, yet "reason" is used in

phrases such as "the reason for earthquakes" (Reason 2019), "28 Reasons for a Sudden or Throbbing Headache and Nausea" (Steinberg and Buoy Medical Review Team 2019), and the "reason for the motion and beat of the heart" (Harvey 1628/1993, as quoted in Ribatti 2009). This can be construed as a parasitic usage that expresses the causal power of reasons without there actually being any reasons. Or, one could on the basis of such usage assert that "reason" is polysemous. Similar options may be available in regard to the use of "function" as causal action.

The reason one might choose to understand CR-function talk as a derivative synecdochical form is out of concern about the ontological superfluousness of adding CR functions to the furniture of science. In his original article introducing CR functions, Cummins (1975) challenges the SE notion by asking, "We know why evolutionary biologists are interested in effects contributing to an organism's capacity to maintain its species, but *why call them functions*?" (756, emphasis added). He argues that one can just describe all of the biological facts without the additional label. Let's agree with the implied "semantic Occam's razor" that one does not want to needlessly introduce terminology into science that does not represent a substantive value-added ontological commitment. If so, then Cummins seems to have a problem if his question, "*why call them functions*?" is directed at his own proposed "CR function" terminology.

"Causal roles" are, after all, nothing more than abstract descriptions of causes within complex systems, an ontology that is already succinctly describable using standard causal language. Every causal analysis, whether explaining systemic capacities or describing billiard ball causation, involves abstracting from the many concrete features of the causal process and selecting causal sequences that are of interest. So, if the label denotes nothing substantive other than causation, *why call them "functions"*? (I will address this question in the case of SE functions below.) For example, in a science news article about the remarkable discovery that a long-known biomarker for pancreatic cancer actually has a causal role in causing the cancer, one author is quoted as explaining, "We serendipitously noted that the mice developed inflammation of the pancreas … and further studies then showed they had accelerated pancreatic tumor progression. CA19–9, in other words, was actively causing pancreatitis, leading to pancreatic cancer," and an independent expert who wrote an editorial about the discovery is quoted as saying, "This is a whole different paradigm. … What used to be a marker is now [taking] a functional role in pancreatic disease" (Blanco 2019). What is notable is that there would be no content or implication lost if "functional role" in the second quote were replaced by "cause" or "causal role" from the first quote because they are both expressing the same idea. This seems true of CR-function mentions in general.

There thus seems to be no ontological profit in introducing CR function as a scientific concept, which as we shall see is decidedly not true of SE functions due to their distinctive feature of what Cummins calls "effect-sensitivity." Cummins seems to implicitly recognize this oddity about CR functions when he states that he is open to not calling

them "functions": "It is, of course, perfectly possible to acknowledge that what we are calling functional analysis is both a useful and ubiquitous form of explanation in science and engineering, while denying that the analyzing capacities appealed to in such explanations are functions" (Cummins and Roth 2009, 75). However, CR functions would have more of a claim on ontological relevance if the concept of medical disorder could be adequately anchored in CR functions, and it is to that question that I now turn.

The Problem of the Objectivity of Biological Dysfunction

In a medical context, the idea that function and dysfunction and thus health and disorder are entirely relative to the interests or concerns of the observer is a nonstarter. There may be vast value-based disagreements about what to do about various functions and dysfunctions, and our ignorance of functions and dysfunctions of body and mind remains vast as well. However, the objectivity of the fact that, for example, the heart's function is to pump blood and that its failure to do so is a dysfunction is part of the scientific foundation on which medicine rests.

Cummins's view of the nature of functions and dysfunctions denies any such objectivity. Cummins (2002) begins one of his papers on functions with the following striking statement: "There are two subpopulations of functional explanation roaming the earth: teleological explanation, and functional analysis. The two are in competition" (2002, 157). Murphy follows Cummins in accepting as a starting point what is now often referred to as the consensus view (Godfrey-Smith 1993), that there are two forms of functional explanation and thus two kinds of function statements of the form, "The function of X is to Y," comparing it to the situation of the term "gene" in biology: "Different areas of biology, or even the same scientist at different times, will use the gene concept to pick out different aspects of the underlying genetic and developmental processes. I think *function* is scientifically polysemous in just this way" (Murphy, this volume). That is, "function" is used in both the CR and SE senses.

One might well ask: if there really are two different (legitimate) meanings of "function" and corresponding forms of functional explanation, then why are they in competition? Murphy's own analogy to the polysemous use of "gene" suggests a more pluralistic stance. If there really are two senses of "function," then the problem would seem to be which sense applies in which contexts. It would make no more sense to say that they are in competition than it would to say that "there are two different subpopulations of water roaming the earth, liquid water and substance water, and the two are in competition." In a restaurant, when I ask for water, I presumably mean the liquid, but in a chemistry lab or when studying molecules floating in the Horsehead Nebula, I probably mean the substance. If someone proposed that because these two related ideas are expressed using the very same word, we must decide on one meaning or the other for all contexts, that would not be taken seriously.

Nonetheless, Murphy follows Cummins in arguing that the CR view of functions is better than the SE view. Given that the major objection to CR functions has been that they are incapable of supporting objective dysfunction attributions and thus cannot be a basis for diagnosis, this is a risky position for Murphy to take. Indeed, Murphy acknowledges that this is a major problem with Cummins's view:

> The bigger problem with Cummins's analysis is that, notoriously, attributions of function on this account are interest relative: if you are interested in the heart's contribution to the circulation of the blood, you can analyze it one way, and if you are interested in its contribution to the sum total of the noises the body makes, you can treat it as a "lub-dup" generator. In both cases, it counts as a part of the overall system you wish to decompose. Cummins thought that the overall system depended on the interests of the investigator. In that sense, he was not really trying to articulate a naturalistic concept of function that captures biological practice but to understand the logic of functional analysis. This is, however, a problem for such a view. … We want an objective characterization of dysfunction rather than a mere acknowledgment that we decompose a system into its contributory parts. If that were all that the account gave us, it would threaten the objectivity that is appealing about … the HDA. … We need a way of characterizing biological systems that sees them as mind-independent components of nature, not just reflections of human concerns.

This problem defines the major task of Murphy's paper, which is to show that CR functions can provide an objective naturalist account suitable to medical diagnosis. Murphy proposes three tests that he will use to determine which of the two views of function is best suited to the foundations of medical diagnosis: "First, it should be able to make precise the idea that attributions of mental illness rely on something being objectively wrong under the hood of a human being. Second, it should capture relevant scientific practice. Third, it should not be too revisionist but make sense of the idea that some conditions may have causes that cannot be shown to depend on failures of selected effects."

Immediately below, I consider the second and third tests, which I think are based on misunderstandings and misguided assumptions. Then, for the remainder of this reply, I focus on addressing the first and conceptually most fundamental test: can CR functions explain the idea central to psychiatric diagnosis of "something being objectively wrong under the hood of a human being," as Murphy puts it? This is the essence of the idea of mental disorder, and if a view of function cannot explain this, then it is not the view of function that is relevant to medical and psychiatric diagnosis. Murphy's other two tests become irrelevant if the first test fails to support the CR approach as a viable alternative.

Is the HDA Too Revisionist?

Murphy proposes that an analysis of function and dysfunction for medical purposes "should not be too revisionist but make sense of the idea that some conditions may

have causes that cannot be shown to depend on failures of selected effects." That is, Murphy accuses the HDA of being too revisionist because, he claims, there are many disorders that are not failures of selected effects, so the HDA would misclassify them as nondisorders. This claim begs the central question at issue unless Murphy can demonstrate that there are such conditions, which he attempts to do. It is a surprising claim given that the HDA attempts to explain actual disorder and nondisorder judgments and so should follow these judgments rather closely.

I am going to answer the charge of revisionism, but I note in passing that the charge is questionable as a claimed deep flaw in the HDA, depending on the kind of revisionism. Revisionism of specific disorder judgments is perfectly acceptable scientifically. Behaviorists, social constructivists, family systems theorists, and antipsychiatrists all had accounts that if true would have radically altered and virtually eliminated our usual disorder judgments. The reason for rejecting those views was not that they were revisionist—that is hardly scientifically problematic, and it was generally agreed that if these theories were correct, then there would be few or no mental disorders—but that they were false theories. If the HDA is correct about the concept of disorder, then it is up to advances in scientific knowledge to tell us which conditions satisfy that concept, and I would let the chips fall where they may.

To support his "too revisionist" claim about the HDA, Murphy compares the HDA and CR views on their potential for revisionism and claims that the "systemic capacity view wins easily." This is because the HDA's evolutionary requirement sets a much higher bar for disorder attribution than the CR view. The HDA "stipulates that there must be some evolved system somewhere that has departed from its historical design and is connected to the current problem," but the CR view "just asks what the underlying system is that is misbehaving." Murphy thus concludes that the CR view is less revisionist because it is "less epistemically committed" and it "has one less bet to make."

Murphy argues that because the HDA classifies much fewer conditions as disordered, it forces us to revise many of our judgments about disorder. This is a patently invalid argument unless one has already independently established that the conditions excluded from the disorder category by the HDA are in fact legitimate disorders. In fact, the real problem—and the real threat of massive unjustified revisionism—is just the opposite and lies with the looseness of the CR view. Because the CR view holds that what is a dysfunction (or what underlying system is misbehaving, as Murphy would have it) depends on researcher or clinician interests, CR diagnostic practice is subject to arbitrary expansion to reflect diagnosticians' preferences. Murphy's view that lack of sufficiently all-encompassing diagnosis is of central concern reveals a failure to appreciate that the goal is not to maximize diagnosable conditions such that all systemic "misbehavior" qualifies as a psychiatric disorder (which was exactly the criticism of the antipsychiatrists that psychiatry has labored to answer) but to get diagnosis right as medically legitimate. Recent debates over *Diagnostic and Statistical Manual of Mental*

Disorders (*DSM-5*) as well as ongoing critiques of psychiatry as in the current "neurodiversity" movement underscore that overdiagnosis remains a major issue. The HDA's higher diagnostic threshold (or, actually, the fact that it has a threshold at all, because the interest-relative CR approach has none other than the diagnosticians' preferences) protects against false positives and the use of diagnosis for social control and locates psychiatric disorder within a legitimate medical domain. The CR-function approach is deeply and catastrophically revisionist because, as antipsychiatrists would quickly note, whether a system is "misbehaving" is in the eye of the beholder and not a matter of science.

Unlike the CR's free-for-all, the HDA's account of the concept of disorder is *appropriately* potentially revisionist in light of what we may discover about the scientific facts. It is important to recognize that revisionism at the level of specific disorder judgments is not intrinsically problematic, as long as one is observing the meaning of "mental disorder." Some scientific theories of the conditions generally labeled as mental disorders are truly massively revisionist at the level of specific judgments because they imply that those conditions generally do not satisfy the requirements to be a medical disorder. For example, certain forms of classic behaviorism and of extreme social constructivism imply that there are few actual dysfunctions underlying the conditions generally labeled mental disorders and so, consistent with the HDA, they claim that few conditions are genuine mental disorders. The HDA more modestly suggests that, in light of the facts, some revision of our bloated nosology is needed, including depathologization of certain types of what is now classified as attention-deficit/hyperactivity disorder (ADHD) (see my reply to De Vreese in this volume), depressive disorder (Horwitz and Wakefield 2007), autism spectrum disorder (Wakefield, Wasserman, and Conrad 2020), anxiety disorder (Horwitz and Wakefield 2012), and others (Wakefield 2016b). The ability to identify such justified revisions correcting diagnostic overreach allows psychiatry to respond to criticisms that specific categories go beyond the domain of medicine proper and have become a tool of social control, thus constructively addressing the legitimate concerns of the antipsychiatrists. The CR approach offers psychiatry no such response to antipsychiatric concerns.

Remarkably, the only example of the HDA's supposedly being "too revisionist" that Murphy offers is the tired (and unrepresentative; see my comments below on Kingma) "dyslexia"-type examples that I have addressed at length elsewhere (e.g., Wakefield 1999a; see also my reply to De Block and Sholl in this volume for further discussion of dyslexia). It is worth commenting on a couple of faulty assertions in Murphy's discussion. First, echoing other critics, Murphy suggests that in my portraying dyslexia in HDA terms as a social harm (inability to learn to read) caused by an inferred (but as yet not clearly established) brain dysfunction, the dyslexia example shows that I simply make up dysfunctions in an ad hoc manner so that my position is unfalsifiable: "Wakefield seems to think that … one can always postulate the existence of a selected

function that is linked to the condition. … This seems grossly ad hoc; … this just makes the HDA immune to counterexample by stipulation … Wakefield can just insist that there must be one somewhere."

This common objection reveals a misunderstanding of the HDA. The HDA is a conceptual analysis that explains when we are justified in our attributions of disorder; it is not a substantive theory of any particular condition's status as a disorder or nondisorder. Dyslexia is considered a disorder because it is believed to be due to a dysfunction, and one can excavate from the literature the reasoning that supports this hypothesis in the minds of clinicians and researchers. The HDA makes no claim that dyslexia actually is caused by a dysfunction; that is a matter of scientific evidence, not conceptual analysis, and nothing about the HDA's validity depends on whether there is or is not such a dysfunction. The HDA asserts that people attribute disorder when they believe there is a dysfunction; it says nothing about whether people are right in any given case. The relevant evidence consists of whether the way that reading disorder specialists and others think about and justify diagnosis of dyslexia is consistent with the HDA's conceptual-analytic claim. This evidence is publicly available and cannot be made up in an ad hoc manner. The question for the CR view is whether specialists believe that diagnosis of dyslexia is justified whenever a child does not learn to read given that we have an interest in children reading and thus failure to learn to read constitutes the "misbehaving" of some system in the child and warrants being labeled a disorder. When I reject that view and claim that dyslexia presupposes a dysfunction in the stronger HDA sense, I am reporting what a close examination of dyslexia diagnosis and research clearly reveals, namely, that dyslexia is commonly distinguished from all other manner of nondysfunction "misbehaving" (e.g., lack of education, emotional distraction, lack of motivation, low general intelligence, etc.) that causes lack of ability to read, any of which could be classified as CR dysfunctions, and moreover that, in the course of diagnosis, symptoms suggestive of neurological dysfunction (such as, in former days, letter reversal, although this particular symptom is now questioned) are accorded special significance as confirming the diagnosis. One can dispute my reading of the evidence, but there is nothing ad hoc or unfalsifiable about this approach to testing the HDA's explanatory power.

Moreover, there is a false background assumption underlying Murphy's dyslexia argument for the HDA's revisionism. Like some other critics (e.g., see Garson in this volume), Murphy presumes that we would continue to consider a condition such as dyslexia a disorder even if we discovered it is not due to a dysfunction, as if the attribution of disorder to a condition is atheoretical and remains fixed no matter how we explain it. The history of psychiatry shows this is false. When problematic psychological conditions come to be understood either as naturally selected or as just something other than failures of natural functions, they also come to be understood as nondisorders (see my replies to Garson, to De Vreese, and to Cooper in this volume).

In response to my claim that clinicians and researchers limit disorder attributions to what they believe to be dysfunctions, Murphy actually agrees but retorts, "This is correct, but it only meets the objection if those researchers are committed to an evolutionary account of function, which is the very point at issue. The fact that they take for granted the existence of a dysfunction … does not show that they take for granted the existence of an evolutionary dysfunction." This objection again confuses substantive essentialist theorizing with conceptual analysis. Obviously, most clinicians and researchers and laypeople are not explicitly thinking about evolution in judging there is a disorder, any more than people are thinking about the chemical formula H_2O when they ask for a glass of water in a restaurant. The concept of disorder was around long before Darwin and is understood by evangelical Christians who reject evolutionary theory, so someone need not have evolution in mind or even know about or believe in evolution to understand the concept of disorder. As I have explained elsewhere (Wakefield 1999a, 1999b; see also my reply to Lemoine in this volume), "disorder" means "harmful dysfunction" where dysfunction is understood intuitively in terms of failure of natural function, that is, failure of how we are biologically designed, which is a notion that has been at the center of biology and available to common sense ever since the discipline began. However, since Darwin, we know that the best theoretical explanation of natural functions and biological design is evolution by natural selection. The HDA proposes that evolutionary dysfunction is the best *theoretical explanation* of what clinicians and laypeople are aiming at when they identify disorders in the medical sense (see my reply to Forest in this volume for further comments on this point). So, in assuming there is a breakdown of some sort in the way the organism is designed to function, clinicians are in effect assuming there is an evolutionary dysfunction, just as in asking for water in a restaurant, one is in effect asking for H_2O even if one does not know anything about chemical theory. Note that the objectivity of dysfunction—Murphy's first test of an account of dysfunction, which is evaluated in detail below—is a salient aspect of disorder judgments that requires explanation, and evolutionary dysfunction explains this feature, whereas CR dysfunction cannot explain it unless, of course, Murphy can make good on his claim that objectivity can be teased out of the CR view, a claim that I evaluate below.

Reply to Kingma's Genetic Linkage Argument for the HDA Being Strongly Revisionist

There remains one central "revisionist" argument deployed by Murphy, which he largely outsources by citing a handbook chapter by Elsilijn Kingma (2013), in which she critiques the HDA as unacceptably revisionist. Kingma's and Murphy's argument is essentially that genetic linkage causes phenotypic effects that are not due to natural selection, so if those effects are valued but go wrong, then there are disorders that are not failures of natural functions to which the HDA denies disorder status, making it highly revisionist

relative to current diagnostic practices. Murphy puts the linkage-based revisionism argument as follows: "Now, as Kingma notes, it is quite possible for a linked trait to be abnormal for independent reasons that have nothing to do with a dysfunction in the system it is linked to. … Kingma argues that Wakefield's position is potentially very revisionist indeed: if a large number of disorders turn out not to be linked to selected effects, but the product of other processes, such as developmental or genetic linkage, then he will have to say that they are not really disorders." This linkage criticism is closely related to a criticism earlier leveled by Murphy himself (Murphy and Woolfolk 2000) to which I replied (Wakefield 2000b). However, given this new more elaborated version, once more unto the breach.

Kingma cites several genetic linkage phenomena in addition to standard genetic linkage due to close positioning on a chromosome. The complexities of, for example, heterozygous advantage, developmental linkage, and antagonistic pleiotropy pose interesting challenges for any theory of health and disorder. These phenomena are so unanticipated that they might have required what I have called "conceptual rectification" with the intuitive concept of natural function, in the way that isotopes required rectification with the intuitive structural concept of substance (see my reply to Lemoine in this volume). However, Kingma acknowledges that the HDA can handle cases of heterozygous advantage (e.g., sickle cell trait versus sickle cell anemia) simply by distinguishing between the positively selected nondisordered trait and the negatively selected disorder, and a similar resolution can work when both effects are within one individual, as in antagonistic pleiotropy. So, I focus on Kingma's primary example of standard genetic linkage, which refers to the fact that a trait not at all selected for—thus either neutral or even deleterious with regard to fitness—may nonetheless be selected at higher than neutral statistical rates due to its proximity on the chromosome to a trait that is naturally selected. This is due to the mechanics of reproduction in which close-together genes tend to be kept together. Such linkage is not sensitive to the nature of the phenotypic results of the gene but is just a matter of the happenstance of where the gene falls on the chromosome, so it is unlikely to display apparent manifestations of biological design. Note that in trying to show that there are disorders without dysfunctions, Kingma has in mind examples in which there are no HDA-type dysfunctions to muddy the waters, so failure of the nonselected trait is neither due to nor causes any HDA-type dysfunction of any selected trait.

Kingma illustrates the possibility of nonselected traits due to linkage with the example of blue eyes, which is known to be linked to a skin coloration gene: "For example, the presence of blue eyes is not explained by an effect of blue eyes, but by the increased ability of lighter skin to absorb ultraviolet B radiation (which helps with vitamin D production). This can happen because the trait 'blue eyes' is linked to the trait 'light skin'" (392). Most people with the light-skin gene that causes lower production of melanin have the usual brown eyes, but in some people, the closely linked gene that regulates

eye color has a genetic "switch" in an "off" position that blocks melanin production in the stroma of the iris and thus yields blue eyes. Kingma assumes that blue eyes are selectively neutral but are maintained or increased in the population because of the blue-eye gene's linkage to the pale-skin gene, which has been selected for in northern populations. (In actuality, there is some evidence that the blue eyes trait has been sexually selected and this explains the spread of blue eyes, but I leave this complexity aside here.)

Now, to Kingma's argument that "such linked traits pose a serious problem for Wakefield's account of disorder" (392). The problem is claimed to arise, first, because linked traits that are not themselves naturally selected have no natural function: "traits that are not selected for their own effect, but are selected because of their linkage to other successful traits, do not have a function on Wakefield's account" (392). Second, despite lacking a natural function, such nonselected linked traits may be socially exploited in various ways: "It is overwhelmingly likely that we have an abundance of traits that fulfill important roles for us and in our culture, particularly in the mental realm, but whose effects may not be what drove their selection" (395). Third, it is in principle possible for a linked trait to be interfered with without causing a problem for the naturally selected gene to which it is linked, and thus without causing any collateral evolutionary dysfunction: "it is not just possible, but in fact overwhelmingly likely, that all manner of things … could affect one out of a pair of genetically or developmentally linked traits without affecting the other" (393). So, for example, "even though the selection of blue eyes is explained by the effects of fair skin, it is entirely possible for something to happen to the blueness of my eyes without my skin being affected" (393). This potential for something to happen to a linked nonselected trait without affecting any selected trait applies to the aforementioned linked traits that we value and socially exploit. However, because these valued linked traits do not have a natural function, if they fail, the failure cannot be a dysfunction on the evolutionary account: "These traits therefore lack functions and, by consequence, the ability to dysfunction" (395). (I note in passing that all these premises would also apply to Boorse's view that Kingma embraces as superior because Boorse limits functions to the recent contributions of a part to survival and reproduction, and most nonselected linked traits will make no such contribution and thus lack Boorse-type functions.)

I agree with Kingma's premises but disagree with the conclusion that Kingma claims to follow from these premises that the HDA is highly revisionist. Kingma does not explicitly take the final steps to her conclusion, but the only route to get there is by adding something like the following suppressed premise: *the concept of disorder is such that any failure of a valued human trait qualifies as a disorder*. She indicates that the nonselected conditions that she thinks are truly disorders but revisionistically not classified as such by the HDA are "traits that fulfill important roles for us and in our culture." The suppressed premise of her argument is more general because there is nothing special

about valued linked traits versus other valued nonselected traits that distinguish them in the argument. Moreover, linkage was unknown to those using the concept of disorder throughout medical history and so cannot play any distinctive conceptual role in the concept of disorder. Rather, linkage is just a vehicle for Kingma to make the correct point that some traits we value have no HDA-type natural function. Her suppressed premise that such conditions are disorders allows the argument to proceed to its suppressed conclusion, as follows: *Therefore, when valued linked traits fail, even though such failures are not dysfunctions, the failures fall under our concept of disorder and are in fact considered disorders. However, the HDA does not classify such failures of valued nonselected traits as disorders. Thus, the HDA is revisionist.*

The only example provided by Kingma is, like Murphy, a speculation that dyslexia might result from the failure of linked genes rather than from a dysfunction of selected brain mechanisms. Generalizing from the dyslexia example, Kingma further argues that such linked traits and their failures must be very common even among traits we (often mistakenly, according to Kingma) take to be biologically designed. This leads her to her unrestrained mincing-no-words conclusion: "Wakefield's account of disorder, it turns out, is very strongly revisionist" (394) (also, "highly," "so very," "terribly," and "so terribly" revisionist and "more revisionist than he realizes" [396, 394, 395]). I will examine each of the three aspects of this argument: the concept of disorder as failure of valued traits, the dyslexia example, and the argument for the frequency of nonselected traits constituting mental modules that we consider selected.

An argument that the HDA is revisionist is by nature comparative and relative to some accepted nonarbitrary baseline. Kingma's argument is built on her key assumption, never explicitly defended, that our baseline concept of medical disorder encompasses all failures of socially valued traits. Kingma relies on this inflated account of disorder to argue for the HDA's supposed revisionism in not properly classifying every socially disvalued trait as a disorder. This claim has no plausibility on its face because there are endless positive traits, the absence of which are not considered disorders. There is nothing special about genetic linkage here. With regard to any traits, socially undesirable normal variation—scoring low, for example, on dimensions of intelligence, height, mathematical or musical ability, verbal skill, social skill, emotional resilience, physical strength, physical attractiveness, and so on—is not considered a disorder as long as it is within a range that is not considered to indicate dysfunction. For those aware of the history of psychiatry, the claim that socially disvalued traits are disorders is not merely false but worrisome because it would be the basis for social control by psychiatry and invites antipsychiatric objections.

Consider, for example, the illustrations in the following passage in which Kingma claims that there are naturally selected disorders (see my supplementary response to Cooper for fuller discussion of this claim):

> The first possible problem for Wakefield is that of "selected disorders." These are selected effects or strategies that have very negative effects in our present society. Possible examples include forms of antisocial behavior: rape, a violent disposition, or dependent or attention-seeking behavior. All of these may have been beneficial in selective terms: serial rape can be a good strategy for increasing one's reproductive output, for example, and violence, dependence, or attention-seeking may all increase one's access to resources that in turn increase fitness. But if it is true that these conditions have been selected, then they are not a disorder according to Wakefield—and this countervenes our current way of thinking about these conditions. (2013, 391)

Her examples, based on the criterion that conditions that "have very negative effects in our present society" are disorders, indicate that it is Kingma who "countervenes our current way of thinking about these conditions." There is no category in *DSM-5* or *International Classification of Diseases* (*ICD-11*) in which these undesirable conditions are generally labeled disorders. They are only labeled disorders in extraordinary, extreme cases in which there is a plausible argument that there is a dysfunction of the system that gives rise to them. For example, aggression and violence per se, even when they involve mass killing, are not necessarily considered a disorder (e.g., Knoll and Pies 2019). Engaging in multiple rapes is a crime, not a disorder (Wakefield 2011b). *DSM-5* did consider a rape-related paraphilic category of coercive sexuality in which the coerciveness itself becomes the primary arousing factor and rejected it partly because it could easily be confused with the nondisorder of serial criminal rape. Regarding antisocial behavior more generally, although universally considered undesirable, it is not in itself considered disordered (e.g., antisocial delinquent behavior and even gang behavior among youth is generally not pathologized). Moreover, as I have explained elsewhere in this volume (see my responses to Cooper and Garson), when researchers conclude that antisocial personality is a naturally selected strategy or the result of social conditions and not an HDA-type dysfunction, they tend to revise their belief that it is a disorder and come to understand it as an undesirable nondisordered normal variant. In *DSM-5*, regarding categories that address antisocial behavior (e.g., intermittent explosive disorder, antisocial personality disorder), *DSM-5* takes pains in the diagnostic criteria to try to separate dysfunctions of aggressive tendencies from the wider issue of antisocial behavior in general that Kingma seems willing to pathologize. Neither attention seeking nor dependence per se are generally considered disorders, despite being looked down upon in our society.

Kingma later presents what she sees as her "most damaging" revisionist objection. She argues that "most if not all of our physical, and the vast majority of our mental traits, fall within the domain of health and disorder … they are either disordered, and if not, they are healthy." She claims that "an evolutionary account of disorder can never bear this out" (395) because only selected traits can have functions and can dysfunction. She concludes that "Wakefield's account of disorder places a substantial portion

of our physiological and mental traits out of the realm of health and disorder altogether" in "a clear violation of one core conceptual element of the health and disorder dichotomy" (396).

First, Kingma's argument that the HDA does not dichotomize health versus disorder is based on an elementary misunderstanding. Like Christopher Boorse, whose view Kingma in her chapter embraces as superior to the HDA, I define health for the purposes of these analyses as lack of disorder. Thus, any selectively neutral trait, whether it is positive or negative in terms of its social valuation, is part of normal variation and thus part of health along with every selected variant. The issue here is not dichotomizing health and disorder but where socially disvalued nondysfunction conditions fall in the dichotomy. Kingma's point in emphasizing the dichotomy is to locate all socially negative traits on the "disorder" side of the dichotomy rather than allowing for normal (nondisordered) variation that is socially disvalued and yet perfectly healthy. Her view fails to comport with standard medical concepts. Nor does it follow that socially disvalued normal variation cannot be treated. I (Wakefield 2015) have argued that if limitations to an individual's opportunity are primarily due to socially negative valuation of parts of normal variation, it is a matter of justice that the individual should be offered treatment, even though the condition is not a disorder.

So, who is the revisionist here? The issue here is not whether undesirable nonselected effects of linked traits may occur but whether such conditions are (or would be) considered *disorders*. Kingma's inclusion of socially undesirable traits within disorder has no foundation in the history or current practice of diagnosis and no coherent relationship to psychiatric diagnosis as it is codified in *DSM-5* or *ICD-11* or as it is generally practiced by mental health professionals. Kingma's expansive view eliminates the divide between immoral, illegal, ignorant, and other undesirable behavior, on one hand, and mental disorder, on the other, undermining psychiatry's distinctive validity and making it the agent of social control that antipsychiatry accused it of being.

Kingma and Murphy both use dyslexia (reading disorder) as their primary—indeed, only—example of the HDA's potential revisionism. They both think that to defend the HDA, I must claim that dyslexia is in fact caused by an evolutionary dysfunction and that this is uncertain: "Wakefield's response seems terribly ad hoc. Of course it may be the case that our ability to read is produced by a mechanism that was selected for a particular effect, and that dyslexia indicates a breakdown of that mechanism. But it is just as plausible that that mechanism is itself a by-product of the selection of a different, linked trait, and therefore lacking in function. Or that both the normal ability to learn to read and dyslexia are on a spectrum of normal variation in non-selected effects produced by a functioning underlying mechanism. … Wakefield, therefore, seems to be making a risky bet" (393).

However, there is nothing to bet on here. The HDA is an explanation of how we think about disorder, not a substantive theory of any particular disorder. Classically,

based on various arguments regarding similar symptoms occurring in brain trauma, brain studies that indicated anomalies, and such recently questioned beliefs as that dyslexia involves such neurologically suggestive symptoms as letter and word reversal, it was generally concluded that dyslexia is due to a brain dysfunction. This is still the majority view by researchers and clinicians. However, reading the literature on dyslexia, in fact there is much disagreement and debate, and the arguments reveal how people think about disorder versus nondisorder. There are some researchers who think dyslexia is a disorder, and they justify this belief with the claim that it is due to a malfunction of certain brain mechanisms. There are others who think that dyslexia is not a failure of brain mechanisms but some form of normal variation—perhaps along the lines of some of the linkage options offered by Kingma—and they deny that it is a disorder. It is clear from the literature that if dyslexia were to be proven to be due to variations in the effects of neutral linkages and thus was a form of normal variation and not something going wrong with the brain, then there would be a move to depathologize it. Indeed, the neurodiversity movement and many dyslexia support organizations already argue for Kingma's third option, that dyslexia is just the lower end of normal variation in the ability to learn to read and there is no dysfunction, but they conclude that, in virtue of this, there is no disorder. The HDA attempts to explain such differences in views of disorder by the differences in belief about dysfunction, and the literature on dyslexia tends to bear out the HDA's predictions. That literature sharply diverges from Kingma's views; all discussants agree that difficulty learning to read is a serious negative condition that warrants intervention in our society, yet there is a sharp and vigorous divide over whether it is a disorder based on differing views of whether there is a brain dysfunction. My discussions of dyslexia in regard to the HDA were never aimed at betting on one outcome or another but on explaining that those who believe that dyslexia is a disorder base that judgment on their belief that it is due to a dysfunction. For this account of the conceptual distinction, no hypothesis, ad hoc or otherwise, is necessary about the actual cause of dyslexia. (For various approaches to dyslexia, see, for example, the following: Armstrong 2015; Ap 2016; Artigas-Pallarés 2009; "Dyslexia Has a Language Barrier" 2004; "Dyslexia Is Not a Disease" n.d.; Habib 2000; Lilienfeld 2010; Protopapas and Parrila 2018, 2019; Schneps 2015; Treiman 2014; Ziegler et al. 2003).

Now, where does Kingma get her frequency estimation that the HDA is "very strongly" revisionist? Without evidence, based on the spurious accusation that postulates an ad hoc bet about dyslexia that can go wrong, Kingma generalizes the accusation to all disorders, for each purported disorder could be caused by failure of a linked trait rather than a failure of a selected effect: "Wakefield can bet against the odds in one case, dyslexia, and either win or lose. But if very many of our mental capacities are like reading—that is, effects of traits that do not themselves explain why those traits were selected, and that are therefore not functions—Wakefield's position starts to look more precarious" (394).

However, dyslexia is atypical precisely because the harm is not a failure of an apparently selected trait and so there is an unusual degree of inference involved in deciding whether or not there is a dysfunction. In contrast, most mental disorders in *DSM* and *ICD* are generally identified as disorders because they are occurring in what are pretty clearly biologically designed systems with complex regulatory features that are not plausibly due to linkage, and the failures that are labeled disorders compromise what appear to be the designed functions of the module. It is thus actually "overwhelmingly likely" that linkage is not the explanation for the vast majority of mental modules. Page through the chapters and categories of *DSM* and what you find are failures of the systems that are most plausibly biologically designed, such as human thought, perception, sadness, joy, grief, fear, psychological development, sleep, sex, eating, excretory function, and other categories of function, all of which are most plausibly the result of selection. This includes the vast majority of currently recognized physical and mental disorders.

Perhaps recognizing the flaw in the sheer "probability" argument for linkage, Kingma attempts to finesse the problem of the evidence for design by casting doubt on whether we are able to tell whether a system is likely biologically designed. She uses the example of reading: "It seems almost impossible that our ability to learn to read would not be designed: it is unique to humans, complicated, widespread, and incredibly useful, so how could it be a fluke? But it is a mistake to think that something is either selected, or a fluke" (395). Quite right, but what this example shows is that there are many sorts of processes that can lead to signs of design, which include biological selection, social construction, individual learning, and artifact-creation. After all, despite the design-like characteristics of our ability to read, no one in the diagnostic community has ever held that reading is a naturally selected trait.

Ignoring the fact that we theorize about mental modules based on evidence of likely design, Kingma tries to portray it as a matter of sheer probability whether a module is designed or not: "Here is one reason to suppose that more rather than fewer of our mental capacities will be like reading. . . . If a selected effect account is to bear out that mental modules have functions . . . those effects should . . . have been the drivers of the selection of those very modules. In other words, every single mental module or capacity must have been 'visible to natural selection' via its own effect rather than through any of the other possibilities discussed. . . . Given the developmental complexity of our mind, that seems extremely unlikely. . . . Wakefield's account of disorder, it turns out, is very strongly revisionist" (394).

I would say that, given the developmental complexity of our minds, selection was critical to getting all the pieces to work and interact correctly. Moreover, it is precisely due to the plausibility of biological design that we select modules, so there is nothing at all surprising or improbable about each module having been subject to selective pressures. It is not a matter of probability whether mental modules happen to fall within the selected for versus nonselected linked category. Mental modules are identified

based on their apparent biologically designed and adaptive nature. Geneticists and evolutionary biologists affirm that natural selection is the only known process that reliably or frequently gives rise to apparent biological design. The nonselected effects of linkage that happen without reference to the content of the produced traits have no causal properties that are likely to produce such features.

Having argued that one can easily be mistaken about design, Kingma concludes, "Therefore the fact that our traits seem beautifully adapted to what they do should not tempt us into thinking that they were selected for what they do" (395). This overstated conclusion does not follow from anything she has said. Should we throw out Darwin and dismiss Aristotle's foolish notion that the acorn turning into an oak tree required a special form of design-like explanation? Should we perhaps resist the temptation of thinking that the artifacts in our homes have been designed? Taking seriously the evidence of design and adaptation has been the basis for the greatest breakthroughs in the history of biology. Just as in every other area of science, the fact that we can be mistaken does not obviate the general reliability and usefulness of our intuitions for generating initial hypotheses that allow us through testing to bootstrap to the truth.

Kingma ends up rejecting my view because "our concepts of health and disorder are not value-free" (397). This is of course my position as well, but Kingma never mentions the evaluative "harm" component of the HDA. Kingma prefers Boorse's (1975) view—which is genuinely nonevaluative—but never considers that Boorse's view is highly revisionist; its statistical criterion allows arbitrary classification of the bottom 50% of the population on any functional variable as disordered, and it classifies as a medical disorder any biological dysfunction no matter how harmless, of which each of us has millions (e.g., genetic mutations) (Wakefield 2014). Kingma prefers the "forward-looking model created by Boorse, not the backward-looking one by Wakefield" (397). Just as contemporary essentialist semantics builds causal history into the meaning of proper names and natural kind terms without thereby rendering the activities based on their meanings in any sense backward-looking, invoking history as part of the meaning of medical disorder has little to do with how forward-looking medical practice is. What is not forward-looking is to open the door wide to bogus diagnostic practices that support social control efforts and undermine the legitimacy of medicine, thus reawakening dormant antipsychiatric concerns that with decades of painstaking effort have been put to rest.

Does the HDA Cohere with Relevant Scientific Practice?

Regarding Murphy's proposed test that the concept of function "should capture relevant scientific practice," the most "relevant scientific practice" in this case is scientific practice that concerns the nature and treatment of disorders. Murphy says, "Insofar as psychiatry is a branch of medicine, the concept of function it needs resembles those

of physiology and biomedicine. Evolutionary considerations are just beside the point." However, the concepts of function and dysfunction underlying medical diagnosis cannot be judged by adjacent disciplines that do not have the unique burdens and aims of medical diagnosis. Even if some other areas of science use "function" differently (which is not at all as clear as Murphy suggests), the real question is how medical diagnostic clinicians and researchers use the concept. Murphy here ignores his own point that "function" is polysemous, which suggests that it may have somewhat different meanings in different subdisciplines.

But what about the integration of different medical and nonmedical parts of biomedicine? As my analyses below will make clear, whether or not other areas of biology sometimes use CR functions, they also assuredly use evolutionary dysfunctions, so there will be no problem with an interface. In any event, a proper interface between psychiatry or medicine and the other biological sciences requires not that "function" is used in precisely the same way across disciplines but only that the various uses of function intermesh in a coherent and scientifically useful way even if they are somewhat different. And they do intermesh because, as Tinbergen (1963) famously observed, proximate causal explanation in terms of how things work and distal evolutionary explanations are complementary integral parts of an overall understanding of the biological design of organisms. Of course, most research on disorder is about how things work, not evolutionary functions, but, as we saw above, that research takes place within a framework set by the concept of biological design.

Murphy asserts that "the questions that medicine asks are not those that a selectionist account of function can answer." A glance at recent disputes about disorder status reveals that evolutionary considerations often do address basic questions of disorder versus nondisorder. For example, is lactose tolerance a disorder, given that it exists in a minority of the world's population and involves loss of an efficiency advantage after weaning? The answer, presented, for example, in the Gluckman (2009) reference that Murphy cites, is that because lactose tolerance has been naturally selected in environments where the domestication of animals made milk available during famines, it is not a disorder. Is ADHD partially caused by the DRD4 7-repeat allele a genetic disorder? Studies of sedentary versus nomadic populations suggest that this gene was naturally selected for interest in novelty and activity and that children who have this gene and display symptoms of ADHD are not suffering from a disorder but rather a normal variation that is mismatched to the demands of our current social environment (Eisenberg et al. 2008), and this information changed people's minds about the pathological status of this subset of ADHD diagnoses (see the discussion of ADHD in my reply to De Vreese in this volume). More generally, such considerations remain implicit because manifest (albeit fallible) indicators of design and dysfunction are taken as sufficient for judgment in most cases.

Murphy further argues, "Normal biomedical ascription of function to a system makes no claims about selective history. It requires only that we can identify the role

played by a system in the overall economy of the organism. How is dysfunction determined? By the use of a biomedical concept of normality that is an idealized description of a component of a biological system in an unperturbed state. It does not rest on the failure of a biological part to replicate as its ancestors did, or to reduce overall fitness, but by its failure to be close enough to the causal contribution of the analogous part in the idealized overall system." Again, ascriptions of function and dysfunction generally operate at a more manifest level with the biological-design evolutionary underpinnings implicit, just as normal ascription that a liquid is water makes no explicit reference to chemical theory. In any event, Murphy's description of the situation begs the crucial objectivity question. It sounds objective—sort of like classic physics—to judge that a condition would exist in an "unperturbed" or "idealized" system. However, within a CR-function account of the kind that Murphy embraces, the notions of "unperturbed" and "idealized" are interest-relative terms that confer no objectivity, so the "overall economy of the organism" can be described in any way the investigator prefers with functions and dysfunctions distributed according to the investigator's interests. Such an account is not related to the medical target of health in the usual sense, unless, again, objectivity can be teased out of the CR view.

In sum, Murphy's claim that the CR view is superior to the HDA on the two tests of diagnostic revision and disciplinary integration fails. But can he make good on his further claim that the CR view passes the crucial test of yielding an objective measure of function and dysfunction corresponding with the way health and disorder judgments are actually made in medicine. I now turn in the rest of this reply to the question of whether the CR view can provide that kind of objective account of functions and dysfunctions.

Cummins on Interest Relativity and the Function of the Sound of the Heart

The interest relativity of CR functions implies that function judgments can occur in contexts in which they go against standard usage and intuition. Cummins unhesitatingly bites the bullet in such cases and insists that such attributions are entirely legitimate and the intuitions can be ignored. He maintains this position even in such classic cases of intuitive nonfunction as the sounds made by the heart: "Evolutionary biologists probably will not say that a function of the heart is to make sounds. But an ethnologist studying medical diagnosis probably wouldn't blink an eye. This relativity to a containing system and target capacity is just what the systematic account would predict, if it were in the business of predicting intuitions" (Cummins and Roth 2009, 83).

There are several problems with this facile claim if taken seriously. First, where is the evidence from the medical diagnostic literature that those studying medical diagnosis actually say the sorts of things that Cummins claims? There are endless sources that describe the functions of the heart, and not one of them that I have accessed, including

the ones concerned with medical diagnosis, says that they include making a sound. The sound and pulse of the heart have played crucial roles in medical diagnosis since ancient times, so if Cummins's claim is true, there ought to be plenty of sources to illustrate its truth. If in fact the sources blink and refrain from such talk and Cummins's claim is unsupported, that reveals something about "function."

Second, according to Cummins's view, it should be a clear dysfunction if one's nose is shaped in such a way that it does not hold up one's glasses (Wang 2017; "Why Asian Fit?" 2019) or if one's blood vessels are hard to find and roll out of the way when a medical person is trying to take blood (as do mine in my left arm). Yet, no one describes these normal-variation situations as dysfunctions.

Third, even if medical diagnosticians were to apply "function" to heart sounds, they likely would not mean that the sounds have a natural or biological function, which is the sense relevant to medicine and relevant to the HDA. "Function" also applies to instrumental means in intentional action, including artifact construction. Medical diagnosis is an intentional human activity, and this creates the possibility of legitimate intentional-function attributions such as "the function of (listening to) heart sounds during a heart examination is to. . . ." The HDA does not block such instrumental function attributions, but it does imply that the failure of an intentional function would not yield judgments of dysfunction and disorder relevant to medical diagnosis. So, the test of the HDA here is whether ethnodiagnostic researchers, when they come across individuals in a diagnostic context whose heart sound is muted or difficult to access for idiosyncratic anatomical reasons and thus the use of heart sounds diagnostically is disrupted, would judge the individual to have a biological dysfunction and thereby a medical disorder. This test would presumably not be passed by the claimed function attributions to heart sounds.

To all the arguments presented above, Cummins answers that they are based on intuitions we have about function attributions, but CR functions violate prior intuitions and so intuitions can be ignored. He compares this to what often happens in proposing breakthrough counterintuitive theories in the sciences: "We agree that the account does not square with intuitions about functions, in many cases. . . . But how seriously should we take this? Scientific treatments of motion have increasingly diverged from intuition since the seventeenth century. . . . Biology should be no more constrained by intuitions concerning functions than physics should be constrained by intuitions about motion. Physics is, in large part, counterintuitive, and so, it would be no knock against biology if it turns out it makes counterintuitive function attributions" (Cummins and Roth 2009, 82–83). Murphy follows Cummins in arguing that revisionism that overthrows common beliefs and intuitions should be no obstacle in science: "If we are trying to capture a scientific concept, on what grounds can we argue that it should be criticized for departing from traditional conceptions? A chemist who was told that traditional usage did not regard objects as made of atoms would be unmoved."

Cummins and Murphy here suggest that the reason we should not get too exercised by the counterintuitive conclusion that "the function of heart sounds is to alert the physician to medical problems" is because we should be ready to adjust to profound scientific insights that turn our view of the world upside down. However, such semantic deployment of "function" is scientifically trivial, based on no remarkable scientific discovery. In contrast, the discovery that biological functions are naturally selected effects is one of the most momentous and counterintuitive scientific discoveries in human history, comparable to Murphy's example of solid objects being made of atoms, and it did overthrow much prescientific understanding, supporting a distinctive term for naturally selected effects. The Darwinian image of "function" *is* the "scientific image" in the biomedical sciences.

Why Call SE Functions "Functions"?

We are now in the position to answer Cummins's "why call them functions?" argument and the answer is simple: functions are effects that are the products of biological design, and the best scientific theory we have of the intuitive concept of biological design is evolutionary theory. Even Cummins recognizes that there is something special along these lines about naturally selected effects:

> According to the selectionist, appeals to function to explain the spread of a trait are legitimate because there is a function-sensitive natural process that spreads traits: natural selection. … We have no problem with natural selection. So, if selectionists see functional explanation as simply a standard application of natural selection, then we can have no objections to selectionist accounts of functional explanation, so understood. (Cummins and Roth 2009, 77)

There are two crucial points. First, natural selection is a *function-sensitive* or, better, an *effect-sensitive* causal process because the *effects of traits* must be cited in explaining why the traits are present and how they are structured. This is a highly unusual situation. Such effect-sensitive processes are neither quite prototypically mechanistic nor quite prototypically teleological. Cummins's term "neo-teleology" is as good as any for labeling this sort of effect-sensitive causal process. Second, effect-sensitivity is not an explanatory desideratum or goal in itself. Rather, it is a necessary element in explaining biological design, and it is biological design that is the ultimate target domain of classic function explanations. As Aristotle already observed, oak trees can be explained mechanically in terms of the CR functions by which an acorn gives rise to an oak tree, but that misses something explanatorily crucial and distinctive, namely, how could it be that something could have those causal properties yielding species reproduction in a mechanical universe? It must be that acorns are the way they are *because* they give rise to oak trees. But, again, how can that be in a mechanical universe? Similarly, for William Harvey, the function of the heart is to pump the blood, and "this is the only

reason for the motion and beat of the heart" (Harvey 1628/1993, as quoted in Ribatti 2009, 2), that is, the function of pumping somehow explains why the heart beats, and how this can be so is a mystery in addition to and transcending the mystery of the CR mechanics of the heart pumping the blood. What all this means is that adding the term "natural function" or, equivalently, "biological function" to the ontology of biological theory has real theoretical content that, while providing an initial minimal explanation sketch, identifies a fundamental and profoundly challenging explanatory target.

As Cummins and Roth (2009) point out, contra Aristotle, mechanics turned out not to be effect-sensitive or have use for a notion of natural function. There was no theoretically interesting sense in which objects were moving as they did because they were seeking to get to their natural place, nor was there any sort of feedback as to how they were doing in reaching that goal that then influenced their subsequent motion. Thus, the teleological effect-sensitivity aspect of ancient mechanics dropped away. Efficient-mechanical causation was deemed sufficient to explain motion and became the primary model for scientific explanation.

In biology, things went differently. The greatest puzzle about biological entities from ancient times was how, in a mechanical universe, organisms can possibly be so well adapted to the environment in a design-like manner and perform the seeming miracles of surviving and reproducing. Note that the process of reproduction takes the puzzle beyond sheer adaptation in the sense of a match between the needs of the organism for survival and the environment, as well as makes it a broader puzzle of biological design. Of course, as Murphy repeatedly points out, biological, anatomical, and biomedical investigations throughout history tried to figure out how things work (or how they function, where function is understood in CR-like terms as causal action) and do not generally refer explicitly to biological design, let alone natural selection (although selection-like theories of adaptation go back to ancient times).

Nonetheless, it is difficult to overstate the centrality of the issue of the design-like nature of organisms in the history of biology and the importance of Darwin's explanation for it in terms of natural selection. This point seems underappreciated by Murphy and Cummins, so allow me a brief historical anecdote. In 1880, the eminent German electrophysiologist and discoverer of the nerve action potential, Emil Du Bois-Reymond (1818–1896), delivered a famous lecture to the Berlin Academy of Sciences on the occasion of Leibniz's birthday, published two years later in *Popular Science Monthly* (Du Bois-Reymond 1882), in which he listed what he considered to be the seven most fundamental unsolved scientific puzzles. These basic mysteries, which he labeled the "seven world enigmas" (or "seven world problems"), ranged from "the origin of motion" and "the origin of life" to "the origin of sense perception [i.e., conscious experience]" and "the question of free will." The paper became a pivotal and enduring statement about the possibilities of scientific progress (e.g., the mathematician David Hilbert was still disputing its claims in a 1930 talk on BBC radio). In his talk,

Du Bois-Reymond identified a subset of the enigmas that he claimed were "transcendent," meaning that they could *never* be solved by science, such as the origin of motion and the origin of consciousness (of these, he famously said "*ignoramus et ignorabimus*," "we do not know and we will not know"), versus those that seemed in principle scientifically resolvable.

Among Du Bois-Reymond's seven enigmas, the fourth was "the apparently teleological arrangement of nature," by which he meant the fact, observed since Aristotle, that the remarkable adaptiveness of organisms' features appeared "inconsistent with the mechanical view of nature" that Du Bois-Reymond himself strongly championed. This particular enigma, however, Du Bois-Reymond did not classify as "transcendent" and beyond science's reach despite the common attribution of the teleology of biology to God. The reason he considered it scientifically resolvable was the recent theory of Charles Darwin: "This difficulty is, however, not absolutely transcendent, for Mr. Darwin has pointed out in his doctrine of natural selection a possible way of overcoming it, and of explaining the inner suitableness of organic creation to its purposes and its adaptation to inorganic conditions … by a kind of mechanism in connection with natural necessity." Du Bois-Reymond not only classified the adaptiveness of biological features as one of the seven most fundamental mysteries in all the sciences but also acknowledged that for a nontheist, it appeared resolvable only because of the theory of natural selection, concluding, "Thus the fourth difficulty is no longer transcendent when it is earnestly, thoughtfully met."

It is this fundamental millennia-long scientific challenge of explaining the design-likeness of organisms that warranted adding "biological function" as an *ontological* rather than sheerly verbal-convenience category to the vocabulary of science and searching for this category's explanatory essence. That essence turned out to be natural selection.

Can SE Functions Be Understood as Fixed-Interest CR Functions?

Cummins tries to make it seem like effect-sensitivity of causal explanation—which is another way of referring to design-likeness in which somehow the causal role of a feature is shaped by the effects that it causes—is just another relativistic context or interest like any other: "You can make an instrumental norm look like a Norm by privileging a particular goal-state, but this is still just instrumental normativity—hence, relativized normativity—thinly disguised. … If you want to account for ('capture') the function attributions—including malfunction and failure of function attributions—of evolutionary biologists talking about natural selection, you can probably get a pretty good fit by relativizing to fitness, in one way or another" (Cummins and Roth 2009, 83).

So, why not take Cummins's suggestion and transmute the lead of interest relativity into the gold of objectivity simply by fixing the relativity to a certain current interest or goal, such as fitness or statistical deviation, yielding (pseudo) objectivity without

reference to history? As Cummins admits, this strategy is just a cosmetic makeover that yields a make-believe objectivity that is a "thinly disguised" relativity, not the genuine objectivity that medical diagnosis requires. To take an analogy, if one believes that "good" relativistically means "good relative to a given culture's values," one cannot evade relativism and achieve value objectivity in any meaningful sense simply by stipulating that absolute good means "good relative to my culture." The question remains of whether there is an objective sense of "function" beyond the ad hoc maneuver of stipulating a fixed relational element.

As to Cummins's suggestion of using fitness as the objectifying interest, medicine was being pursued for over two millennia before the concept of fitness ascended to its current Darwinian explanatory perch, so that approach fails to get at what justifies the formation of the concept of "function." In any event, fitness or reproductive success in the current environment has complex relationships to biological design. As Murphy says, "Health and fitness are different concepts"; reducing fitness is not necessarily a disorder (if it was, then deciding not to have an additional child would be a disorder). Even if being blind or being unable to walk or being schizophrenic makes no difference to fitness and does not hinder reproduction in our current or future environment, nevertheless these conditions are objectively dysfunctions, that is, failures of parts to perform functions they were biologically designed to perform. Nor can one say that CR functions in the medical sense are those that support health, because health is absence of disorder, and disorder is harmful dysfunction, and dysfunction is—according to the present proposal—any component's causation of a reduced capacity for health, and thus there is a vicious circularity. In sum, you just can't get from CR functions to the objectivity of medical diagnosis that Murphy demands.

Can CR Functions Account for the Objectivity of Dysfunction?

How, then, does Cummins apply his CR-function approach to medical dysfunction and disorder, where there is a failure of biological function? Here is how: "Systematic accounts relativize failures to function properly to a target explanandum: component x is failing to function properly, relative to a capacity C of the containing system S, if (other things equal) S fails to have C (or has a relatively diminished capacity) because of what x is doing" (Cummins and Roth 2009, 79). They explain, "Thus, systematic accounts allow for a kind of relativized or instrumental normativity: what something needs to do for the containing system to exercise the target capacity" (79).

Thus, if one is interested in the effect of a certain infection (such as the "brain-eating" amoebic infections), and that type of infection almost always causes death, but in 1 out of 100 cases the infected individual survives due to idiosyncratic immunologic features, then in that case, the researcher would, according to Cummins, be justified in saying that the infected individual suffered from an immunological dysfunction that

caused him or her to survive the infection. This view of CR dysfunction is a *reductio ad absurdum* of Cummins's view of functions as a plausible foundation for medicine.

Cummins further addresses the CR account of dysfunction in a discussion of blindness. Blindness is a medical disorder because the eyes are not capable of doing what they are biologically designed to do—that is, they are not capable of performing their natural function. However, Cummins and Roth (2009) state that the CR-function view cannot go along with this simple description because it cannot refer to natural functions that are by their nature historical, and the essence of CR functions is to eschew all reference to history:

> A perhaps more serious objection is that this sort of instrumental normativity—viz., the you-ought-to-do-x-to-achieve-g sense—will not accommodate the fact that a blind person's eyes are still for seeing. . . . The objection is that since the eyes of blind people never perform the function of enabling sight, systematic accounts should deny that a blind person's eyes are for seeing (and, thus, deny that the eyes are not functioning properly). To us, this appears to rely on a type–token ambiguity. Eyes generally (the type) enable seeing. A blind person's eyes (here, the token) are not for anything in that individual. . . . The sense that the eyes of a blind person are for seeing is simply the recognition that other humans do see, and that the eyes are an essential part of the human visual system. Thus, the blind person's eyes are not functioning properly (assuming here that the problem is really with the eyes) because they are not functioning in the way required for humans to see. (Cummins and Roth 2009, 79; a similar approach is taken by Boorse 2002)

These comments make it clear that CR functions, when applied to medicine, are not only not objective, but because they are ahistorical, they also must be essentially statistical and/or value based. Yet, the statistical and value views are precisely the views that must be rejected because they cannot answer the antipsychiatric challenge that has been faced by psychiatry for over a half century and would leave psychiatry without a solid conceptual foundation. The Soviet dissidents and runaway slaves and sexual Victorian women were statistically deviant and disapproved of and disvalued in their social contexts, yet were not disordered. Today, there are debates over many conditions as to their diagnostic status, yet the conditions are statistically deviant and disvalued (e.g., ADHD-type rambunctious behavior in schoolchildren). The CR-function approach simply abandons the project of making sense of a coherent and nonoppressive psychiatric medical specialty, which is the point of conceptual analysis of "medical disorder," to chase dubious philosophical ideological will-o'-the-wisps. The application of CR functions as a basis for medical diagnosis is a thoroughly value-laden approach that leaves no ultimate scientific ground for disputing interest-driven function claims and thus has nothing to say about opposed diagnostic judgments grounded in different interests. In contrast, SE functions offer a solid value-free scientific foundation in evolutionary theory and factual claims about biological design. To call the CR approach more scientific than the SE approach is without foundation.

Bock and von Wahlert and the Biological Origins of the CR- versus SE-Functions Distinction

One might ask: If Murphy's quest for an objective sense of CR dysfunction and the rejection of SE function cannot find a plausible foundation in Cummins's work, is such a rationale provided by some of the other prominent proponents of CR functions? I consider two of the most cited papers defending CR functions to show that they both accept SE functions and offer no solution to the medical objectivity challenge to CR functions.

Murphy heavily cites Amundson and Lauder's influential (1994) paper (which I will consider in due course), which in turn relies heavily on a classic paper by evolutionary biologists Walter Bock and Gerd von Wahlert (1965). That paper, despite not using the "CR" terminology (which, however, I will use in describing it), can be considered the origin of the contemporary CR-function notion. I believe that Bock and von Wahlert's position sometimes has been misunderstood as supporting Cummins's position, so I take some time to comment on it.

Bock and von Wahlert start from the position that there are two common uses of "function" in biology that need to be disambiguated: "A major source of ambiguity stems from the several meanings of function in biology. Function is used in the sense of the physical and chemical properties of the feature and in the sense of the role the feature has in the life of the organism. A review of the literature of functional anatomy will reveal that both meanings are employed. … We feel that these two concepts must be separated sharply" (1965, 276–277).

In constructing their terminology, Bock and von Wahlert choose to use the term "function" to label all of a feature's physiological causal actions on any aspect of the organism: "Basically the function of a feature is its action or how it works … which include all physical and chemical properties arising from its form (i.e., its material composition and arrangement thereof)" (274). This use of "function" is emphatically independent of any selection-related or other teleological implications: "We wish to stress that the definition of function as given above does not involve any aspect of purpose, design, or end-directedness. Moreover, this definition of function is free, as it should be, of any form of teleology" (274).

As Amundson and Lauder (1994) point out, Bock and von Wahlert's treatment of CR function diverges from Cummins's approach in that there is no interest relativity of CR functions in their account. This is because they do not relativize their CR functions to a given analysis but rather simply encompass within that category every possible action under every possible circumstance, including highly artificial stressors that would not occur in a natural habitat. Otherwise, Bock and von Wahlert and Cummins are on the same page: "Apart from the issue of unutilized functions, Cummins's concept of function matches the anatomists'" (1994, 450).

In contrast to "functions" so defined, Bock and von Wahlert (1965) label as "biological roles" the traditional SE functions directly linked to natural selection: "The biological role of a faculty … may be defined as the action or the use of the faculty by the organism in the course of its life history. … Each biological role of the faculty is under the influence of a set of selection forces" (278–279); "An evolutionary adaptation is … formed by a biological role coupled with a selection force. … The interaction between the organism and its environment is through a couple formed by the biological role and the selection force" (296). The notion of biological role is closely linked to the central explanatory puzzle of biology, the organism's adaptation to its natural environment, which Bock and von Wahlert agree is an SE-type evolutionary notion: "Clarity of meaning would be increased if the general term 'adaptation' were restricted to evolutionary adaptation" (285).

Bock and von Wahlert diverge from Cummins in considering SE function, or "biological role," to be a scientifically central sense of "function" linked directly to explanation by natural selection, with CR functions an instrumentally useful step toward that goal. They thus consider CR and SE functions as complementary and not in competition. Their analysis implies that they would have rejected outright anything like the attempted purge of SE functions from biomedical theory that Cummins and Murphy propose. They allow that the SE-function "biological role" usage is traditional: "We agree that many biologists formulate functional statements in a teleological framework. … Most workers discussing this problem probably use the term function in the sense of biological role" (274–275).

The functions that do not correspond to biological roles are considered nonutilized functions, and they, unlike functions with biological roles, have no SE functions: "Some of these faculties would be non-utilized ones corresponding to the non-utilized functions. Each utilized faculty of a feature is controlled by a different set of selection forces and hence each would have a separate evolution" (276). Thus, Bock and von Wahlert do not accept the Cummins zero-sum approach that SE functions are dispensable and "function" just means CR function. For them, SE functions are a central biological concept.

The distinction between function and biological role yields a division of labor within biological research between those doing laboratory work that may explore functions in dimensions never seen in the wild and those studying the natural life history of the organism:

> The function of a feature may be studied and described independently of the natural environment of the organism as is done in most studies of functional anatomy. The animal is placed in an experimental device which allows ascertainment of the functions of the feature with various degrees of precision. But the conditions are almost always highly artificial. In addition to these studies of pure function are investigations of biological anatomy in which the "function" (= biological role) of a structure is studied with the animal living freely in its natural habitat. Both types of studies are required to obtain different, but related sets of information which are prerequisites for the study of adaptation. (274)

The last sentence of this passage underscores that Bock and von Wahlert understand that the essential scientific challenge of biology concerns teleology and that, even for them, the study of CR functions ultimately is aimed at elucidating SE functions linked to adaptation and natural selection. By formulating the notion of CR functions, Bock and von Wahlert want to make the science of SE functions more effective. They argue that the focus on CR functions is valuable due to its epistemological usefulness in the pursuit of SE functions. Rather than the usual procedure of identifying a proposed SE function and then following it back to the causal actions of a feature that has that effect as its function, one can start with all "functions" (in the sense of causal actions) of the organism's features, devoid of premature teleological assumptions, and follow those effects forward, thereby discovering unsuspected biological roles: "Usually the biological roles of a feature are the guides to the function that are studied; however, this procedure may hinder the clarification of important functions of the feature which may be utilized in some or all cases" (n. 1, 274).

Thus, Bock and von Wahlert's point is not to eliminate or even downplay the importance of naturally selected biological roles of features. Rather, it is to provide a framework that allows biologists to be optimally open-minded in discovering biological roles by actually seeing which CR functions yield biological roles rather than either assuming from a CR function what the biological role must be ("the biological role of a feature cannot be predicted with any certainty from the study of the form and the function of the feature" [278]) or reasoning back from plausible biological roles to what CR functions are for.

Contrary to the usual impression, for Bock and von Wahlert, the analysis of CR functions is ultimately about identifying SE functions. But, what, then, is the importance of studying nonutilized CR functions, even ones that go beyond anything that occurs under usual conditions in the wild? The primary answer lies in Bock's interest in preadaptation, where, in response to changing environmental pressures, nonutilized CR functions are exploited in novel ways to support new biological roles. Formerly defined as "a structure is said to be preadapted for a new function if its present form which enables it to discharge its original function also enables it to assume the new function whenever need for this function arises" (292), they suggest that the definition can now be reformulated more clearly as "a feature is said to be preadapted when its present forms and functions (both utilized and non-utilized ones) allow … [it] to acquire a new biological role … whenever the need (= appearance of the selection force) for this new adaptation … should arise" (292). They explain that, as a result, "preadaptation should not be construed of as a change in functions as has been expressed in earlier papers, but as a change in biological roles. With the origin of a new adaptation, a new selection force acts upon the feature" (292).

Bock and von Wahlert explain that discovery of biological roles justifies concern about CR functions: "A worker may consciously or unconsciously study functions of features that never occur during the life of the organism. … These non-utilized

functions cannot be ignored because we generally do not know which functions are utilized and which are not utilized by the organisms and because the non-utilized functions form an important basis of the phenomenon of preadaptation" (274). Bock and von Wahlert's notion of preadaptation is rightly equated by Amundson and Lauder with Gould's notion of "exaptation," but their analysis possesses a crystal clarity about the close relationship between preadaptation and natural selection that stands in sharp contrast to Gould's befuddled claims about exaptation somehow undermining natural selection explanations. (For discussion of Gould's confusions, see Wakefield 1999a, 2016a.)

There can be no solace in Bock and von Wahlert's analysis for Cummins's and Murphy's "competition" view of the relationship between CR and SE functions in biology or for Cummins's and Murphy's claims for the priority of CR functions. It is clear that Bock and von Wahlert's article is intended as a corrective to overly simplistic approaches to SE functions that they agree are at the center of biological theorizing and not at all as a critique of SE functions.

Amundson and Lauder's Defense of CR Functions

Murphy several times cites an influential article by Amundson and Lauder (1994), "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology," in support of his position that psychiatric diagnosis can be based wholly on CR functions. Amundson and Lauder's position in turn rests on Bock and von Wahlert's (1965) influential argument considered above but moves beyond it to address the CR versus SE function debate. Their paper has become the locus classicus of the defense of CR functions in biology, so I examine whether or not it supports Murphy's position.

The title of Amundson and Lauder's paper looks promising from Murphy's perspective because it seems to suggest that, even within natural selection's citadel of evolutionary biology, the CR formulation may hold sway. However, it turns out that the title alludes to a more modest claim. They want to dispute not the importance or necessity of SE functions but the claim that CR functions are eliminable in favor of SE functions or somehow subordinate to SE functions. Their paper does not primarily focus on medical diagnosis and is not a repudiation of the fact that there are domains within biomedical theory in which SE function is the primary function concept. They acknowledge that even functional anatomical studies are often guided by the desire to understand SE functions: "Functional anatomists typically choose to analyze integrated character complexes which have significant biological roles" (450).

When they do discuss medically related functional notions, Amundson and Lauder take a position directly opposed to Murphy's. They assert unequivocally that "purpose and dysfunction" are "concepts to which CR function doesn't apply" (451). That is, CR functions are objectively neutral between health and disorder.

Amundson and Lauder consider a related objection to CR functions that they call "the problem of pathological malformations of functional items" (452–453). The background to this discussion is that SE theorists have attempted to outflank CR-function arguments by claiming that categories of anatomical features are themselves SE-defined concepts, so that one cannot even conceptually define a type of organ (e.g., a heart) for CR functions without first specifying an SE-type natural function (e.g., pumping blood) that it was biologically designed to perform and that defines its category. That is, hearts just are organs for pumping blood, so SE functions logically precede anatomical generalizations that use CR functions. SE theorists generally argue that a major goal of a theory of function is to explain dysfunction, and that poses the problem of how we recognize pathological specimens (or for that matter radically different cross-species specimens) as the organs that they are, and the functional account is claimed to resolve this problem.

> According to these theorists, only SE function can categorize parts into their proper categories irrespective of variation and malformation. It does so by defining "function categories." CR function (like other non-historical theories) cannot define appropriate function categories, and so is unable both to identify diseased or malformed hearts as hearts, and to identify the same organ under different forms in different species.
>
> On pathology, Millikan points out that diseased, malformed, and otherwise dysfunctional organs are denominated by the function they would serve if normal. "The problem is, how did the atypical members of the category that cannot perform its defining function *get* into the same function category as the things that actually can perform the function?" A CR analysis of a deformed heart which cannot pump blood obviously cannot designate its *function* as pumping blood, since it doesn't have that causal capacity. On the other hand, even the organism with the malformed heart has a selective history of ancestors which survived because *their* hearts pumped blood. So the category "heart" which ranges over both healthy and malformed organs must be defined by SE, not CR, function. (453)

Amundson and Lauder argue that "SE functionalists are simply mistaken in this claim" that evolutionary history is conceptually prior to functional anatomy. They claim that rather than either SE or CR functions, "the classifications come from a third, non-functional source" (453). That source consists of anatomical, morphological, and histological evidence that allows identification of organs within and often across species, and it can identify both functional and dysfunctional instances of an organ:

> Even a severely malformed vertebrate heart, completely incapable of pumping blood (or serving any biological role at all), could be identified as a heart by histological examination. … Anatomical categorizations of biological items already embrace interspecies and pathological diversity without any appeal to purposive function. Anatomical distinctions are not normally based on CR function *either*, to be sure. Functional anatomists *per se* do not categorize body parts. Rather they study the capacities of anatomical complexes which have already been categorized by comparative anatomists. Causal role functional anatomy proceeds unencumbered by demands to account either for the categorization or the causal origins of the systems under analysis. (457–458)

Murphy reiterates Amundson and Lauder's point that "it is entirely possible to identify anatomical units by anatomical considerations, regardless of proper function … on morphological and physiological grounds, regardless of history." On this point I agree; Amundson and Lauder do persuasively correct SE theorists' conceptual overreach in suggesting that anatomy must be based on SE-function categories. After all, anatomists were identifying hearts from obvious morphological features long before Harvey's discovery of the heart's function.

However, there is a serious problem with the sheer morphological solution that Amundson and Lauder downplay by placing it in a note. In discussing the problem of identifying hearts across very differently structured species in which histological, morphological, and homological comparisons may break down, Amundson and Lauder admit that there are limits to their account that leave an opening for the functional analysis:

> There is one felicitous application of Neander's claim about the inadequacies of morphological criteria to designate hearts. Since the category `heart' is used across major taxonomic differences, a vertebrate taxonomist unfamiliar with mollusks might well not be able to use *vertebrate* morphological criteria to identify a *molluscan* heart. And, to get only slightly bizarre, it is possible to imagine discovering a new taxon of animals which has organs functionally identifiable as hearts, but which fit the morphological criteria for hearts of no known taxon. We agree with the SE functionalist's point in this rather limited set of cases. (n. 4, 467)

This note appears to directly contradict Amundson and Lauder's contention that hearts are strictly anatomically identifiable across species and malformations. However, I believe one should resist fleeing back to the SE theorists' claims. Amundson and Lauder's morphological argument remains compelling within a certain sphere. Moreover, the argument that hearts are by definition blood pumpers comes dangerously close to making it a conceptual truth that the function of hearts is to pump the blood, whereas that is one of the greatest empirical discoveries in the history of physiology. In addition, the problem of recognizing mollusk hearts may not lead quite as straightforwardly back to the SE position as it might seem. Davies (2001) argues that the SE approach to organ identification fares no better than the CR approach in regard to including malfunctions within the appropriate category because the functional category depends on a definition in terms of the evolutionary history of hearts that did pump adequately. Thus, it is unclear how pathological hearts that do not pump adequately and did not contribute to that history get into the historically relevant set: it seems possible that a careful definition of the historically determined set might overcome this obstacle, but Davies's point is that there is in fact a certain arbitrariness to the decision as to how to draw the boundaries of that functionally defined historical kind and so an arbitrariness as to whether it includes malfunctioning instances (see Allen and Neal 2019; Sullivan-Bissett 2017).

To thread this needle, the account of organ categories may have to be more subtle than either a simple SE or simple morphological account allows. A compromise

conceptualization using a black-box essentialist approach (Wakefield 1999b, 2000a, 2004) may be able to resolve these issues in a way that explains the intuitions of both SE and CR theorists. One can take seriously Amundson and Lauder's claim that "even a severely malformed vertebrate heart, completely incapable of pumping blood (or serving any biological role at all), could be identified as a heart by histological examination," while realizing that this is true of *some* deformed or malfunctioning hearts but not all (e.g., Amundson and Lauder's example of a deformed heart displaced to the knee). What the mollusk example shows is that despite the powerful anatomical considerations that can be brought to bear in recognizing hearts that are normal and abnormal or across species, there is a limit to that approach where we can be genuinely surprised that something not initially recognizable as a heart is indeed a heart on functional grounds. This sort of surprising extension of a theoretical category to new instances that do not share superficial features with standard cases is what the black-box essentialist account is designed to explain. Such an account would allow that we can initially identify a base set of hearts—normal and abnormal—by the morphological criteria described by Amundson and Lauder and use that morphologically defined base set to define a broader functional category. Thus, by "heart," we mean any organ across species that has the same natural function—defined in relation to the process of natural selection that led to the presence of *those* organs, both normal and abnormal—as the base set of hearts identified morphologically and histologically in the human species.

This black-box essentialist approach to organ definition allows that morphological criteria are sufficient for recognizing a base set of hearts in humans and analogous creatures without reference to SE functions, and this allows functional anatomy to get under way with no reference to natural selection. It then appropriately becomes an empirical discovery rather than a conceptual truth that the base set of human hearts has the natural function of—and, after Darwin, was naturally selected for—pumping blood, implying that hearts in general are organs naturally selected for pumping blood. These discoveries are not part of the concept of heart but are what is referred to indirectly in saying that "hearts are those anatomical parts with the same biological-design functional essence as the base set of morphologically recognizable hearts." The category of hearts can then be extended across species to very different creatures with hearts that are not at all morphologically like ours based on this understanding of the functional essence of hearts. And, contra Davies, this analysis gets both pumping and nonpumping instances of hearts into the base set and thus into the set to be explained by the essential SE function.

Can CR Functions Be Saved from the Promiscuity Objection?

A major objection to CR functions as interest-relative causal roles is that such attributions would seem to be applicable across the sciences, yet such function attributions

commonly occur only in the few scientific domains with SE functions, namely, the biopsychosocial, medical, and artifactual domains. This apparent mismatch between the predicted "liberality" or "promiscuity" of CR functions and actual scientific practice suggests that CR functions in fact depend on a background framework of biological design, as, for example, Philip Kitcher (1993) argues:

> Without recognizing the background role of the sources of design, an account of the Cummins variety becomes too liberal. Any complex system can be subjected to functional analysis. Thus we can identify the 'function' that a particular arrangement of rocks makes in contributing to the widening of a river delta some miles downstream, or the "functions" of mutant DNA sequences in the formation of tumors—but there are no genuine functions here, and no functional analysis. The causal analysis of delta formation does not link up in any way with a source of design; the account of the causes of tumors reveals *dysfunctions*, not functions. (Kitcher 1993, 390)

Similarly, Millikan (1989) points out that the contributions of clouds to the rain cycle should qualify for CR functions, and Neander (1991) observes that the plate tectonics yielding earthquakes satisfy Cummins's criteria, yet clouds and tectonic plates don't literally have the natural function of producing rain or earthquakes.

Davies (2001) tries to save CR theory from falsification by arguing that SE domains are the only ones that are hierarchically organized, and CR functions are applied only to hierarchical organizations. However, function language was used long before a modern understanding of hierarchical systems, and in any event, there are subdomains within the SE domains that are not hierarchical but to which function language is still applied (see below).

Amundson and Lauder (1994) admit that "whimsical 'functional analyses'" of meteorological or geological systems "are indeed counterintuitive results" (448, 452) but initially dismiss the problem because "the criticism simply does not apply to the real world of scientific practice" (452), a response with which Murphy expresses some sympathy. This of course misses the point. The CR account of function predicts that scientists *should* be making these attributions, so the lack of such attributions—the very fact that they are "whimsical"—falsifies the CR analysis, rendering the CR theory itself whimsical.

Amundson and Lauder have a more substantive response to the promiscuity objection. Cummins puts forward criteria for what makes a good or interesting functional analysis, such as that the system is of interest, the analyzing capacities are simpler or different in type from the analyzed capacities, and the system is complex. Amundson and Lauder argue that the whimsical functions do not meet Cummins's criteria and thus do not occur: "By Cummins's own evaluative criteria (and given the facts of the real world) functional analyses of these systems would have no interest. Analyzing capacities would not be significantly simpler or different in type from analyzed capacities (are plate movements simpler than earthquakes?) nor would the system's

organization be notably complex. … All of the interesting causal role functions have a history of natural selection. … Earthquakes and rainfalls … have no such history, and so no complex functional organization" (452).

The claim that causal analyses of earthquakes and the water cycle are of no interest and are not complex is implausible, to say the least. But, there is a more basic problem. To shore up their argument, Amundson and Lauder demonstrate how Cummins's guidelines for good functional analyses apply to a specific case:

> In a valuable functional analysis, the analyzing capacities will be simpler and/or different in type from the analyzed, and the system's discovered organization will be complex. Suppose the capacity to crush of the hypothetical jaw derives from the extremely simple fact that objects between the two bones are subjected to the brute force of muscle X forcing the bones together. Here the "organization" of the system is almost degeneratively simple, and the force of the muscle hardly simpler or different in kind from the crushing capacity of the jaw. A functional analysis of very low value. On the other hand, suppose that the jaw is a complex of many elements, muscle X is much weaker than the observed crushing capacity, the crushing action itself is a complex rolling and grinding, the action of muscle X moves one of its attached bones into a position from which the bone can support one of the several directions of motion, and that this action must be coordinated with other muscle actions so that it will occur at a particular time in the crushing cycle. Here X's function is much simpler than the analyzed capacity, is different in kind (moving in one dimension in contrast to the three dimensional motion of the jaw) and the organization of components which explains jaw action is complex indeed. A functional analysis of high value. (1994, 451)

This illustration inadvertently but decisively refutes Amundson and Lauder's reply to the promiscuity objection by revealing that function attributions in science do not at all depend on Cummins's criteria for "good functional analysis." In the example, Amundson and Lauder explain that the simple jaw's mastication muscle lacks all of the Cummins-specified properties, thus yielding a bad functional analysis or not being amenable to functional analysis. What they fail to observe is that that muscle still would be described unequivocally by any biologist—or nonbiologist—as having the biological function of enabling the chewing of food, and would be considered to have this function to no lesser degree than the muscle in the example of a "good" functional analysis. Consequently, the claimed low-value status of functional analyses in other scientific disciplines fails to explain away the lack of function attributions, and the promiscuity objection remains unaddressed.

For Amundson and Lauder's response to be persuasive, every function attributed to an organic or artifactual system must satisfy Cummins's "goodness" criteria to a greater degree than any system in any other domain, a claim that is not credible. For example, there is a little extendable piece of plastic on the output tray of my printer that is linked to absolutely nothing else, yet clearly has the function of stopping the pages of larger printed files from falling to the ground. This analysis does not fulfill any of the "good"

complexity criteria, but it is still a clear instance of an item having a function because the aforementioned part is clearly designed to contribute to the printer's designed purpose of providing a convenient and clean printed outcome. There are many simple biological mechanisms as well that work directly to fulfill their attributed SE function and are less complex than the workings of plate tectonics or meteorological phenomena but can be recognized as designed, as Amundson and Lauder seem to acknowledge: "Given a simple trait with a known biological role, the evolutionist might feel justified in ignoring anatomical details" (450). In any event, even if one absurdly dismisses all the rest of science as being of less interest or not as systematically complex as biological systems, one cannot do the same with medical pathology. Pathological conditions often depend for their existence on systemic relationships every bit as complex as those in health, implying that we should be attributing functions—rather than dysfunctions—to disordered parts, contrary to scientific reality. The claim that functions have nothing to do with design and that it just so happens that the only systems complex enough to warrant function attributions are designed ones is just as ad hoc and spurious as it seems.

The promiscuity objection thus remains a perfectly effective objection that casts doubt on CR functions. I suspect that Cummins understands this situation better than his defenders do, for rather than trying to concoct a rationalization to explain why CR functions are not attributed in most domains of scientific research, Cummins, as we saw, accepts such counterintuitive cases as legitimate function attributions and suggests that we jettison out intuitions along with our standard scientific practices—an extraordinary degree of revisionism on which Murphy has no comment.

Biological Design as the Target Explanandum for Function Attributions

The black-box essentialist analysis of function addresses a further much-debated question, namely, the nature of the relationship between the concept of a natural function and Darwin's theory of natural selection. For convenience, in this reply, I have used the standard "SE" abbreviations for selected effects functions, but this misleadingly suggests that "function" just means "selected effect." Both Neander (1991) and Millikan (1989), in well-known papers, claim on different grounds that "function" means "naturally selected effect." Neander argues that this is what biologists and others commonly mean today by "function," and Millikan says it is simply a theoretical definition that has little to do with ordinary usage. Cummins, too, writes as if the SE account proposes that "function" *means* or is *theoretically defined as* "naturally selected" and in his arguments exploits the various paradoxes and confusions that result.

However, as I have argued elsewhere (Wakefield 2000a), "function" cannot mean "naturally selected effect," for two reasons of a type familiar from the natural-kind/

essentialist literature: (1) biologists have understood the concept of a natural function going back to Aristotle and Harvey but had no idea of the theory of evolution, and (2) it is an empirical discovery that natural functions are naturally selected effects and it could have turned out—indeed, in principle, it could still turn out—otherwise (e.g., we might find that the theists are right and that natural functions are determined by the intentions of a Divine Creator rather than evolution). Darwin did not redefine the concept of function that had existed since Aristotle and Harvey; he explained that natural functions are in fact naturally selected effects.

So, precisely how is "function" linked to naturally selected effect? "Function" is a shared concept based on prototypical examples of intuitively nonaccidental (biologically designed) beneficial effects like sight and on the idea that some common underlying process or processes must be responsible for such remarkable phenomena. These are notions shared by Aristotle, Harvey, and us. It is a scientific discovery, not a conceptual truth, that functions exist because of natural selection. So, function is not directly linked to natural selection either by a conceptual analysis or by a theoretical definition. The link consists instead of two steps: first, a conceptual analysis that identifies functions—as understood by Aristotle, Harvey, and us—as effects that share an essential explanatory process with prototypical nonaccidental benefits like sight and, second, the modern discovery that the essential process is natural selection.

Surprisingly, the black-box essentialist view fits well with the typical descriptions of the empirical, conceptual, and theoretical situation in the history of biology even by those who also support the option of attributing CR functions. For example, Bock and von Wahlert (1965) aptly summarize what amounts to a black-box essentialist account of biological function, which holds that biological design—or what they refer to as adaptation—is a fundamental observation that already presupposes teleology and that motivates the search for an account of how there can be teleological functions:

> The idea of a close correlation between the features of living organisms and the conditions of their environment predates by many years the general acceptance of any theory of organic evolution by biologists. Pre-evolutionary biologists understood the general notion and many of the details of this correlation between organisms and environment as well as we do today; what they lacked was a solid scientific explanation of the how and the why of adaptation. Rather than the notion of adaptation being a consequence of the acceptance of organic evolution, the search for an explanation of these observations was a major impetus in the development of a scientifically acceptable theory of organic evolution. (282–283)

Similarly, when discussing the homology of structures across species as the foundation of functional anatomy that does not rely on SE functions—instead relying on indicators such as similarity in structure, identical connections or position within an overall structural pattern, and common developmental origin in the embryo—Amundson and Lauder offer an analysis that fits well with the black-box essentialist analysis:

> Comparative anatomy, morphology, and the concept of homology predate evolutionary biology. They provided Darwin with some of the most potent evidence for the fact of descent with modification. … So the evolutionary definition of homology mentioned above is a theoretical definition. As with other theoretical definitions, it is subject to sniping from practitioners of conceptual analysis. A philosopher could argue (pointlessly) that "homology" cannot mean "traits which characterize monophyletic clades," since many 1840s biologists knew that birds' wings were homologous to human arms but disbelieved in evolution (and so disbelieved that humans and birds shared a clade). SE advocates' usual reply to the William Harvey objection is applicable here. Just as Harvey could see the marks of biological purpose without knowing the origin or true nature of biological purpose, preDarwinian anatomists could see the marks of homology without knowing the cause and true nature of homology itself. (1994, 454–455)

Amundson and Lauder here imagine a conceptual objection to the standard Darwinian nonfunctional definition of homology as common derivation, namely, that homology cannot mean anything that presupposes Darwinian theory of descent with modification because the concept of homology was understood and homologies were recognized by biologists well before Darwin wrote. Despite Amundson and Lauder's dismissive remarks about conceptual analysis (which flow from their earlier rebuttal of SE theorists' incorrect conceptual-analytic arguments), they seriously respond to this objection, revealing that the conceptual analyst's "sniping" is not all that "pointless" after all. They respond, "Just as Harvey could see the marks of biological purpose without knowing the origin or true nature of biological purpose," so "pre-Darwinian anatomists could see the marks of homology without knowing the cause and true nature of homology itself." Thus, what Amundson and Lauder, following Millikan, label the Darwinian "theoretical definition" of homology is not a definition at all but the empirical discovery of a hypothesized essential nature of a previously recognized phenomenon. The black-box essentialist analysis of "homology" would fit this account quite well. But, more to the point, their characterization of Harvey's recognition of biological design without knowing its Darwinian "true nature" or essence acknowledges what is best understood as a black-box essentialist structure to the notion of biological design.

Peter Godfrey-Smith (1993) versus the Harmful Dysfunction Analysis: Reply to the Editors

Finally, I need to briefly address support for Murphy's position that comes from an unexpected quarter. Despite my boundless gratitude to the editors of this volume, Denis Forest and Luc Faucher, I do want to disagree with them on one important claim they make in their introduction to this volume. In the course of discussing Cummins's causal-role (CR) model of function, they assert that "it is not impossible to derive an account of dysfunctions from this view of functions (Godfrey-Smith 1993)," thus agreeing with Murphy and citing Peter Godfrey-Smith's influential paper in support of their

assertion. I too admire the clarity, perspicuity, and depth of insight that Godfrey-Smith consistently brings to topics in philosophy of biology and philosophy of science more generally. However, in this case, the editors' endorsement of Murphy's claim overlooks what I believe is the entirely vacuous nature of Godfrey-Smith's claim.

Here is what Godfrey-Smith says:

> On Cummins analysis, functions are not effects which explain why something is there, but effects which contribute to the explanation of more complex capacities and dispositions of a containing system. Although it is not always appreciated, the distinction between function and *malfunction* can be made within Cummins' framework, as well as within Wright's. If a token of a component of a system is not able to do whatever it is that other tokens do, that plays a distinguished role in the explanation of the capacities of the broader system, then that token component is malfunctional. The concept of malfunction is context dependent on Cummins' view, just as the concept of function in general is. (Godfrey-Smith 1993, 200)

Of course, one can always arbitrarily *stipulate* some interest-relative meaning of "malfunction" within a CR-function perspective. For example, one can stipulate that by "malfunction," one will mean whatever is statistically unusual, or whatever does not bring about the outcome that one is interested in studying, or whatever one doesn't want to happen. However, the standard problem in the function literature is whether, given that Cummins-type functions are explicitly relative to the interests of the observer (or "context dependent," as Godfrey-Smith describes it), there is an account of malfunction definable within the Cummins-style view of function that has a coherent relationship to the standard objective meaning of malfunction within psychiatric and medical diagnosis.

The problem is that the primary goal in characterizing malfunction or dysfunction is to make sense of medicine, and as we have seen above with various deployments of the CR notion, Godfrey-Smith's characterization of CR malfunction has no relationship to what we mean by malfunction in medicine. Godfrey-Smith states that a malfunction occurs when "a component of a system is not able to do whatever it is that other tokens do." This claim can be interpreted in various ways. It might mean that there is a malfunction in a component whenever there exists *any* other instance of the component that can do *anything* that the target component cannot do. However, on this interpretation, every instance of every mechanism would be "malfunctioning" simply in virtue of normal variation. Moreover, a cancerous cell can do things that a normal cell cannot do, so on this broad interpretation of Godfrey-Smith's criterion, the normal cell would be malfunctioning. Indeed, any genetic mutation that can cause disease would thereby render the normal genes malfunctioning.

One might try to fix this problem by more plausibly interpreting Godfrey-Smith as claiming that an instance of a component is malfunctioning only when it cannot do something that the *majority* of instances of the component can do. This is a common strategy among those trying to defend the CR function analysis. However, this notion of malfunction would be based purely on statistical deviance. If there is one thing on

which there is a near-consensus among philosophers of psychiatry, it is that a purely statistical definition of dysfunction and disorder does not get at our intuitive medical concepts. Such an account does not adequately distinguish normal variation from malfunction, excellence from pathology, epidemics and statistically common disorders from normality, and so on. (The exception is Christopher Boorse's [1977] biostatistical theory, which does place its faith in statistical deviance and fails for this reason among others; see my reply to Lemoine in this volume.)

However, such statistical or technical attempts to save Godfrey-Smith's claim are beside the point due to a more basic problem, revealed at the end of the above passage: "The concept of malfunction is context dependent on Cummins' view, just as the concept of function in general is." That is, Godfrey-Smith acknowledges that because any capacity or disposition of a system that is of interest can be the target of a CR-functional analysis, the notion of CR malfunction must be interest relative as well. This is a claim that even CR defender Murphy (this volume) now belatedly questions as conflicting with the objectivity of medical dysfunctions.

As we have seen, Cummins himself (Cummins and Roth 2009) accepts that on his interest-relative systems or CR approach to function, dysfunction must be defined in interest-relative terms: "Systematic accounts relativize failures to function properly to a target *explanandum*. … Thus, systematic accounts allow for a kind of relativized or instrumental normativity: what something needs to do for the containing system to exercise the target capacity" (2009, 79). Cummins's analysis thus gives up any pretense to explaining the objectivity of medical diagnosis. Whether your heart ceasing to pump is a dysfunction in the medical sense is not dependent on your interests. Even if you are suicidal or tired of life, the heart's lack of pumping is a medical dysfunction. If you are a medical researcher and the capacities of certain genes or bodily conditions to cause diseases are what interest you (e.g., the capacity to form new tumors in certain forms of cancer; the capacity of the heart to cause edema in congestive heart failure), then whatever bodily component actions lead to that outcome will have that outcome as their CR functions, and within that context, the failure to cause the tumor formation or edema, respectively, becomes a malfunction. This implication constitutes a clear *reductio ad absurdum* of the CR notion of malfunction put forward by Godfrey-Smith, at least as a notion of malfunction relevant to medical diagnosis.

So, if the question is how one can start from CR functions and get to a conceptually adequate account of malfunction in the objective sense relevant to medicine, the answer is that you can't get there from here. The interest relativity of CR functions and CR dysfunctions, acknowledged by Godfrey-Smith as well as Cummins, makes it impossible for any attempted definition of CR malfunction to approximate the objective nature of the intuitive medical notion of malfunction.

In the same paper, Godfrey-Smith (1993) offers a way out of the problem posed by his passage above if "malfunction" is interpreted in its medical sense. He famously insists

that CR and SE functions are two different kinds of functions that are dominant in different subdisciplines of biology, and they are not to be confused even while both are to be accepted as legitimate: "We should accept both senses of function, and keep them strictly distinct. All attempts to make one concept of function work equally for behavioral ecology and physiology are misguided. On this view, 'Wright functions' and 'Cummins functions' are both effects which are distinguished by their explanatory importance. The difference is in the type of explanation" (Godfrey-Smith 1993, 200–201).

Suppose that for the sake of argument, we accept Godfrey-Smith's famously proposed "consensus" that both SE and CR forms of functional explanation are real and legitimate ways of identifying functions. This consensus implies that the relevant question is not which approach is the superior one to adopt for all of the biomedical sciences (this is the unfortunate way that Murphy approaches the issue). Rather, the relevant question is which of the two approaches to function attributions, SE or CR, is the appropriate one specifically for the function and dysfunction attributions underpinning medical and psychiatric diagnosis of disorder versus normal variation. This approach obviates the need for a forced answer to how the CR approach can define malfunction, because malfunction in the medical sense can be a domain in which SE functions are the more appropriate of the two approaches.

Conclusion

Murphy accepts that medical judgments of dysfunction must be objective and acknowledges that CR functions on their face are not objective due to their interest relativity. So far, neither Murphy's own arguments nor Cummins's presentation nor the arguments and positions of the writers cited by Murphy—several of whom are biologists and ought to know—nor the biological literature itself support Murphy's contention that CR functions can and should supplant SE functions. Most important, there is no evidence in anything presented in Murphy's paper for his claim that CR functions can yield an appropriately objective sense of dysfunction of the kind that can support the practice of medical diagnosis in anything like the form it currently exists.

However, this does not entirely resolve the issue of whether such an explanation exists. Although putting forward no solution to this conundrum himself, Murphy in effect outsources the solution to the problem of dysfunction objectivity by citing philosophers Carl Craver and Marcel Weber as having put forward theories that can explain how CR functions can provide the foundation for objective medical judgments of dysfunction and disorder. Murphy thus suggests that my work "overlooks the extensive work done in recent years by philosophers of biology who have developed [the CR approach] into a causal-explanatory account."

Thus, to complete a fair assessment of Murphy's argument for CR functions as a foundation for medical diagnosis, it is necessary to go beyond the arguments laid out in

Murphy's chapter and directly consider—and not once again overlook—the work of the authors he cites as having the solutions he seeks. I therefore provide in a supplement to this reply a close examination of whether Craver's or Weber's analyses cited by Murphy provide the solution Murphy requires to warrant his CR-function approach to medicine.

References

Allen, C., and J. Neal. 2019. Teleological notions in biology. In *The Stanford Encyclopedia of Philosophy,* E. N. Zalta (ed.). https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=teleology-biology.

Amundson, R., and G. V. Lauder. 1994. Function without purpose: The uses of causal role function in evolutionary biology. *Biology and Philosophy* 9: 443–469.

Ap, T. 2016. This is what reading is like if you have dyslexia. *CNN Health.* https://www.cnn.com/2016/03/05/health/dyslexia-simulation/index.html.

Armstrong, T. 2015. The myth of the normal brain: Embracing neurodiversity. *AMA Journal of Ethics* 17(4): 348–352.

Artigas-Pallarés, J. 2009. Dyslexia: A disease, a disorder, or something else? *Revue Neurologique* 48(suppl. 2): S63–S69.

Blanco, D. B. 2019. A molecule long thought harmless plays a role in pancreatic cancer, could hint at cure. *Discover*, June. http://blogs.discovermagazine.com/d-brief/2019/06/20/a-molecule-long-thought-harmless-plays-a-role-in-pancreatic-cancer-could-hint-at-cure/#.XbN5WZNKgWo.

Bock, W. J., and G. von Wahlert. 1965. Adaptation and the form-function complex. *Evolution* 19: 269–299.

Boorse, C. 1975. On the distinction between disease and illness. *Philosophy and Public Affairs* 5(1): 49–68.

Boorse, C. 1976. Wright on functions. *Philosophical Review* 85: 70–86.

Boorse, C. 1977. Health as a theoretical concept. *Philosophy of Science* 44: 542–573.

Boorse, C. 2002. A rebuttal on functions. In *Functions: New Essays in the Philosophy of Psychology and Biology,* A. Ariew, R. Cummins, and M. Perlman (eds.), 63–112. Oxford University Press.

Cummins, R. 1975. Functional analysis. *Journal of Philosophy* 72: 741–765.

Cummins, R. 2002. Neo-teleology. In *Functions: New Essays in the Philosophy of Psychology and Biology,* R. Cummins, A. Ariew, and M. Perlman (eds.), 157–173. Oxford University Press.

Cummins, R., and M. Roth. 2009. Traits have not evolved to function the way they do because of a past advantage. In *Contemporary Debates in Philosophy of Biology,* F. Ayala and R. Arp (eds.), 72–86. Blackwell.

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

Du Bois-Reymond, E. 1882. The seven world-problems. *Popular Science Monthly*. https://en.wikisource.org/wiki/Popular_Science_Monthly/Volume_20/February_1882/The_Seven_World-Problems. April 25, 2018.

Dyslexia has a language barrier. 2004. *The Guardian*, September 22. https://www.theguardian.com/education/2004/sep/23/research.highereducation2

Dyslexia is not a disease. n.d. Davis Dyslexia Association International. https://www.dyslexia.com/question/dyslexia-is-not-a-disease/.

Eisenberg, D. T. A., B. Campbell, P. B. Gray, and M. D. Sorenson. 2008. Dopamine receptor genetic polymorphisms and body composition in undernourished pastoralists: An exploration of nutrition indices among nomadic and recently settled Ariaal men of northern Kenya. *BMC Evolutionary Biology* 8: 173.

Engle, D. D., H. Tiriac, K. D. Rivera, A. Pommier, S. Whalen, E. T. Oni, et al. 2019. The glycan CA19–9 promotes pancreatitis and pancreatic cancer in mice. *Science* 364: 1156–1162.

First, M. B., and J. C. Wakefield. 2010. Defining 'mental disorder' in *DSM-V*. *Psychological Medicine* 40(11): 1779–1782.

First, M. B., and J. C. Wakefield. 2013. Diagnostic criteria as dysfunction indicators: Bridging the chasm between the definition of mental disorder and diagnostic criteria for specific disorders. *Canadian Journal of Psychiatry* 58(12): 663–669.

Gluckman, P., A. Beedle, and M. Hanson. 2009. *Principles of Evolutionary Medicine*. Oxford University Press.

Godfrey-Smith, P. 1993. Functions: Consensus without unity. *Pacific Philosophical Quarterly* 74: 196–208.

Habib, M. 2000. The neurological basis of developmental dyslexia: An overview and working hypothesis. *Brain* 123: 2373–2399.

Halbrook, C. J., and H. C. Crawford. 2019. Hiding in plain sight: A common biomarker of pancreatic disease has a functional role in pathogenesis. *Science* 364(6446): 1132–1133.

Harvey, W. 1628/1993. *On the Motion of the Heart and Blood in Animals*. Trans. R. Willis. Prometheus Books.

Horwitz, A., and J. C. Wakefield. 2007. *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford University Press.

Hubel, D. H., and T. N. Wiesel. 1959. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology* 148: 574–591.

Hubel, D. H., and T. N. Wiesel. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160: 106–154.

Jorgenson, E., N. Matharu, M. R. Palmer, J. Yin, J. Shan, T. J. Hoffman, et al. 2018. Genetic variation in the SIM1 locus is associated with erectile dysfunction. *Proceedings from the National Academy of Sciences of the United States of America* 115(43): 11018–11023.

Josberger, E., P. Hassanzadeh, Y. Deng, J. Sohn, M. J. Rego, C. Amemiya, et al. 2016. Proton conductivity in ampullae of Lorenzini jelly. *Science Advances* 2(5): e1600112–21600112: 1–6.

Kareklas, K., D. Nettle, and T. V. Smulders. 2013. Water-induced finger wrinkles improve handling of wet objects. *Biology Letters* 9(2): 1–3.

Kingma, E. 2013. Naturalist accounts of mental disorder. In *The Oxford Handbook of Philosophy and Psychiatry,* K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (eds.), 363–384. Oxford University Press.

Kitcher, P. 1993. Function and design. *Midwest Studies in Philosophy* 18: 379–397.

Knoll, J. L., IV, and R. W. Pies. 2019. Cruel, immoral behavior is not mental illness. *Psychiatric Times*, August 19. https://www.psychiatrictimes.com/couch-crisis/cruel-immoral-behavior-not-mental-illness.

Lilienfeld, S. O., S. J. Lynn, J. Ruscio, and B. L. Beyerstein. 2010. Myth #17: The defining feature of dyslexia is reversing letters. In *50 Great Myths of Popular Psychology: Shattering Widespread Misconceptions about Human Behavior*, 37–38. John Wiley. https://www.psychologicalscience.org/media/myths/myth_17.cfm.

Lloyd, E. 2005. *The Case of the Female Orgasm: Bias in the Science of Evolution*. Harvard University Press.

Millikan, R. 1989. In defense of proper functions. *Philosophy of Science* 56: 288–302.

Murphy, D. 2006. *Psychiatry in the Scientific Image*. MIT Press.

Murphy, D., and R. L. Woolfolk. 2000. The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, and Psychology* 7(4): 241–252.

Neander, K. 1991. Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science* 58(2): 168–184.

Payn, N. M., and J. C. Wakefield. 1982. The effects of group treatment of primary orgasmic dysfunction on the marital relationship. *Journal of Sex and Marital Therapy* 8: 135–150.

Protopapas, A., and R. Parrila. 2018. Is dyslexia a brain disorder? *Brain Sciences* 8(61): 1–18.

Protopapas, A., and R. Parrila. 2019. Dyslexia: Still not a neurodevelopmental disorder. *Brain Sciences* 9(9): 1–5.

Reason. 2019. In *Merriam-Webster.com*. https://www.merriam-webster.com/dictionary/reason.

Ribatti, D. 2009. William Harvey and the discovery of the circulation of the blood. *Journal of Angiogenesis Research* 1(3): 1–2.

Schneps, M. H. 2015. Dyslexia can deliver benefits. *Scientific American Mind*, January 1. https://www.scientificamerican.com/article/dyslexia-can-deliver-benefits/.

Smith, W. L., E. Everman, and C. Richardson. 2018. Phylogeny and taxonomy of flatheads, scorpionfishes, sea robins, and stonefishes (Percomorpha: Scorpaeniformes) and the evolution of the lachrymal saber. *Copeia* 106(1): 94–119.

Spitzer, R. L. 1997. Brief comments from a psychiatric nosologist weary from his own attempts to define mental disorder: Why Ossorio's definition muddles and Wakefield's "harmful dysfunction" illuminates the issues. *Clinical Psychology: Science and Practice* 4(3): 259–261.

Spitzer, R. L. 1999. Harmful dysfunction and the *DSM* definition of mental disorder. *Journal of Abnormal Psychology* 108(3): 430–432.

Steinberg, P. L., and Buoy Medical Review Team. 2019. 28 reasons for a light or severe headache and nausea: A comprehensive guide on headaches, whether it's a headache after eating, a workout, or due to an illness. *Buoy Health Blog*, March 8. https://www.buoyhealth.com/current/28-types-of-headaches-and-causes.

Sullivan-Bisset, E. 2017. Malfunction defended. *Synthese* 194(7): 2501–2522.

Symonds, D. 1979. *The Evolution of Human Sexuality*. Oxford University Press.

Szanto, A., V. Narkar, Q. Shen, I. P. Uray, P. J. A. Davies, and L. Nagy. 2004. Retinoid X receptors: X-ploring their (patho)physiological functions. *Cell Death & Differentiation* 11, S126–S143.

Taylor, C. 1964. *The Explanation of Behavior*. Routledge and Kegan Paul.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20: 410–433.

Treiman, R., J. Gordon, R. Boada, R. L. Peterson, and B. F. Pennington. 2014. Statistical learning, letter reversals, and reading. *Scientific Studies of Reading* 18(6): 383–394.

Wakefield, J. C. 1992a. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47: 373–388.

Wakefield, J. C. 1992b. Disorder as harmful dysfunction: A conceptual critique of *DSM-III-R*'s definition of mental disorder. *Psychological Review* 99: 232–247.

Wakefield, J. C. 1993. Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria of mental disorder. *Journal of Abnormal Psychology* 102: 160–172.

Wakefield, J. C. 1995. Dysfunction as a value-free concept: A reply to Sadler and Agich. *Philosophy, Psychiatry, and Psychology* 2: 233–46.

Wakefield, J. C. 1997a. Diagnosing *DSM-IV*, part 1: *DSM-IV* and the concept of mental disorder. *Behaviour Research and Therapy* 35: 633–650.

Wakefield, J. C. 1997b. Diagnosing *DSM-IV*, part 2: Eysenck (1986) and the essentialist fallacy. *Behaviour Research and Therapy*: 35: 651–666.

Wakefield, J. C. 1997c. Normal inability versus pathological disability: Why Ossorio's (1985) definition of mental disorder is not sufficient. *Clinical Psychology: Science and Practice* 4: 249–258.

Wakefield, J. C. 1997d. When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder. *Development and Psychopathology* 9: 269–290.

Wakefield, J. C. 1998. The *DSM*'s theory-neutral nosology is scientifically progressive: Response to Follette and Houts. *Journal of Consulting and Clinical Psychology* 66: 846–852.

Wakefield, J. C. 1999a. Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology* 108: 374–399.

Wakefield, J. C. 1999b. Mental disorder as a black box essentialist concept. *Journal of Abnormal Psychology* 108: 465–472.

Wakefield, J. C. 1999c. The concept of mental disorder as a foundation for the *DSM*'s theory-neutral nosology: Response to Follette and Houts, part 2. *Behaviour Research and Therapy* 37: 1001–1027.

Wakefield, J. C. 2000a. Aristotle as sociobiologist: The "function of a human being" argument, black box essentialism, and the concept of mental disorder. *Philosophy, Psychiatry, and Psychology* 7: 17–44.

Wakefield, J. C. 2000b. Spandrels, vestigial organs, and such: Reply to Murphy and Woolfolk's "The harmful dysfunction analysis of mental disorder." *Philosophy, Psychiatry, and Psychology* 7: 253–269.

Wakefield, J. C. 2001. Evolutionary history versus current causal role in the definition of disorder: Reply to McNally. *Behaviour Research and Therapy* 39: 347–366.

Wakefield, J. C. 2004. The myth of open concepts: Meehl's analysis of construct meaning versus black box essentialism. *Applied & Preventive Psychology* 11: 77–82.

Wakefield, J. C. 2006. What makes a mental disorder mental? *Philosophy, Psychiatry, and Psychology* 13: 123–131.

Wakefield, J. C. 2007. The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry* 6: 149–156.

Wakefield, J. C. 2009. Mental disorder and moral responsibility: Disorders of personhood as harmful dysfunctions, with special reference to alcoholism. *Philosophy, Psychiatry, and Psychology* 16: 91–99.

Wakefield, J. C. 2011a. Darwin, functional explanation, and the philosophy of psychiatry. In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory,* P. R. Andriaens and A. De Block (eds.), 143–172. Oxford University Press.

Wakefield, J. C. 2011b. *DSM-5* proposed diagnostic criteria for sexual paraphilias: Tensions between diagnostic validity and forensic utility. *International Journal of Law and Psychiatry* 34: 195–209.

Wakefield, J. C. 2014. The biostatistical theory versus the harmful dysfunction analysis, part 1: Is part-dysfunction a sufficient condition for medical disorder? *Journal of Medicine and Philosophy* 39: 648–682.

Wakefield, J. C. 2015. Psychological justice: *DSM-5*, false positive diagnosis, and fair equality of opportunity. *Public Affairs Quarterly* 29(1): 32–75.

Wakefield, J. C. 2016a. The concepts of biological function and dysfunction: Toward a conceptual foundation for evolutionary psychopathology. In *Handbook of Evolutionary Psychology,* D. Buss (ed.), 2nd ed., vol. 2, 988–1006. Oxford University Press.

Wakefield, J. C. 2016b. Diagnostic issues and controversies in *DSM-5*: Return of the false positives problem. *Annual Review of Clinical Psychology* 12: 105–132.

Wakefield, J. C. 2019. The harmful dysfunction analysis of addiction: Normal brains and abnormal states of mind. In *The Routledge Handbook of Philosophy and Science of Addiction,* H. Pickard and S. H. Ahmed (eds.), 90–101. Routledge.

Wakefield, J. C., and M. B. First. 2003. Clarifying the distinction between disorder and nondisorder: Confronting the overdiagnosis ("false positives") problem in *DSM-V*. In *Advancing DSM: Dilemmas in Psychiatric Diagnosis*, K. A. Phillips, M. B. First, and H. A. Pincus (eds.), 23–56. American Psychiatric Press.

Wakefield, J. C., and M. B. First. 2012. Placing symptoms in context: The role of contextual criteria in reducing false positives in *DSM* diagnosis. *Comprehensive Psychiatry* 53: 130–139.

Wakefield, J. C., D. Wasserman, and J. A. Conrad. 2020. Neurodiversity, autism, and psychiatric disability: The harmful dysfunction perspective. In *Oxford Handbook of Philosophy and Disability,* A. Cureton and D. Wasserman (eds.), 501–521. Oxford University Press.

Wang, C. 2017. Let me introduce you to "Asian fit" glasses. *Refinery29*. https://www.refinery29.com/en-us/asian-glasses.

Why Asian fit? It's not just for Asians. 2019. *TC Charton*. https://www.tc-charton.com/asian-fit-eyewear.aspx.

Wright, L. 1973. Functions. *Philosophical Review* 82: 139–168.

Ziegler, J. C., C. Perry, A. Ma-Wyatt, D. Ladner, and G. Körne-Schulte. 2003. Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology* 86: 169–193.