12 Is the Harmful Dysfunction Analysis Descriptive or Stipulative, and Is the HDA or BST the Better Naturalist Account of Dysfunction? Reply to Maël Lemoine

Jerome Wakefield

I thank Maël Lemoine for his provocative and nuanced critique of my harmful dysfunction analysis (HDA) of the concept 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, 2000a, 2000b, 2001, 2006, 2007, 2009, 2011, 2014, 2016a, 2016b; Wakefield and First 2003 2012). There are more points of contention raised by his wide-ranging paper than I can address here, and some are best pursued in future personal interactions in which I look forward to extending the enjoyable interchange that started when we met at a conference in honor of Christopher Boorse some time ago. In this reply, I focus on the most important questions raised by his paper's main line of argument challenging the HDA. Some of these are questions I have not dealt with before, and I thank Lemoine for prodding me to address them.

A brief word about terminology and abbreviations: there is a general "selected effects" account of functions that is applied across domains (e.g., biology, artifacts) and is commonly abbreviated as "SE" functions. But, here I am concerned only with biological functions and with the theory of natural selection, and when the SE account is so restricted, I will label it the "evolutionary" or "NSE" (i.e., naturally selected effects) approach to functions. Cummins puts forward the very broad view that functions are simply the causal roles played by various mechanisms, commonly labeled "CR" functions (see my response to Murphy for a detailed analysis of CR functions). Boorse employs a form of CR functions that he terms the "general goal contribution" (GGC) account, which restricts relevant causal roles to the ones that contribute to goals, and he applies the GGC across domains. When applied to biology, Boorse claims that the goals of organisms are survival and reproduction, so the GGC view becomes the view that a biological function is the causal contribution made by a mechanism to survival and reproduction, which he labels the "S&R" view of biological functions.

First, it is important to clarify that a central point raised by Lemoine as a criticism is in fact one on which he and I entirely agree. Lemoine argues that when I call the HDA a descriptive conceptual analysis, I "conflate two good ideas into one" because the HDA's evolutionary analysis of "dysfunction" is in fact not a descriptive conceptual analysis but something different, which he calls a "stipulation" (I will come back to the stipulation issue later): "there is indeed something stipulative in Wakefield's position … but also something descriptive. Yet what is stipulative is not the general framework for the concept of mental disorder … : it is rather the evolutionary concept of dysfunction." Lemoine further claims that the descriptive versus stipulative distinction reveals a tension within the HDA: "Wakefield's arguments hide an incompatibility (at some point) between two purposes: on one hand, to give an account of what is usually meant by 'mental disorder'; on the other hand, to give a satisfactory scientific account of the concept of dysfunction … : while a descriptive account must not stipulate, a scientific account has to."

Now, if we momentarily put aside the question of whether the proposed nonconceptual-analytic part of the HDA should be interpreted specifically as a "stipulation," *Lemoine's account is precisely the position I have always put forward*! "Harmful dysfunction" is a conceptual analysis prior to the evolutionary interpretation of "dysfunction," and the evolutionary interpretation of "function" is an essentialist theoretical move that is not conceptual-analytic or sheerly descriptive but a theoretical identification. Thus, if I can be excused a lengthy self-quote to make this point clear, near the beginning of my main article defending the evolutionary part of my account (1999a), which Lemoine cites in his references, I said,

> One technically must distinguish the analysis of disorder as harmful dysfunction from the evolutionary theory of dysfunction, which together comprise the HD view. The HD analysis cannot directly define disorder in evolutionary terms because the analysis aims to capture a widely shared, intuitive medical and lay concept that existed long before evolutionary theory and is shared by many who are ignorant of or who reject evolutionary theory. Simply put, you do not have to understand or accept evolution to possess the concept of disorder. It is a momentous scientific discovery, not a matter of definition, that natural selection is the essential process that explains functions and dysfunctions. So, harmful dysfunction is the meaning of disorder, and evolution is the most incisive theory of the nature of functions and dysfunctions.
>
> The HD analysis may be thought of as arriving at the evolutionary account in three steps. First, … a disorder exists only when an internal mechanism is dysfunctional, specifically in the sense that it is incapable of performing one of its natural functions (at this stage of the analysis, natural function is used in an intuitive sense that has existed for millennia, not in a technical evolutionary sense). … Second, … natural function[s] … like the intentionally designed functions of artifacts, must somehow be part of the explanation of why the underlying mechanisms exist and are structured as they are. (By analogy with artifacts, such functions are often said to be what the mechanism is "designed" to do.). … Disorders, then, are failures of mechanisms to perform their natural functions. …

> Strictly speaking, these two steps complete the conceptual analysis of disorder. However, the analysis inevitably leads to the question, What kind of underlying process could possibly be responsible for such seeming design in natural systems without any designer? … Evolutionary theory provides the only plausible scientific account that presently exists of how the natural functions of a mechanism can explain the existence and structure of the mechanism. … This third, theoretical argument leads to the conclusion that disorders are failures of mechanisms to perform functions for which they were naturally selected. (1999a, 374–375)

So, I have been quite explicit about this, and Lemoine is simply agreeing with me about the dual nature of my argument for the HDA. As Lemoine observes, I link the conceptual analysis and evolutionary theory via a "black-box essentialist" analysis of "function" (Wakefield 1999b, 2000a); a natural function is, conceptually, whatever is due to the same essential process that brings about a base set of obvious examples of apparent biologically designed features, such as eyes seeing, hands grasping, thirst causing us to drink needed water, fear causing us to flee danger, and so on. It is an empirical discovery that that essential process is natural selection, implying that a dysfunction (in the sense relevant to medical disorder) is failure of a mechanism to be capable of having its naturally selected effect. Note that there is no comparable unified and direct essentialist account of dysfunctions, only the indirect one that they are failures of natural functions, because such failures occur for myriad diverse reasons.

In fairness to Lemoine, in some of my writings, when it seemed not to matter, I have been sloppy about distinguishing the conceptual-analytic and scientific-theoretical aspects, misleadingly compressing my description to characterize the entire HDA as a conceptual analysis. However, in my more careful theoretical writings, I have clearly drawn the distinction that Lemoine accuses me of ignoring and have even critiqued (e.g., Wakefield 2000a) both Neander (1991a) and Millikan (1989) for making the error, in different ways, of interpreting the link between "function" and natural selection as a conceptual one, and this should have made my position clear.

The Conceptual Analysis (Strictly Speaking) of "Medical Disorder"

Given the agreed division of the HDA into a conceptual-analytic and scientific-theoretical component, Lemoine's discussion raises two important questions: (1) Exactly how far can conceptual analysis take us before we must turn to scientific theory? (2) What degrees of freedom to "stipulate" do we have in moving via the black-box essentialist structure from the conceptual analysis to a theory of functions and dysfunctions? Lemoine answers both questions in ways I will dispute.

Lemoine agrees with what he interprets as the conceptual-analytic part of the HDA. He says that other than the evolutionary interpretation of dysfunction, the HDA "is probably the best account of the concept of mental disorder" and "is indeed a highly faithful account of what both laymen and psychiatrists mean by 'disorder.'" He also

seems to accept or at least acknowledges the advantages of my black-box essentialist account of function and dysfunction as allowing a theory-free conceptual analysis: "Thanks to this black-box view, the concept of dysfunction can avoid theoretical stipulation."

For the HDA's conceptual-analytic part prior to the evolutionary interpretation of function and dysfunction, Lemoine coins the term "harmful abnormality." The use of the vague term "abnormality" is specifically aimed at stepping back from a more specific notion of "dysfunction." Lemoine does not offer any actual counterexamples to support his rejection of "dysfunction" in favor of the broader notion of "abnormality" but rather relies on an assertion by Derek Bolton: "What we know as mental disorder—or at least, as mental health problems—can involve factors other than *dysfunction*. Among the most important and readily understood key ideas in an evolutionary theoretic framework that point in this direction are (1) design/environment mismatches, (2) highly evolved design features of human beings, (3) defensive strategies, and (4) strategies that involve disruption of function" (Bolton 2000, 146). Citing Bolton's assertion, Lemoine concludes that the conditions that fall under the intuitive concept of disorder are of many kinds that go well beyond dysfunction.

Lemoine's conclusion is mistaken and based on lack of attention to actual intuitive examples (see, e.g., Wakefield 1999a). For example, defensive strategies (e.g., coughing in response to dust in the air), functions that are designed to disrupt subsidiary functions (e.g., impairment of a man's ability to urinate when sexually aroused), and design/environment mismatches (e.g., problems with desiring high-fat and high-sugar foods in modern environments in which they are all too readily available) are not generally judged to be disorders. Additionally, if organism-environment mismatches are classified as disorders, that immediately makes every problematic social deviation into a mental disorder, which is one of the main outcomes that the analysis of mental disorder is meant to prevent.

In fact, Bolton hedges his claim by specifying that his list consists of conditions that "we know as mental disorder—*or at least, as mental health problems*" (emphasis added). Elsewhere in the same article, Bolton cites my arguments and notes that my analysis leads to the conclusion that not all such potentially treatable problematic mental health conditions are literally disorders: "there may be a broader class of behaviors relevant to 'mental health' than the class of disorders defined by Wakefield. … Wakefield acknowledges that *disorder* and *treatable conditions* do not coincide" (145). Note that the scope of Bolton's term "mental health problems" goes well beyond failures of health in the medical sense and encompasses almost any negative psychological state. For example, a very extensive list of Z coded mental conditions that are not disorders but are frequent targets of clinical intervention is included in the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5*) (American Psychiatric Association 2013) and the *International Classification of Diseases* (*ICD-11*) (World Health Organization 2018).

Writers sympathetic with the HDA (e.g., Cosmides and Tooby 1999), including me (Wakefield 2015), have argued that there are many problems that the medical professions ought to be mandated to treat but that are not mental disorders, ranging from substance abuse without addiction to marital conflict. Consequently, Lemoine's use of the broad category of abnormality instead of dysfunction in his rendition of the HDA conceptual component is unsupported.

Having rejected dysfunction as a conceptual requirement for disorder, Lemoine elaborates his broader analysis:

> So a description of the general framework of the concept of mental disorder ought to be some sort of deflationary or downgraded version of the harmful-dysfunction analysis (I propose "HAA" for harmful abnormality analysis). By "abnormality" here, I mean a much broader concept than that of dysfunction and one that is not restricted to the statistical concept of abnormality. In a nutshell, "abnormality" addresses the notion of the objective basis of the concept of mental disorder, whether it is a dysfunction, a mismatch, a strategy, and so on. Abnormalities are observed facts; they are not supposed to be spotted after value judgments and *they* are expected to limit arbitrary disease entities and false-positive cases. The preceding sums up an analytic or descriptive approach to the HAA. The HAA is the only uncontroversially descriptive general framework for the concept of mental disorder: every further specification is stipulation and constitutes a theoretical move.

Lemoine does not explain how such an extraordinarily broad notion that allows any problematic objective internal state to be a dysfunction can possibly serve to "limit arbitrary disease entities and false-positive cases." In any event, as we have seen, there are ample reasons why this analysis cannot be correct. Beyond the earlier examples of nondisordered defenses, mismatches, and so on, "abnormality" in the sense that Lemoine defines it encompasses a vast terrain of problematic nondisorders (e.g., ignorance, lack of talent, negative personality traits, social deviance), defeating the point of an analysis of disorder. (For further explanation of why such a view fails, see my reply to De Vreese in this volume.)

The history of medicine from Hippocrates to our own time indicates that the concept of a medical disorder involves more than problematic states caused by internal conditions, for there are many normal problematic states caused by internal conditions. It involves a presupposition that, as Robert Spitzer used to put it, *something has gone wrong,* which, I have argued (and Spitzer eventually agreed [Spitzer 1997, 1999]), involves the presupposition that there is a failure of some internal mechanism to perform as it was biologically designed to perform. Biological design, which is apparent to laypersons and scientists alike, has been the central puzzle of biology from Aristotle to Darwin, and it is anchoring in the objective feature of biological design that allows disorder to transcend values and gives it a distinctive social status. (For further comments on the historical centrality of biological design to biology, see my reply to Murphy in this volume.) Thus, conceptual analysis can take us further than Lemoine allows. The

purely conceptual-analytic meaning of "disorder" is "harmful dysfunction" in which "dysfunction" is not evolutionarily interpreted but understood in intuitive biological-design natural-function terms.

Bolton (2000), in the same article cited by Lemoine, seems to understand better than Lemoine the logic and appeal of natural function as part of the foundation for the prescientific concept of disorder:

> Wakefield (1999a) clarifies his position as being that "harmful dysfunction is the meaning of disorder, and evolution is the best theory of the nature of functions and dysfunctions." …
>
> This is an important and helpful line of thought. The suggestion is that we have first an analysis of a folk concept of disorder, or its reasonably close cognates, appealing essentially to what is natural—or of our nature. This we may plausibly suppose not only captures the principles of common usage of terms related to disorder, but is also at work among physicians, including mental health professionals in the clinic. (145)

So, the answer to the first question raised by Lemoine's critique—namely, what the conceptual analysis of "disorder" can yield—is that "disorder" means "harmful dysfunction" where "dysfunction" means "failure of a natural function" (or "failure of a biologically designed function") in a preevolutionary sense, and "natural function" is understood in terms of a black-box essentialist descriptive naturalist definition. Note that this answer to the first question imposes limits on the answer to the second question regarding the potential scope for stipulation in specifying a theory of dysfunction, a point to which I will later return.

Modest Black-Box Essentialism

Before addressing stipulation, I need to clarify my understanding of essentialism—a term to which some critics seem allergic—because the question of stipulation will be considered within the context of the black-box essentialist analysis of "function." I construe an essentialist view of natural kind concepts in a minimalist way that avoids the metaphysical doctrinal loading of Kripke's (1980) account and the strawman formulations of critics (e.g., Kendler, Zachar, and Craver 2011). On most issues, I adopt Putnam's nonmetaphysical and more scientific and pragmatic account that, simply stated, is that "to be water, or gold, or some other natural kind, is to have *the same nature* as 'this,' where the 'this' can be any one of the [majority of the] paradigms we point to, *and the 'sameness of nature' is a scientific or protoscientific concept, not a metaphysical one*" (Putnam 2015c, 359). (Putnam more frequently calls the paradigm cases "stereotypes" and I call them the "base set.") In particular, the "nature" or "essence" of the category is a nonobservable feature that is explanatorily and theoretically potent, in that it plays a major role in formulating theories that explain the important features of the base set, including the salient observable features that made us pick out the base set to define a broader category in the first place. I also accept a flexible version

of Kripke's notion of "baptism," in which one identifies an initial set of instances as the basis for reference fixation of the overall category. This construct seems necessary to make good on Putnam's reliance on a reference-fixing sample of "stereotypical" category members.

The essence is not explicitly identified in the category's definition but is rather referenced via the indirect description, "whatever is the same in nature as the base set." This allows the category to be defined without any explicit reference to, and often without any knowledge of, the specific hypothesized essential property that determines its members and explains their salient properties. The basic point is that new category members are added based not on their superficial similarity to the members of the base set but on the basis of the judgment that they share the relevant underlying nature with the base set. The base set may be defined descriptively in terms of observable features, but that description does not define the entire category: "The stereotype, that gold is yellow, precious, etc., is *not* analytic; it may well turn out to be wrong; but nevertheless the shared stereotype plays a role in stabilizing the use of 'gold'" (Putnam 2015a, 77). The kind can always transcend the stereotypical observable properties (e.g., the base set of tigers are large striped felines, but there are still dwarf albino tigers). Despite this definitional structure, essentialistic categories, like most categories, generally have fuzzy boundaries in virtue of the fuzziness of the various component concepts.

Admittedly, the term "essence" has a disturbing resonance with bygone metaphysical doctrines. However, it is generally used today as a philosophical term of art shorn of such doctrines. As noted above, it refers to hidden explanatorily potent structures, as in gold being the element with atomic number 79 or water being the chemical H_2O. Moreover, the "hidden structure" characterization cannot be taken too literally. We can of course use terms in all sorts of ways for varying purposes, including positivist observable-property meanings. However, especially in science, the conceptual and theoretical undertow pulls us toward structures beyond observable properties because they generally support a more perspicuous theory and deeper understanding. "Hidden structure" is best understood as simply a term of art covering almost any not-directly-observable property that determines category membership. So, on this modest interpretation, if the intentions of designers determine artifact category membership or the history of an interbreeding population in addition to its genetics partly determines a species, these can still be essentialist concepts.

Essentialism in this modest form has a number of benefits, from correcting positivist accounts of concept meanings in terms of superficially observable or operationalizable properties, to offering a path to reject positivist-meaning holism and thus an escape from Kuhnian incommensurability. The descriptive elements in the identification of the base set preserve a link to observables but in a nonpositivist way that, reflecting scientific reality, allows category reference to go beyond any reductionistic tie to the observable. Consequently, the common scientific occurrence of being surprised by the

novel way things with certain observable properties are theoretically categorized is explained by modest essentialism.

There is a problem for essentialism known as the "*qua*" problem (Devitt and Streleny 1987) that might be mistaken for indeterminacy or an opportunity for stipulation but is quite different and reveals the importance of an additional element of natural kind definitions. The *qua* problem arises from the fact that the very same baptized base set has many different levels and types of hidden explanatory structures that, if identified as the sought-after "nature" of the base set, determine various distinct categories: "Any sample of a natural kind is likely to be a sample of many natural kinds; for example, the sample is not only an echidna, but also a monotreme, a mammal, a vertebrate, and so on … a term refers to all objects having the same underlying nature as the objects in the sample. But which underlying nature? The samples share many" (Devitt and Streleny 1987, 73).

Putnam proposes that the relevant essence of a kind in a given context is determined by an additional often-implicit element of the definition, a "semantic marker"—or what I call an "ontological marker"—that specifies what kind of thing is being defined. I agree with Putnam that many terms are used with a multiplicity of varying ontological markers indicating varying related kinds. Putnam embraces the resulting diversity and observes that for the same term and base set, the category reference can vary but remain determinate in each context based on the interests of those doing the defining ("in one context, 'water' may mean *chemically pure water,* while in another it may mean the stuff in Lake Michigan"), and he labels such judgments "interest-relative and context-sensitive" (Putnam 2015a, 80). These divergences occur in science as well as ordinary language: "'Is it part of the essence of dogs that they are descended from wolves?' The answer seems to be 'yes' from an evolutionary biologist's point of view and 'no' from a molecular biologist's point of view" (Putnam 2015b, 333).

It is important to keep in mind here that multiplicity is not the same as indeterminacy. Despite the plurality of ontological markers, there is not a stipulative free-for-all because in each context there is an anchoring in specific aspects of reality: "I believe that given the interests that structure the various natural sciences, some classifications are objectively more natural than others. This does not mean that all the natural sciences must use the same classification: a molecular biologist may legitimately classify organisms differently than an evolutionary biologist" (Putnam 2015c, 359).

Several critics raise the question of whether the concepts with which I deal are ordinary folk concepts or experts' scientific concepts. In the case of "medical disorder," I see no decisive *conceptual* separation between ordinary and professional technical concepts, although there are of course vast differences in ordinary and expert beliefs. This view is consistent with the essentialist view that science is often filling in the black boxes in vernacular essentialist concepts, but the concept itself generally stays the same and only the identification of the indirectly-referred-to essence is at stake: "Our language is a cooperative venture; and it would be a foolish layman who would

be unwilling to ever accept correction from an expert on what was or was not water, or gold, or a mosquito, or whatever. … Ordinary language and scientific language are different but *interdependent*" (Putnam 2015c, 361–362; see also Putnam 2015b, 333).

With this elaboration of a modest essentialism in hand, I turn to the question of whether or in what senses identifying the essence of natural functions allows for stipulated choice between the HDA and the biostatistical theory (BST).

Do the HDA and BST Offer Competing Theories of "Function"?

Because the HDA's evolutionary account of function and dysfunction is not a conceptual analysis, Lemoine leaps to the conclusion that it is no longer linked tightly to an analysis of meaning and that the evolutionary component of the HDA is in fact a sheer stipulation and should be explicitly pursued as such: "it is suggested that the 'conceptual analysis' approach be replaced by a full-fledged naturalist approach of mental disorder that is openly stipulative." To fill the proposed space open to stipulation, Lemoine thinks he can simply choose Boorse's (1987) "biostatistical theory" of disorder over the HDA as a better stipulation.

From the fact that the evolutionary account of function is not a conceptual analysis, it does not follow that it is a stipulation. According to the HDA's black-box essentialist analysis of "natural function," the evolutionary account of "natural function" is an explanatory scientific theory of the nature or essence of natural functions. Such theories are not stipulations in any usual sense of the word. They are scientific discoveries that are embraced on the basis of evidential support, explanatory power, and the scientific goals and methodological canons of a given scientific discipline. For example, the identity of the essence of water as the chemical structure H_2O is not a conceptual analysis, but neither is it a stipulation. Rather, it is a scientific discovery about the nature of water. It is embraced on the basis of a judgment that it is evidentially the best supported theory. Similarly, it is a matter of the scientific evidence and not stipulation whether evolution versus, say, creationism better specifies the essential nature of the process that explains biological design. In sum, the HDA's evolutionary component is best construed not as a stipulation or choice of how to think about functions but as a scientific theory of the nature of natural functions yielding biological design.

However, even if Lemoine accepts the black-box essentialist framework for formulating a theory of natural functions (in a later section, I will consider these issues apart from any essentialist assumptions), he could claim that the HDA and the BST are competing essentialist theories of this domain. If the BST is a viable alternative theory of natural functions, perhaps it could legitimately be "stipulated" or selected as the best available theory consistent with the constraints of scientific methodology and the goals of theory formation. Scientific domains are, after all, filled with rival theoretical formulations that compete for evidential vindication.

From a black-box essentialist perspective, the statistical-contribution-to-S&R and evolutionary theories in principle can be construed as essentialist scientific theories that attempt to explain adaptation, natural functions, and biological design, and they can be compared in terms of explanatory power and evidential support. Lemoine does not attempt such a comparison. However, on Lemoine's behalf, we might consider the following scenario. If Boorse's S&R theory of biological function and the HDA's evolutionary theory are construed as rival scientific theories of the same target domain of natural functions, but no clear deciding evidence exists, then one might imagine a "stipulation" of one or the other theory for certain purposes. Analogously, for example, at a point when there was genuine scientific uncertainty about whether the phlogiston theory or the oxidation theory of fire would prove to be true, in a discussion of fire, various theorists might simply have stipulated one theory or the other for the sake of that discussion.

However, to have such a situation in which stipulation might enter into theory selection, both theories would in fact have to be attempting to explain roughly the same domain of phenomena, and both would have to do a reasonably good job. That domain, which we saw in the discussion above, is the nature or essence of whatever explains the presence of biological design. The BST simply identities S&R-productive organismic features. Contrary to the "stipulation" scenario, it is a confusion to think that the HDA's evolutionary account and the BST's S&R account are plausibly construed as rival accounts of "natural function." They are in fact attempts to explain different things. In keeping with the analysis presented earlier, the target of the evolutionary account is to explain biological design. Whereas the BST account is explaining how the organism's nature causes it to have capacities for survival and reproduction, which, as it happens, are the two most salient domains of biological design. These are different explanatory problems of how biological design works.

What, then, even in very rough schematic form, would a black-box essentialist definition of "natural function" look like, and how is it different from a BST-type explanation? Let me approach this question via a well-known example of Aristotle's, that acorns have the remarkable ability under suitable circumstances to grow into oak trees, which in turn produce such acorns. That, one intuitively judges, can't be a mechanistic accident; this is such an unlikely and remarkable process that acorns in some sense must be "designed" to grow into oak trees. Aristotle understood that in explaining such a puzzling phenomenon, there are two different causal explanations required: efficient and final causes. Aristotle had no idea of the details of either explanation but understood that both types of explanation must be involved. The efficient cause of the acorn's turning into an oak tree is a standard causal explanation of how it works, addressing the puzzle of how an acorn can possibly produce an oak tree. The explanation will be couched in terms of the nature of the acorn's internal structures and parts, its interaction with the soil's nutrients, its positioning relative to sunlight, and so on. This explanation will be complex, and it will involve many initial conditions given

that, according to some estimates, perhaps only 1 in 10,000 acorns actually grows into an oak tree even under standard conditions. The efficient causal explanation of how the parts of an acorn contribute to its ability to become an oak tree is the type of explanation provided by Cummins's causal role (CR) functions. And, because the acorn turning into an oak tree that in turn produces acorns is the act of reproduction of the oak tree that produced the acorn, the efficient causal explanation provides the Boorsean S&R biological functions of the parts of the acorn—that is, their contributions to the oak tree's S&R. A full efficient causal account will allow us to understand how this biologically designed acorn-to-oak-tree process works.

What is lacking here? How the acorn's nature explains its capacity to grow into an oak tree is one major scientific mystery, but it is not the only scientific mystery. Aristotle saw that a second scientific mystery—and perhaps the more profound one—is to explain what in nature shapes organisms to have parts with the specific causal powers that enable them to produce such design-like effects and contribute in such unlikely ways to S&R. This is a second-level explanatory mystery, and the inferred cause by which the end shapes the means is what Aristotle referred to as the "final cause." The problem Darwin addresses is analogous to this Aristotelian puzzle of final causation, which is the second-order causal puzzle of what causes organisms to have efficient causal properties that are instances of and produce biological design—for example, why things like acorns have an unlikely, coordinated, and remarkably complex system of causal powers such that they yield things like oak trees, where the entire process appears biologically designed. The final cause is not about how an acorn's structure gives it the power to become an oak tree but about how an acorn comes to have the kind of structure that enables it to become an oak tree when that very structure must in some way have been shaped by the very fact that they have that oak tree outcome. The challenge of natural functions, then, is the explanation of biological design, and natural functions are in the first instance the category of effects that are design-like. The mystery that Aristotle labeled "final causation" is what explains why the efficient causation by the acorn's parts has effects that are design-like. Functions as causal contributions to S&R offer an efficient-causal analysis of the most salient acknowledged domains of biological design. Functions as *naturally selected* contributions to S&R offer a final-causation analysis of biological design itself and explain why there are so many substantial S&R functions.

So, in a black-box essentialist vein, one might say: for an effect of an organism's feature to be a natural function of that feature is for it to be due to the right effect-sensitive causal process (i.e., a process in which the effect somehow causally shapes the feature's mechanisms that lead to it and where, in this case, the effect is part of the organism's biological design; I here alter Cummins and Roth's [2009] terminology of "function-sensitivity" that explains standard examples of biological design such as eyes seeing, hands grasping, and acorns growing into oak trees). The hypothesized process that explains why organisms have so many S&R functions that are structured in an

apparently biologically designed way to produce biologically designed outcomes could have been some mystical inherent final-causation principle in the universe, or God's handiwork, or many other processes that have been proposed through the ages, but it turned out (based on the best scientific theory we have at present) to be natural selection. But, of course, that discovery of the essence of natural functions and biological design does not enter into the definition of the category in a black-box essentialist definition.

Consequently, whereas S&R functions are standard causal properties, natural functions presuppose second-order causal properties—that is, natural functions presuppose that there is a distinctive effect-sensitive causal explanation of the organism's features having certain distinctive kinds of the standard effects they have. This distinctive type of second-order causal explanation is precisely what evolutionary theory and natural selection provide. So, a first approximation to a black-box essentialist definition of "natural function" might look something like this: "*A natural function N of an organismic feature X is an S&R-function of X, the presence of which is at least partly explained by the same (presumptively) effect-sensitive natural process that explains the base set of instances of biological design, such as the presence of the eyes' S&R function of enabling sight, the hands' S&R function of enabling grasping, fear's S&R function of taking us away from danger, thirst's S&R function of causing us to seek out and drink water needed to survive, and an acorn's capacity to grow into an oak tree.*"

If this is correct, then there is no opportunity for the sort of stipulation Lemoine suggests, at least not within an essentialist framework for theorizing about the relevant function concept. This is because the BST and HDA are not rivals in explanatory competition. Rather, evolutionary theory and the BST's S&R functions address two different questions. The BST's S&R notion of function, basically a restriction of causal-role functions to those that contribute to the S&R of the organism, attempts to describe how the processes that constitute biological design work but has no capacity to explain what the concept of natural function addresses, namely, why such an ample number of S&R functions constituting biological design exist in the first place. So, one cannot choose to stipulate the BST as the account of "natural function" because it simply does not address that issue.

I provisionally conclude that S&R functions are not rivals to the etiological account of natural functions but rather address a different first-order domain of causal relations. If so, there is no room for stipulation of one theory over another because they are not competitors. Only evolutionary theory attempts to identify the essence referenced in the definition of natural function.

Boorse on the Explanatory Power of the BST

Naturally selected effect (NSE) function theorists hold that "function statements are intrinsically explanatory: to ascribe a function to a device is to offer an explanation

of its presence" (Price 1995, 153). They thus object to the S&R account of "function" because, they claim, it offers no such explanations. However, Boorse (2002) disputes the claim that S&R functions "cannot accommodate functional explanation" (63) and have "insufficient explanatory power" (78) to match that of the NSE account aimed at explaining a feature's presence. Thus, the conclusion arrived at earlier that the NSE and S&R accounts are not rival explanatory accounts of "natural function" must remain provisional until Boorse's arguments that the two approaches provide similar explanatory power are considered.

Boorse first says, "In the first place, however, there was never any basis for assuming function statements to be inherently explanatory of anything, any more than statements about organisms, cells, … or most other objects of biology" (2002, 78). This is clearly not true specifically for natural functions, for which a 2,300-year tradition concerning final causation provides such a basis. We saw that in the case of natural functions, there is explanatory content both in the presumed effect sensitivity that explains a feature's nature or presence or maintenance and in the reference to an inferred natural process that explains why so many of such features yield apparent biological design. Boorse here simply begs the question.

Boorse proceeds: "In the second place, even if function statements have to be inherently explanatory, a satisfactory kind of non-etiological functional explanation is available: Cummins's functional analysis (1975), undeniably prominent in biological fields like physiology" (2002, 78). The attribution of a CR function does imply a causal role in producing the organisms' capacities and thus is explanatory in that way. However, as we saw, such CR functions are presupposed and built upon by the concept of natural function, which involves second-order causal attributions of the process that causes CR functions to come about in a way that yields biological design and is presumptively effect-sensitive. Citing the fact that CR functions have some causal explanatory power is thus a non sequitur with regard to the question of the specific forms of causal power attributed to natural functions.

Boorse further argues, "In the third place, that there are unselected biological functions is part of the current 'consensus'. … To attribute such functions is not to offer any etiological explanation. … Once one recognizes unevolved functions, their prevalence or rarity is, of course, an empirical question" (2002, 78). That is, the acknowledgment of CR functions (which I do acknowledge; see my response to Murphy in this volume) implies that not all things labeled "functions" are etiologically defined, and thus the general claim that functions must be etiological is defeated. However, this argument depends on interpreting the term "function" as univocal in meaning and as referring to S&R functions, where some functions as conceptual accidents just happen to have natural-function properties as well. This interpretation ignores the obvious possibility that S&R and NSE functions are two distinguishable senses of "function" with somewhat different meanings. Taking the latter approach, the "current consensus" is best

understood as a consensus that biology uses "function" in two distinct senses (or perhaps in a primary NSE sense and a secondary derived synecdochical S&R or CR sense; see my reply to Murphy in this volume). If so, throwing all the uses together into one bin and pointing to the CR sense as evidence that functions need not be explanatory is a confused non sequitur that makes no more sense than, say, rejecting the assertion that "water is a liquid" on the grounds that the substance concept, "water," covers ice and steam as well as liquid water. Clearly, there is an alternative sense of "water" that refers strictly to the liquid (thus the waiter really has made a mistake if in response to a request for a glass of water, he delivers a glass of ice, although in a chemistry class, maybe that would be fine). In citing features of CR functions to dispute claims about natural functions, Boorse is similarly confusing matters by running together broader and narrower meanings of "function." This is further problematic in that it potentially obscures the precise senses of "function" and "dysfunction" that form the basis for medical judgments of health and disorder.

As a further point, Boorse says of S&R functions, "What the analysis does not do is to write even the existence of such an [evolutionary] explanation, let alone its details, into the meaning of the function statement itself. But there is no reason why it should. … There is no reason why all the premises of a full evolutionary explanation, including background theory and initial conditions, must be part of a function statement's meaning" (2002, 80). I agree with Boorse (contra some etiological theorists) that it would be inappropriate for "a full evolutionary explanation" to be written into the concept of natural function. The concept "natural function" existed long before those evolutionary details were known, and surely its meaning makes no reference to those details. However, Boorse is incorrect in claiming that the analysis of "natural function" should not include reference to the existence of some effect-sensitive type of explanation. The concept of natural function rests on an inference to the existence of some such process that explains the bewildering existence of so many apparently adaptive S&R features that contribute to the quintessential biologically designed outcomes of survival and reproduction. The existence of effect-sensitive feature shaping is built into the meaning of the concept via the specification of the kind of shared "nature" that defines the category using base-set instances that exemplify biological design.

Boorse mounts another argument in defense of the explanatory power of S&R functions when he considers common function attributions, such as "the function of the giraffe's long neck is to reach up into the trees for food." Such a statement seems to be a way of explaining why the giraffe has a long neck, yet the S&R function attribution seems to offer no such explanations. Boorse (2002) argues that the explanation is indirect, arrived at by uniting the function attribution with what we know about the link between S&R functions and natural selection: "What I believe, with nearly all other current writers on biology, is the following: a disposition D of a trait type T causally to contribute to the goal of individual fitness can, via evolutionary theory, explain the

prevalence of present tokens of T by D's manifestation in past tokens. Since such contribution is a GGC function, that immediately solves … the 'Explanation Problem' of 'understanding how ascribing a function to a biological trait can help explain the trait's existence'" (79); "However one explains the origin of traits like … long necks via their fitness benefits, any such evolutionary explanation is in terms of these organs' GGC functions—namely, their causal contributions to goals of the organism, survival and reproduction. So the GGC account has no defect of missing explanatory power" (80).

Boorse's argument here is, again, a non sequitur. He is of course correct that, generally speaking, certain S&R functions influenced natural selection and thus over evolutionary history explain biological design via natural selection. So, he is correct that if you conjoin the S&R function of a feature with evolutionary theory, then maybe you will get an explanation of the presence of the feature (but, as noted earlier, not always, because lots of current S&R functions are not the S&R effects that actually shaped the selection of a feature). However, first, S&R functions in themselves have no *conceptual* implications that involve any effect-sensitive process let alone natural selection as a directly or indirectly referenced process or outcome. This is why Aristotle felt the need to add the final cause to the efficient cause. If the goal is a conceptual analysis of "function," Boorse has now left that domain and is linking "function" to etiology via a scientific discovery, thus failing to place the claim of an effect-sensitive causal process within the meaning of "natural function." Second, this supposed solution depends on the details of Darwin's discoveries and thus cannot possibly explain how it was that for the 2,100 or so years between Aristotle and Darwin, there was a teleological notion of "natural function" and a continuing deep mystery of biological design, the unknown solution of which was implicitly referenced in the meaning of "natural function."

Boorse notes that in some contexts such as evolutionary biology, "one can *agree* to mean by 'the function of X' *the evolutionary function of X*. If so, one's function statements will have essential etiological explanatory force—they will be 'equivalent to' or 'tantamount to' an evolutionary explanation" (2002, 80). This seeming concession evades the issue. Of course, one can always stipulate a meaning of a term. However, this dispute is about the conceptual analysis of an existing meaning, not the possibility of stipulating a deviant meaning. Moreover, "function" had explanatory force before we understood evolutionary theory and so we would like to understand that meaning, which was not determined by an evolutionary context.

I conclude that Boorse's attempted rebuttal fails to counter the explanatory-power objection, and so the earlier conclusion can stand. The HDA and the BST are not explanatory rivals, and so there is no option to choose or stipulate between them. This conclusion does, however, pose an interesting question. If the BST and HDA are not theoretical rivals in the way most observers have assumed, then what precisely is their relationship? I will return to this question in my conclusions.

Essence Indeterminacy and the Limits of Stipulation

The idea that one might stipulate an essence is not an idiosyncratic notion of Lemoine's but the topic of an active philosophical literature. Before proceeding to the remainder of Lemoine's argument, I briefly consider the implications of that literature for Lemoine's claim.

When I argued for the evolutionary theory of function and dysfunction, I gave no thought to possible stipulated choices because I thought of the essence of "natural function" as being firmly fixed by Darwin's theory, which provides the only respectable scientific explanation of biological design. However, the philosophical literature raises the possibility that in linking a vernacular essentialist concept to a scientific theory of the relevant essence, the process can be more complicated than simply discovering *the* essence. As Wilson (1982) puts it, "The 'natural kind' doctrine makes the uniqueness of this [essential] property seem more likely than is reasonably plausible" (579).

Keith Donnellan (1983/2014) offers perhaps the most influential example of a possible indeterminacy and stipulation in essence identification. He argues that, when there are multiple ways of dividing things up to form essences, different background senses of what is important can yield different judgments about how best to translate vernacular concepts into the terms of a new theory. Donnellan constructs his argument in the context of an imagined Putnam-style "Twin Earth" situation, but to simplify matters—and for those unfamiliar with the Twin Earth scenario—I paraphrase Donnellan's thought experiment without his parallel-worlds apparatus.

Each atom of a given element has the same characteristic number of positively charged protons in its nucleus, which is the element's "atomic number" and largely determines its chemical properties. For this reason, the periodic table of elements is organized by periodicities in atomic number that determine similar patterns of chemical reactions. In addition to protons, the nucleus can also contain varying numbers of neutrally charged particles, "neutrons," where each number of neutrons determines an "isotope" of the element. Neutrons are about as massive as protons, so the "atomic weight" of the same element's isotopes varies (electrons are of negligible weight in this context). Although chemical reactions are generally similar across the isotopes of an element, there can be significant nonchemical differences. For example, some isotopes of an element can be radioactive and others not, and some isotopes can be unstable and break down into another element, whereas others are stable.

Given the possible differences among isotopes of an element, Donnellan argues that, despite the overwhelming importance of elements' atomic numbers for chemical reactions, in principle, for those with interests different from ours, "it might be a close question as to whether isotope number or atomic number has more importance" (197). Thus, he claims, it was a *choice* whether the periodic table was organized so that atomic numbers or atomic weights are the essences of elements.

Interestingly, this thought experiment comes close to describing an actual historical occurrence. When the elements' atomic weights were calculated for Mendeleev's original periodic table of elements, the average of available samples was used and so the atomic weights represented an amalgam of the weights of isotopes readily available on earth. This led to some anomalies in the grouping of elements in terms of chemical properties. In a momentous scientific advance, Henry Moseley figured out in 1913 that the anomalies were eliminated if atomic number rather than atomic weight was used as the organizing principle, yielding the modern periodic table.

Of more relevance to the translation of vernacular terms, Donnellan extends his argument to the chemical substance, water. Scientists have given the names "protium," "deuterium," and "tritium" to the three isotopes of hydrogen in which the nucleus, along with hydrogen's one proton, have zero, one, and two neutrons, respectively. Since all three can combine chemically with oxygen, there are three types of H_2O or "water": protium, deuterium, and tritium water (deuterium water is known as "heavy water"). Donnellan argues his analysis "can, obviously, be extended to the vernacular term 'water.' … In my story, because isotopes are taken more seriously for one or another practical or historical reason, we can suppose that [we] will identify water with protium oxide and exclude what we call 'heavy water'—deuterium or tritium oxide" (Donnellan 1983/2014, 198). That this might have been an option makes some sense because earth's water is almost entirely protium water (99.98%), with a little deuterium "heavy water" thrown in (tritium is unstable and rare). Moreover, although protium and deuterium water have the same basic pattern of chemical reactions, there are some differences in the rate of some chemical reactions that can make a difference to the health of an organism. A glass of heavy water is harmless, but about 50% replacement of protium water with heavy water can be lethal. Also, heavy water plays a unique role in certain types of nuclear reactors. These practical nonchemical differences led Donnellan to argue that despite the importance of atomic numbers in explaining chemical reactions, in principle, scientists might judge that isotope number has more importance in identifying the essence of water. Thus, he claims, it was to some degree a *choice* to identify water (as understood in the vernacular) with the substance H_2O (including all hydrogen-isotope variations) rather than identifying water exclusively with protium water and leaving deuterium and tritium H_2O outside of the "water" category altogether.

Donnellan concludes his analysis with the following thoughts:

> What do I conclude from my story? I do not draw the conclusion that Putnam has failed to describe how natural kind terms in the vernacular function. The story does not show that. But … there is a certain slackness in the machinery which Putnam does not, I feel, prepare us for. … The slackness comes from how ordinary language terms for kinds are mapped onto the same classifications. In my story I have envisaged only a small wobble; how much latitude there might be in theory I do not know. … The "slackness" I have talked about seems to allow that from the very same linguistic base we may, after the very same scientific discoveries, move in different directions. (1983/2014, 199–200)

Donnellan's argument leaves the door ajar for possible stipulation in essence identification. However, his conclusion does not support Lemoine's exuberant claims. The indeterminacy Donnellan identifies merely indicates "a certain slackness … from how ordinary language terms for kinds are mapped onto" new scientific classifications and represents a "small wobble" from the standard determinate essentialist story (although he leaves open how large a wobble is possible). Rather than freely selecting from among alternative competing theories of water, there is one overwhelmingly supported chemical theory of water, but the discovery of isotopes, unanticipated by the vernacular concept's pretheoretical background assumptions, led to ambiguities in precisely how to map water onto chemical theory. If Donnellan is correct, then some rectification of chemical theory and the vernacular concept of water is required. Whether the rectification allows for stipulation in accordance with external interests as Donnellan suggests or is decided for chemists by the canons of scientific theory formation as Putnam holds remains disputed, as we shall see. Either way, the ambiguity is limited by the fact that it occurs within a theory that overall is understood to identify the nature of water. Although Donnellan argues that it is open to stipulation whether water is all of H_2O or just protium H_2O, he accepts that science has discovered, not stipulated, that the water is H_2O in some form that includes protium.

Joseph LaPorte (2004) systematically expands the Donnellan type of argument to additional areas of science and makes the case that essence indeterminacy often dominates over scientific discovery in identification of essences of vernacular kinds. Going beyond the standard element and substance-type examples, LaPorte argues, for example, that biologists, consistent with the discoveries they have made but contrary to their actual decisions, might have chosen to classify whales as fish and guinea pigs as rodents, yielding different essences of vernacular kinds. LaPorte distinguishes such classificatory decisions from changes in the meaning of a term, which everyone agrees can occur. Rather, there are areas of inherent boundary fuzziness in the vernacular concept yielding a choice of how to form categories. A theory of essence can be stipulated to resolve those ambiguous fuzzy cases one way or another, thus "precisifying" the concept rather than changing the term's meaning.

Alexander Bird (2010) lucidly summarizes LaPorte's position as follows:

> Concentrating on theoretical identities such as 'water is H2O', LaPorte argues that there is considerable vagueness in the use of kind terms, especially vernacular kind terms. … For a kind term 'K', some things will be determinately K and other things will be determinately not K. But there will be a boundary of things for which there is no determinate fact of the matter whether they are K or not. … According to LaPorte, when a natural kind identity is established as being determinately true, that is because scientists have made a *decision* to adopt the identity as true. In so doing, it will now be determined of items that were previously in the boundary (neither K nor not-K) whether they are K or not. For example, we now regard heavy water (deuterium oxide [D_2O]) as a subspecies of water; but scientists could have decided to exclude deuterium

> oxide from the extension of 'water'. So 'water is H2O' is true in virtue of a decision. That truth … is not the discovery of some previously hidden essence. Rather, it is an empirically motivated *stipulation*. (2010, 125)

Bird objects to LaPorte's view that "there is rather less room for conceptual choice and stipulation than LaPorte supposes" (125), for two reasons. First, the fuzzy boundary area is limited, so there is limited freedom for stipulative precisification: "Vagueness between red and orange leaves it determinate nonetheless that a ripe tomato is red. … The concept water may have open texture so that it is not determinate whether D_2O is water. But that is consistent with its being determinate that all water is H_2O" (Bird 2010, 135).

Second, science is more determinate than it seems because it eschews the kinds of practical concerns cited by LaPorte and others (e.g., Zachar 2002) as grounds for essence indeterminacy. Here Bird follows Putnam in holding that scientists have methodological standards for determining essences that are distinct from general interests. Bird argues that any choice other than identifying water with H_2O would violate the scientific canon that requires chemists to determine substances based strictly on chemical theoretical properties: "it will be chemical facts that determine the identity of substances. The chemical facts class D_2O with other kinds of H_2O" (2010, 127). Bird argues that, because the differences between D_2O and protium water primarily "come from outside chemistry," they "are not pertinent to the science whose job it is to investigate the nature of and to classify water" (2010, 127–128).

Bird's point seems fundamental. It is difficult to imagine how science could progress otherwise. The introduction into chemical theory of considerations of human concerns or practical uses would introduce myriad issues distant from what is needed to identify the classification that offers the deepest and most perspicuous understanding of how the world works. Such an approach would hobble the kind of scientific theory development that is chemistry's task. Practical concerns should be and are reflected in other available concepts and terminology but not in chemical theory per se. Thus, the nature of science's own standards for successful theory sets a limit on the scope for stipulation in essence identification.

Whatever one thinks of Bird's or Putnam's responses, the examples presented by Donnellan and LaPorte at least raise the possibility—if not in the discussed examples, then perhaps elsewhere—of the need for rectification of some degree of indeterminacy and stipulation in essence identification due to theoretical anomalies relative to vernacular background assumptions. However, the scope of indeterminacy suggested by these arguments is quite limited and does not alter the big picture of scientific accounts of essence. Even for Donnellan and LaPorte, water is a form of H_2O. None of the surveyed arguments open Lemoine's stipulationist spigot to radically different theories. There is nothing in this fascinating literature that would cast doubt on the overwhelming scientific primacy of the evolutionary explanation of the nature of natural functions. However, even if the evolutionary account is inevitably the theory of the essence

of natural functions, these debates suggest that there could there be a degree of indeterminacy and possible stipulation in the precise specification of the evolutionary essence of function. If so, this remains unexplored territory.

So, where does this leave Lemoine's first core claim that, because the HDA's evolutionary theory of function and dysfunction is not a conceptual analysis, one can choose to stipulate Boorse's biostatistical theory (BST) of function rather than the HDA's evolutionary account of function, as one pleases? The literature suggests that even if in rare cases stipulation is an option, its scope is quite limited and intratheoretic, concerned with nuanced indeterminacies in how the terms of a dominant essentialist theory are precisely mapped onto pretheoretical concepts and not a matter of freely selecting among theories. There can be little question that evolutionary theory is the dominant theory that explains the essential nature of natural functions and biological design. We have seen that the BST cannot be considered an alternative theory of the essence of "natural function" in the relevant sense because it does not correspond to the right kind of explanatory essence. The BST, being a variant of Cummins's CR-function approach, is by its nature not explanatory at the same level, Boorse's claims to the contrary notwithstanding. So, there is no competition and no support for Lemoine's proposed stipulative choice of the BST over the HDA as an account of "function" or his suggestion of a stipulative free-for-all. If there is any minimal domain of detail of terminological mapping open to stipulation, it is insufficient to support Lemoine's ambitious claims and poses no threat to the HDA's evolutionary component.

Is the BST a Better Naturalist Account of Dysfunction Than the HDA?

In the final section of his paper, Lemoine pivots from his focus on the HDA's account of dysfunction as a mere stipulation to a straightforward argument that the BST's statistical account is superior to the HDA's evolutionary component as a naturalistic account of dysfunction: "*Boorse's account of dysfunction is better than Wakefield's as a naturalist explication of the concept of dysfunction*." For those who, like Lemoine, have naturalist aspirations, this question of whether the evolutionary or statistical approach provides the best naturalist account of dysfunction is a crucial issue.

Lemoine argues that although the concept of evolutionary dysfunction is in principle not dependent on the concept of harm ("It is easy to understand how an ultimate evolutionary theory of what is dysfunctional could be contrasted with what we consider to be harmful or not"), we are unable, due to our relative ignorance of psychological and physiological causal mechanisms, to make dysfunction claims that do not rely on harm as a heuristic for dysfunction ("it is not easy to understand how we can consider dysfunctional states and harmful states to be different things in an imperfect state of knowledge"). (This criticism was also posed by De Block and Sholl in their chapter and addressed in my reply in this volume.) Lemoine then asks whether the HDA

has any strictly naturalist approach to identifying dysfunction without reliance on harm. He answers that, once divorced of the harm criterion, and given our ignorance of mechanisms and evolutionary history, the evolutionary account must rely on the BST's statistical approach to get off the ground, so the BST is the superior naturalist account.

Lemoine can think of only two possible harm-independent naturalist ways to recognize that a mechanism is failing to perform its natural function. The first way is by a statistical comparison of the mechanism's performance to the performances of analogous mechanisms in promoting survival and reproduction (S&R) in some reference class of other individuals; this is the naturalist statistical criterion proposed by Boorse. The second way is by directly appraising the degree to which the actual performance replaces and runs contrary to the presumed natural function of the underlying mechanism; this naturalist failure-of-function criterion is the one proposed by me. However, Lemoine argues that the failure-of-function criterion presumes knowledge of the proper natural functioning of the underlying mechanism, but "in an imperfect state of knowledge" (as Lemoine describes our situation of ignorance about most internal mechanisms and their functions), we do not have such knowledge, and so the HDA's distinctive failure-of-function approach cannot be used. Thus, the only way to appraise failure of function is to use Boorse's statistical comparison method to establish what is normal functioning: "I think that the only naturalist way to distinguish between non-natural and harmful effects of a mechanism in an imperfect state of knowledge is to adopt Boorse's biostatistical views on dysfunction."

Lemoine's analysis seems to run together conceptual and epistemological issues. Both the BST and the HDA aim to address the conceptual question of the nature of normal function and dysfunction. Lemoine's argument concerns the epistemological question of how we identify a dysfunction in circumstances of ignorance. However, Lemoine might reply that if the BST's conceptual analysis better illuminates the epistemology of dysfunction identification as we know it, then to this extent, it is a better account of the concept we actually use. So, I will take Lemoine's epistemological analysis as an indirect conceptual claim based on the claimed superiority of the BST in explaining dysfunction identification. I will return to the epistemological issue at the end of my analysis.

The HDA versus the BST on Setting the Mean and Range of Normal Functioning

Lemoine's argument that the evolutionary view ultimately must depend on statistics is based on his intuition that there is no other way to decide what is a normal function versus dysfunction. However, his argument needs elaboration and evaluation. Fortunately, the view that the evolutionary view ultimately rests on the statistical view has also been defended by Boorse himself, who claims that there are "reasons to think that no evolutionary approach can analyze biomedical normality without appealing

to statistics, as I do" (2002, 101). So, to understand how one might defend Lemoine's claim, I examine Boorse's arguments for the same point.

Boorse's primary argument is that a purely evolutionary account, unlike a statistical account, cannot "determine the mean and endpoints of normal function. … Even theoretically, it seems impossible … to avoid statistics" (2002, 101). He first considers the mean, asking, "How can evolution alone locate the mean of, say, normal human visual acuity?" (2002, 101). He pursues this question by analyzing a discussion of Neander's (1991b) on why penguins have poor vision when on land, Neander's answer being that their vision is primarily designed for seeing underwater in order to catch fish, and poor vision on land is a by-product of the way penguins' eyes were biologically designed in response to selective pressure for water vision. In response, Boorse says, "Neander confuses two questions: what the normal level of penguin vision is, and how one explains its origin. Penguin land myopia is normal because it is typical of penguins, not because it is somehow endorsed by evolution as a byproduct of something else, underwater visual acuity" (101–102).

However, it is Boorse who confuses two questions: what is the current statistically typical level of penguin visual acuity on land, and what is the biologically designed normal level of acuity? To that extent, the BST represents a confusion of the concepts of statistical normality and functional normality. That penguins statistically see with modest visual acuity on land can have several explanations, only one of which is that their eyes are biologically designed to see with that acuity on land. Alternative hypotheses tend to be improbable, and given the reach of biological design, we generally rely on what is statistically typical as the defeasible default hypothesis to tell us what is likely normal. However, various alternatives lurk in the background and reveal that the typical need not be the normal. For example, one alternative hypothesis—analogous to there being almost universal gum disease and tooth decay—is that penguins are subject to an almost universal eye infection that limits visual acuity on land. Another is that environmental conditions have changed in a way that creates vision-obscuring atmospheric distortions or allergens in the penguin habitats that has reduced penguins' previously much sharper terrestrial vision. A third is that penguins now suffer from a critical-period developmental dysfunction due to lack of some expectable stimulation that triggers development of greater land-based visual acuity, in the way that early close-focus visual experiences may cause near-sightedness in humans.

Boorse asks, "If it is as easy as Neander thinks for whole species to be diseased, why are penguins not diseased for not seeing well both on land and in the sea? … Why was that genetic deficiency not itself a pandemic penguin disease?" (2002, 101). The answer is that there was no "genetic deficiency." The level of visual acuity on land results from or is a by-product of the way penguins' visual system is biologically designed through natural selection and can be explained by the pattern of selective pressures exerted

on the penguin population. The answer to Boorse's first question is that the mean of normality on a dimension is determined by the mean of the range for which it was naturally selected.

Boorse's second question concerns the range of normal function: "How can a purely evolutionary concept set boundaries to the normal range? If we seek to capture the biomedical idea of normality, it must be possible to have even pathologically myopic penguins. … So how does penguins' evolutionary history determine a lower limit of normal penguin myopia? Pending such an explanation, I conclude that … even an etiological theory requires a concept of statistical normality to match basic logical features of biomedical concepts" (2002, 102).

Again, the answer at a theoretical level is obvious. Although there will be a fuzzy boundary zone as there is for most legitimate conceptual distinctions (day/night, red/orange, child/adult), the setting of the normal range is basically a matter of judging the range over which positive selective pressures played a significant role in shaping the capacity in question through an effect-sensitive causal process. Boorse's "in theory" challenge is thus answered, although Lemoine's question about how this works in a state of ignorance still needs to be addressed (see below).

There is an irony in Boorse's critique of the evolutionary account for not specifying the range of normality. Boorse fails to consider how well the statistical view does in comparison, and it seems assumed that a statistical view must automatically resolve statistical-like questions about range. However, in fact, Boorse's statistical view has no answer to the question of how to set the range of normality of a function, declaring the boundary between normal function and dysfunction to be wholly arbitrary: "the lower limit of normal functional ability—the line between normal and pathological—is arbitrary" (1987, 371); "the term 'normal functional ability' had been defined dispositionally, as the readiness of an internal part to perform all its normal functions on typical occasions with at least typical efficiency. 'Typical efficiency' of a part-function, in turn, is efficiency above some arbitrarily chosen minimum in its species distribution" (1997, 8); "the BST is consistent with disease prevalence of 35%, 20%, 5%, 1%, or, I suppose, even 0%, and with prevalence varying from disease to disease. What it is inconsistent with is prevalences ≥50%" (2014, 714). According to the BST, the prevalence of dysfunction for any function can be decided arbitrarily anywhere from 0% to 49%, and there is no further conceptual reason or justification for favoring any given level, only pragmatic reasons extraneous to the concept of dysfunction itself.

These limitations of the BST mean that accepting Boorse's—and Lemoine's—position would be disastrous for achieving one of the primary goals that motivated the search for a definition of disorder in the first place: to limit false-positive diagnoses in which social deviance is mislabeled mental disorder and thus to respond to antipsychiatric claims that psychiatric diagnosis is misused for social control purposes by creating overly inclusive categories. Boorse's view provides a conceptual warrant for arbitrarily

pathologizing up to half the population on every single functional variable without any conceptual recourse, a breathtakingly ill-considered approach.

The Pandemic Disease Objection to a Statistical Criterion for Dysfunction

The above discussion of the range of normality indicates a bewildering feature of the BST's statistical approach to dysfunction. The BST's statistical subtypicality account of dysfunction implies that a dysfunction cannot be typical; that is, it cannot occur in more than half of the population. Critics have rightfully taken this claim to be demonstrably false and argued that there is nothing contradictory or even puzzling about statistically common disorders: "There is nothing incoherent in the idea of typical dysfunction, as our concepts of epidemics and pandemics attest" (Neander 2012, 2). If correct, this objection implies that, although most dysfunctions are subtypical, there must be some criterion beyond statistics that forms a backbone for the concepts of function and dysfunction and overrides the statistical approach.

Lemoine ignores this problem, whereas Boorse, to his credit, squarely confronts this problem, admitting that, contrary to the BST, the concept of medical disorder clearly allows for dysfunctions occurring in a majority of a species: "Any account of normality must concede that medicine recognizes a tiny number of diseases that are typical or even universal, either in the whole species (atherosclerosis) or in an age group (osteoarthritis or prostatic cancer in men of a certain age)" (2002, 102–103).

The problem is larger than Boorse suggests. First, the list of such pandemic conditions could easily be expanded beyond Boorse's "tiny" number. For example, tooth decay and gum disease afflict about 90% of humanity and apparently have for a long time, judging from jawbone and tooth remains, and Boorse elsewhere notes that, on his view, if children generally suffer bruised knees, that pandemic condition would not be a disorder. (I note in passing that if Boorse were correct that any dysfunction is a disorder, then the problem would be much larger because there are many pandemic dysfunctions, such as mutations in skin DNA due to sunlight exposure or some number of dysfunctional sperm, that would then constitute pandemic disorders, but since he is plainly incorrect—a pathologist would label a sperm without a tail as a dysfunctional sperm but not thereby necessarily label the individual as medically disordered because there is no harm—I leave this problem aside.) Boorse uses the example of prostate cancer, but the disorder of benign prostatic hyperplasia, which can obstruct urine flow to the point of retention and can lead to kidney damage, is a better example, with about half of men in their fifties and about 90% of men in their eighties suffering from this condition. Second, given that Boorse claims that the BST is a conceptual analysis, an additional problem is the endless number of *possible* pandemic diseases one can easily imagine developing or being discovered. For example, one can easily imagine humanity generally suffering from a dramatic increase in antibiotic-resistant infections

coincident with the worldwide spread of infectious disorders due to global warming or the discovery of formerly unrecognized almost universal parasites. Though counterfactual, such examples of what is clearly possible represent legitimate counterexamples to the BST.

So, how does Boorse address this issue? Rather than conceding that the existence of medically acknowledged pandemic disorders requires abandonment or modification of a statistical view of disorder, he says that "the question is how we should explain this fact" (2002, 102–103). He notes that he has changed his mind more than once on the explanation and offers his then-latest view: "I currently favor the view that medicine is wrong to recognize any universal diseases, since it lacks any coherent concept of pathology that can make them pathological. On this view, what is pathological is only age excessive atherosclerosis, premature prostate cancer, and so on … I will embrace the conclusion of my analysis. If nearly all human left legs have been broken throughout human history … , then that is their normal condition" (2002, 103). Thus, in the face of seemingly conclusive counterexamples, Boorse offers no explanation but rather simply insists on his view and rejects the judgments of the medical field. Nothing Lemoine says extricates him from this problem with the statistical view. The most plausible "explanation" of the facts is simply that Boorse's analysis is incorrect. As to Boorse's claim that there is no coherent view that explains the conceptual possibility of pandemics, of course he is wrong there too, for the HDA's evolutionary account readily explains such judgments.

Reference Classes and the Myth of a Statistical Theory of Dysfunction

I now come to the foundation of the statistical approach to distinguishing normal function from dysfunction endorsed by Lemoine, namely, the process by which the typical and subtypical are statistically identified. To understand why Lemoine's faith that evolutionary judgments must rest on the BST's statistical judgments is misguided, one needs to examine the conceptual bedrock underlying the statistical judgments themselves. So, in this section, I examine Boorse's notions of normality and dysfunction and the BST's critical notion of "reference classes" that underlies the identification of the statistically typical and subtypical.

Boorse defines normal function and dysfunction in strictly statistical terms of typical versus atypical levels of contribution to survival and reproduction (S&R): "normal function of a part or process is a statistically typical contribution by it to their individual survival and reproduction" (1977, 555); "medically normal function of any token item (for example, a single human heart) is analyzable as an output within a statistically typical range of contributions to survival and reproduction by tokens of that type in an age group of a sex of a species" (2002, 72); "what is pathological in medicine is statistically subnormal … function" (2002, 94).

A basic challenge for the BST's statistical conception of dysfunction is that statistical typicality and deviance measures vary depending on the "reference class" that forms the background for the measure. For example, relative to sighted people, a blind person has eyes that make a subtypical contribution to S&R, but relative to other blind people, a blind person's eyes may make a typical contribution to S&R. So, for a statistical account of dysfunction as subtypical S&R contribution to make sense, an account must be provided of the reference classes on which the statistical claims are based. The reference class cannot simply be the entire human race because, for example, then children who normally have less capability than adults could be labeled as dysfunctional and women (who comprise slightly less than half the world population) could be classified as having a dysfunction due to their lack of various male organs and processes. Thus, some more refined reference class must be defined to distinguish normal function from dysfunction. Boorse is well aware that his account of normal function and dysfunction requires the specification of such reference classes: "I make normal function in physiology or medicine a statistical concept, involving generalization over a reference-class. I defined medical normality as 'the readiness of each internal part to perform all its normal functions on typical occasions with at least typical efficiency'—that is, at an efficiency level not far below the reference-class mean" (2002, 90).

The problem for the BST is how to identify such reference classes in a way that yields results consonant with medical intuitions without invoking evolutionary theory. For example, benign prostate hyperplasia (BPH) is considered a disorder, according to the HDA, based on the judgment that in BPH, the biologically designed functioning of the urinary and prostate system is harmfully failing. This evolutionary judgment, Lemoine would claim, is based on the statistical abnormality of the features of urinary and prostate functioning in BPH. However, relative to what reference class is BPH statistically subtypical functioning? One cannot use all human beings as the reference class because half are women without prostate glands, and that lack is not a dysfunction. Nor can one use all males because children do not yet have fully functioning prostates and would skew the results. At the other extreme, if one uses just those males who present for urinary flow problems, then what is truly a dysfunction would be classified as statistically typical and thus normal. So, one has to choose the group that is "just right." If that class is all male adults, then BPH is statistically infrequent. If it is limited to the age-sex class of males over fifty years old, then more than half have some degree of BPH and (counterintuitively) it would not be a disorder in that age-sex cohort, according to the statistical approach. So, the identification of reference classes is crucial for the statistical view.

"For medical purposes," Boorse defines a reference class as "a natural class of organisms of uniform functional design: specifically, an age group of a sex of a species" (Boorse 1977, 555; Boorse 2002, 90). Note that the use of the term "functional" in "uniform functional design" is not circular or begging the question of how to establish

normal function versus dysfunction because all that "functional" means for Boorse is the causal contributions made by various parts under various circumstances to S&R. Thus, "design" with its evolutionary connotation is a bit misleadingly teleological, and elsewhere Boorse uses "functional organization" in explaining what he means: "Once one knows that functions are causal contributions to goals of the organism, one can classify the functional organization of different individuals—that is, the ways their parts contribute to their survival and reproduction—as similar or dissimilar" (2002, 91). In theory, the levels of such contributions under varying circumstances can be established independently of and without any reference to normality versus dysfunction, which are identified at a later stage of the analysis by the levels of contribution to S&R in the reference class that are typical and subtypical, respectively.

Boorse's above definition of reference classes contains two criteria. He first specifies that the class must be of "uniform functional design" and then elaborates that it is, "specifically, an age group of a sex of a species." This raises the question of the relationship between the two criteria, and which is primary. The "uniform functional design" criterion is clearly intended as the rationale for the specific sex and age dimensions of the reference classes; otherwise, the specification of age and sex appears arbitrary. However, in "specifically" indicating age and sex subgroups, Boorse appears to be claiming that these are the only dimensions that yield reference classes of uniform functional design, or at least the ones he is selecting as relevant to medical judgments. (He at times has suggested that perhaps race would be another such dimension, but I set that issue aside here.) However, it is not obvious that limiting reference classes to sex and age dimensions follows from the "similar functional organization" criterion. Consequently, given the potential divergence between the two criteria, in evaluating Boorse's approach to reference classes, it seems charitable to consider separately each of the two possible approaches he suggests, one that specifies that reference classes are limited to age and sex categories, and the other that relies on the general characterization of "uniform functional organization."

I start by evaluating the specification of age and sex categories as the unique reference classes. The idea that age and sex divisions are legitimate reference class divisions may seem innocent enough given that these groupings (e.g., male versus female, child versus adult) have evolutionarily shaped differences. However, using age and sex reference classes independently of evolutionary judgments allows diagnostic absurdities that are inconsistent with medical thinking and thus falsify the statistical view.

First, if reference classes are limited to age and sex categories, then all those naturally selected features that result from such processes as niche selection or balancing selection are in danger of being considered dysfunctions by virtue of their comparison to typical functioning in species-level age and sex categories. Yet, these processes produce specific adaptive features in response to specific environmental contexts—for example, lactose tolerance in cultures with the availability of milk, sickle cell trait in

environments with endemic malaria, blood alterations in high altitudes—that are considered normal functioning and not disorders. Age and sex reference classes cannot take account of more fine-grained normality based on such niche natural selection.

Another domain of falsifications concerns the BST's prediction of allowable age-dependent medical diagnoses. If the BST were correct, then whether a kind of condition could be considered a disorder in a given age cohort would depend on the statistics of the condition at that age, and the answer can vary from age to age. One could delay coming to the physician for a year and find that one's condition, which has stayed entirely constant, was a disorder the year before but is no longer so because it afflicts the majority at your older age. Aside from Boorse's examples mentioned earlier of children with bruised knees and prostate cancer or atherosclerosis in the elderly not being disorders, it is possible that, before the advent of vaccination, children of a certain age who contracted measles and elderly who contracted pneumonia were not disordered by the statistical criterion. None of these predictions comport with medical thinking.

Moreover, for many developmental stages, from puberty and menopause to children learning to walk or speak, there is enormous normal variation in the age at which the changes occur and physicians do not consider such variations in themselves to be disorders. However, the BST, using age to define reference classes, potentially pathologizes almost half of the normal-variation population. For example, the age of onset of puberty in females and thus the capacity for childbearing, which has direct S&R implications, varies greatly, and there is an age at which puberty has occurred in the majority of females and thus has become species typical, so the BST using age cohorts for reference classes allows all those who are still not pubescent because they fall in the later part of the onset curve to be classified as disordered. (Thus, age-cohort reference classes undermine any advantage Boorse might claim for the statistical account's ability to yield means for typical functioning.)

Needless to say, this is simply not how medical diagnosis works. Medicine considers such differences in age of reaching developmental milestones up to a point to be normal variation and generally does not change a diagnosis in response to the changing statistics at the patient's age. Given that Boorse presents the BST as a conceptual analysis of medical thinking, these radical divergences from medical thinking are legitimate counterexamples that falsify the BST.

The age-and-sex account of reference classes is thus both too refined and too unrefined. It is falsified both by the existence of normal naturally selected categories (e.g., lactose tolerance) that require more refined reference classes and by the pathologization of normal variation (e.g., age of puberty) that occurs with the use of overly refined age cohorts as reference classes. These falsifications are avoided by the evolutionary natural-function approach.

This suggests that the more charitable interpretation of Boorse's view may be to abandon the age-and-sex approach to reference classes and give priority to the more abstract

"uniform functional organization" characterization of reference classes, to which I now turn. Despite Boorse's equating of the two approaches, they diverge because functional organization in Boorse's sense of a part's contribution to S&R can vary enormously within sex and age categories, yielding "similar functional organization" reference classes that do not correspond to age and sex categories. For example, twenty-five-year-old women who are blind do not have the same functional organization as normally sighted women of the same age, given that their eyes and their senses of hearing and touch perform such different roles in their survival and reproduction, nor does the pancreas play the same functional role in those with and without diabetes. This problem is not limited to disorders that confer distinct functional organizations but also applies to myriad normal variations. For example, physical appearance plays a different functional organizational role in those who are attractive versus those who are homely, and anxiety plays a different functional role in those high and low on normal-range neuroticism. So, given Boorse's definition of "uniform functional design" as similar contributions of the parts to S&R, reference class distinctions can go well beyond sex and age differences.

The implications of this proliferation of potential reference classes for the statistical view of normal function versus dysfunction are devastating. For example, consider again the distinction between blind and normally sighted twenty-five-year-old females. If they were included in the same reference class of twenty-five-year-old females, then the former women could be judged as having a vision dysfunction because of a lesser contribution to S&R by their eyes relative to others' eyes in that reference class. (It is not clear what one would say about females born without eyes, but I leave that sort of complication aside.) However, as noted, it would seem that the two groups of women do not have similar functional organizations by Boorse's criterion because of the radically different roles that their eyes and other senses play in promoting S&R. Thus, according to the uniform-functional-organization criterion, the blind women would (or could) be placed in a distinct reference class of blind female adults and not the same reference class with sighted female adults. This creates a major problem for the statistical view, because within the reference class of blind female adults, a given blind woman's eyes may well make a typical (low) level of S&R contribution and so she would not be judged to be subtypical and would not be considered to have a vision dysfunction—even though in fact she is blind and has an obvious dysfunction and medical disorder. There are endless variations in functional design—indeed, almost any significant genetic variation may define a distinct functional design (in the sense of a distinct pattern of the parts' contribution to S&R)—implying that reference classes will proliferate, dividing along myriad lines ranging from darkly pigmented skin versus less pigmented skin to pygmy versus Watusi height and morning people versus night people, and so on.

The problem this poses for the statistical view is that if reference classes can be so fine-grained, then the entire system for recognizing dysfunctions as subtypical

functional contribution collapses because dysfunction, from kidney failure to schizophrenia, generally implies a different functional organization than those with normal functioning. That is, the uniform-functional-organization criterion for reference classes threatens to create separate baselines of "normal" functioning for every significant human variation, including every significant genetic normal variation but also every significant dysfunction, thus effectively erasing the distinction between medical disorder and health and undermining the very point of the statistical analysis.

Note that it would be circular for Boorse to try to escape this problem by saying that the blind and the sighted are of uniform design and belong in the same reference class but only seem different because one group has a dysfunction that interferes with that design, because at this stage of the analysis, there is not yet a concept of dysfunction available; the whole point of independently defining the reference classes is to then identify normal function versus dysfunction. Thus, Neander argues that Boorse's definition of reference classes "involves an intolerable circularity" (Neander 1983, 94): "For instance, those with adult onset diabetes might be said to have a 'uniform functional design', if we consider just the actual causal roles of traits. But we would not want to count those with adult-onset diabetes as a distinct reference class or else it will follow that their condition would then count as normal. How to discount this reference class, however, but for the fact that the peculiarities in functioning of those in the class are dysfunctions? One can speak of what is normal in the sense of typical for adult-onset diabetes, but what is normal in the sense of typical for those with adult-onset diabetes involves malfunction" (Neander 2012, 1).

If it seems remotely possible that diabetics and nondiabetics might be seen as having similar functional organizations, then consider more extreme cases such as the class of individuals with the lack of both sight and hearing, such as Helen Keller, or the class of individuals born without multiple limbs (e.g., due to their mothers taking thalidomide). There seems to be no way, without prior reference to natural functions and dysfunction, to argue that such individuals have the same functional organization as the functionally normal individuals without those infirmities and thus no noncircular way to distinguish the dysfunctions from the normal functions using the statistical approach. The only sense in which those without sight and hearing are similar in functional organization to those with intact senses is if one equates their functional organization with their evolutionary biological design and observes that the deviations are not part of biological design. That, however, would be to admit that the statistical view cannot succeed without reference to evolutionary presuppositions.

Boorse attempts to answer the reference-class circularity objection by explaining that he doesn't need to know what a dysfunction is prior to judging uniformity of functional design and forming reference classes: "The circularity charge rests on a confusion. Once one knows that functions are causal contributions to goals of the organism, one can classify the functional organization of different individuals—that is, the

ways their parts contribute to their survival and reproduction—as similar or dissimilar" (2002, 91).

This answer does not resolve the problem. It is true, as noted above, that there is no circularity in the reference to "functional design" because Boorse defines that as the pattern of S&R contributions, which does not presuppose anything about normality versus dysfunction. The problem lies with "uniform." It begs the question for Boorse to assume without any explanatory account that, say, the blind will be judged similarly enough functionally organized to be placed in a common reference group with the sighted so that the they come out statistically dysfunctional relative to the others in their reference group. Some severe pathologies are likely to be judged organizationally different from normality no matter where the "similarity" boundary is located, and then they will be placed in their own reference classes in which they are statistically typical, misclassifying them as normally functioning.

The statistical view using the uniform-functional-organization criterion for reference class formulation, like the age-and-sex account of reference classes considered earlier, fails to explain the distinction between dysfunction and normal function in a way that is noncircular, properly reflects medical judgments, and does not rely on prior evolutionary considerations of biological design. I conclude that the statistical account of dysfunction on which Lemoine rests his case against the HDA has no coherent foundation without reference to the HDA's evolutionary considerations.

Conclusion

Lemoine is surely correct that statistical comparisons are important to intuitive function judgments. When we look around and intuitively judge what for our species are normal functions versus dysfunctions, we often depend to some extent on statistical observations and comparisons. Eyes seeing, hands grasping, fear taking us away from danger, and so on impress us as biologically designed partly because they are obvious, common features of our species. Generally speaking, naturally selected functions become species typical, so species typicality is a powerful guide in assessing normal function. Statistical commonality is thus a useful epistemological adjunct to basic intuitions about candidates for biological design and normal function, especially when evolutionary history is unknown.

However, one has to distinguish epistemological from conceptual claims. Epistemologically, of course, initial judgments about what is a natural function often do rely on some statistical observation based on species typicality. However, conceptually the issue is whether the feature is biologically designed, that is, whether natural selective pressures led to the feature. Any evidential pathway, statistical or otherwise, that you can take to gain such understanding can lead to legitimate function judgments. Indeed, modern genetic analysis techniques are giving us new ways to establish what

was and was not naturally selected (e.g., Ding et al. 2002; Lind et al. 2019; Polimanti and Gelernter 2017) without the need for a time machine. It is true that such methods involve such statistical methods as cross-species and intraspecies comparisons of mutation rates around target loci, but such tests of genetic stability are used to establish the likelihood of positive natural selection that may then be used in inferences about normality and dysfunction; they are not used to directly infer normality and dysfunction from statistics alone without reference to natural selection. Such methods are at this time largely limited to single-loci analyses, but it is an open and growing area of research. Consequently, Lemoine's single-minded focus on the epistemological situation of our being in a perpetual state of gross ignorance about natural selection is scientifically out of date and excessively pessimistic. Additionally, regarding Lemoine's claim that the purely statistical view of normal function and dysfunction is presupposed by the HDA's evolutionary approach and thus a better naturalist account of dysfunction, the analysis reveals that, due to challenges in identifying reference classes, the statistical approach cannot get off the ground on its own steam without an implicit appeal to biological design. Thus, the statistical approach implicitly depends on the evolutionary approach.

In his attempt to find an alternative to an evolutionary account of natural functions, Lemoine fails to critically examine the coherence and evidential plausibility of the statistical approach before embracing it. Close examination reveals that the statistical account is not viable as an exclusive stand-alone statistical foundation for medical judgments for a variety of reasons. Biological design literally overrides statistics in judging dysfunction in ways that provide persuasive counterexamples to the statistical subtypicality account's necessity and sufficiency for dysfunction. Regarding necessity, statistically typical and pandemic conditions (e.g., tooth decay and gum disease, atherosclerosis) can be medically classified as dysfunctions, and regarding sufficiency, many conditions that are subtypical (e.g., high neuroticism, homeliness) or are present in a minority (sickle cell trait, lactose tolerance) are medically classified as normal variations. The absurdity of the age dependency of dysfunction accepted by Boorse (e.g., knee bruises and prostate cancer are not dysfunctions at certain ages), the conceptual allowability according to the BST of arbitrarily pathologizing almost half of the population on any functional variable without any further conceptual justification, and most of all the inability of the proposed reference classes to adequately separate dysfunction from normal function all provide compelling objections and counterexamples to anything like the BST's statistical theory. Contrary to what initially might seem like common sense, one cannot simply go out and do statistical analysis to decide the normal versus dysfunctional status of a feature without some further guidance. Given that Lemoine embraces the BST precisely because it is by far the most sophisticated statistical theory around, the most reasonable conclusion is that the statistical approach to dysfunction must be rejected and that the idea that a purely statistical approach can escape theoretical loading while adequately identifying dysfunction is a myth.

Finally, let me return to a question raised earlier: if Lemoine is wrong in arguing that the evolutionary account of dysfunction rests on the statistical account, then what is the correct relationship between the naturalistic components of the BST and the HDA? Boorse has often been criticized for his selection of survival and reproduction as the effects that determine a function as an implicit value judgment. One might also argue that in selecting these criteria, he is trying to mimic evolutionary theory without acknowledging its role. In my view, the explanation of Boorse's choice is more benign. Boorse has correctly observed the fact that long before evolutionary theory and reaching back to antiquity, survival and reproduction are the two domains that were routinely used to justify function attributions.

What Boorse fails to observe is that there is a deeper rationale for why the two specific areas of survival and reproduction are linked to function judgments. The reason is that they are the areas that most distinctively manifest the mystery of biological design, from eyes seeing to acorns growing into oak trees. That is, while eschewing the theoretical explanatory implications of "natural function" linking it to biological design, Boorse has nonetheless picked as his criterion for the overall function category the causation of precisely those outcomes that form the most plausible base set of biologically designed phenomena for the definition of "natural function."

Thus, the relationship between the S&R concept of function and the etiological concept of function is simply that some S&R functions—where the choice is guided by intuitions about biological design—form the base set for defining natural etiological functions. This would make natural functions and S&R functions complementary parts of one account rather than competitor analyses, like the relationship between the base-set description "clear thirst-quenching liquid in lakes and rivers" and the category of (substance) "water" as anything with the same chemical essence as the base-set examples. According to this integration of the views, the S&R view would descriptively pick out some presumptive instances of a natural kind of "natural functions" and thus would indeed be an epistemologically accessible subset of the larger essentialistically defined category of such functions, where the larger category defined in terms of the base set can encompass instances that do not fit that initial base-set descriptive criterion.

Interestingly, this possibility is hinted at in a passing prescient remark of Lemoine's. Near the end of his paper, he mentions that one possibility is for the HDA to "adopt Boorse's view on dysfunction, at least as a temporary stage, in providing a naturalist account of dysfunction." If one interprets that "temporary stage" as the formulation of the base set of apparently biologically designed effects that forms the foundation for the fuller black-box essentialist definition of "natural function," then the two views can be integrated, with Boorse's view capturing the more readily observable base set of species-typical intuitively biologically designed features that requires further explanation in terms of some explanatory nature that forms the essence of natural functions.

References

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association.

Bird, A. 2010. Discovering the essences of natural kinds. In *The Semantics and Metaphysics of Natural Kinds,* S. Beebee, N. Sabbarton-Leary, and F. Longworth (eds.), 125–136. Routledge.

Bolton, D. 2000. Alternatives to disorder. *Philosophy, Psychiatry, and Psychology* 7(2): 141–153.

Boorse, C. 1987. Concepts of health. In *Health Care Ethics: An Introduction,* D. Van DeVeer and T. Regan (eds.), 359–393. Temple University Press.

Boorse, C. 1997. A rebuttal on health. In *What Is Disease?* J. M. Humber and R. F. Almeder (eds.), 1–134. Humana Press.

Boorse, C. 2002. A rebuttal on functions. In *Functions,* A. Ariew, R. Cummins, and M. Perlman (eds.), 63–112. Oxford University Press.

Boorse, C. 2014. A second rebuttal on health. *Journal of Medicine and Philosophy* 39: 683–724.

Cosmides, L., and J. Tooby. 1999. Toward an evolutionary taxonomy of treatable conditions. *Journal of Abnormal Psychology* 108(3): 453–464.

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

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.

Ding, Y. C., H. C. Chi, D. L. Grady, A. Morishima, J. R. Kidd, K. K. Kidd, et al. 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proceedings of the National Academy of Sciences of the United States of America* 99(1): 309–314.

Donnellan, K. 1983/2014. Kripke and Putnam on natural kind terms. In *Essays on Reference, Language, and Mind,* J. Almog and P. Leonardi (eds.), 179–203. Oxford University Press.

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.

Kendler, K. S., P. Zachar, and C. Craver. 2011. What kinds of things are psychiatric disorders? *Psychological Medicine* 41(6): 1143–1150.

Kripke, S. A. 1980. *Naming and Necessity*. Harvard University Press.

LaPorte, J. 2004. *Natural Kinds and Conceptual Change*. Cambridge University Press.

Lind, A. L., Y. Y. Y. Lai, Y. Mostovoy, A. K. Holloway, A. Iannucci, A. C. Y. Mak, et al. 2019. Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nature Ecology & Evolution* 3(8): 1241–1252.

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

Neander, K. 1991a. Functions as selected effects: The conceptual analysis defense. *Philosophy of Science* 58: 168–184.

Neander, K. 1991b. The teleological notion of function. *Australasian Journal of Philosophy* 69: 454–468.

Neander, K. 2012. Biological function: Statistical theories of function. *Routledge Encyclopedia of Philosophy*. https://www.rep.routledge.com/articles/thematic/biological-function/v-1.

Polimanti, R., and J. Gelernter. 2017. Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genetics* 13(2): e1006618.

Price, C. 1995. Functional explanations and natural norms. *Ratio* 7: 143–160.

Putnam, H. 2015a. Intellectual autobiography of Hilary Putnam. In *The Philosophy of Hilary Putnam,* R. E. Auxier, D. R. Anderson, and L. E. Hahn (eds.), 3–110. Open Court.

Putnam, H. 2015b. Reply to Alan Berger. In *The Philosophy of Hilary Putnam,* R. E. Auxier, D. R. Anderson, and L. E. Hahn (eds.), 332–336. Open Court.

Putnam, H. 2015c. Reply to Ian Hacking. In *The Philosophy of Hilary Putnam,* R. E. Auxier, D. R. Anderson, and L. E. Hahn (eds.), 358–364. Open Court.

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.

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. 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. 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. 2011. 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. 2012. Are you as smart as a 4th grader? Why the prototype-similarity approach to diagnosis is a step backward for a scientific psychiatry. *World Psychiatry* 11: 27–28.

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. 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., 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.

Wilson, M. 1982. Predicate meets property. *The Philosophical Review* 91(4): 549–589.

World Health Organization. 2018. *International Statistical Classification of Diseases and Related Health Problems*. 11th rev. https://icd.who.int/browse11/l-m/en.

Zachar, P. 2002. The practical kinds model as a pragmatist theory of classification: Comment. *Philosophy, Psychiatry, and Psychology* 9(3): 219–227.

Defining Mental Disorder

Jerome Wakefield and His Critics

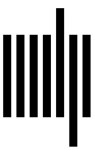
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl

Edited by: Luc Faucher, Denis Forest

Citation:

Defining Mental Disorder: Jerome Wakefield and His Critics
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl
Edited by: Luc Faucher, Denis Forest
DOI: 10.7551/mitpress/9949.001.0001
ISBN (electronic): 9780262362931
Publisher: The MIT Press
Published: 2021

The open access edition of this book was made possible by generous funding and support from Arcadia – a charitable fund of Lisbet Rausing and Peter Baldwin



The MIT Press

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Faucher, Luc, 1963– editor. | Forest, Denis, editor.
Title: Defining mental disorder : Jerome Wakefield and his critics / edited by Luc Faucher and Denis Forest.
Description: Cambridge, Massachusetts : The MIT Press, [2021] | Series: Philosophical psychopathology | Includes bibliographical references and index.
Identifiers: LCCN 2020016671 | ISBN 9780262045643 (hardcover)
Subjects: LCSH: Wakefield, Jerome C. | Psychiatry--Philosophy. | Mental illness--Philosophy. | Mental illness--Diagnosis. | Mental illness--Classification.
Classification: LCC RC437.5 .D434 2021 | DDC 616.89--dc23
LC record available at https://lccn.loc.gov/2020016671