Defining Mental Disorder

Jerome Wakefield and His Critics

The title-level DOI for this work is:

doi:10.7551/mitpress/9949.001.0001

24 Is Indeterminacy of Biological Function an Objection to the Harmful Dysfunction Analysis? Reply to Tim Thornton

Jerome Wakefield

I thank Tim Thornton for his chapter exploring the possibility and limits of naturalism in my harmful dysfunction analysis (HDA) of medical, including mental, disorder. The HDA claims that "disorder" refers to "harmful dysfunction," where dysfunction is the failure of some feature to perform a natural function for which it is biologically designed by evolutionary processes and harm is judged in accordance with social values (First and Wakefield 2010, 2013; Spitzer, 1997, 1999; Wakefield 1992a, 1992b 1993, 1995, 1997a, 1997b, 1997c, 1997d, 1998, 1999a, 1999b, 2000a, 2000b, 2001, 2006, 2007, 2009, 2011, 2014, 2016a, 2016b; Wakefield and First 2003, 2012). I have long benefited from Tim Thornton's extensive work not only in philosophy of the mental health professions but also philosophy of science and mind as well, and he brings his sophisticated understanding of these multiple domains of philosophy to his chapter's exploration of the HDA's naturalist evolutionary analysis of "biological function."

At the beginning of his paper, Thornton astutely draws attention to a cardinal feature of the HDA. In opposition to value-based approaches such as Fulford's (1989) that see values as foundational and pervading every component of "disorder," the HDA distinguishes a purely naturalist necessary component of "disorder" from a value component. This allows for a naturalist value-free zone of assessment and debate over diagnostic validity, offering a partial escape from values in the sense that one can critique disorder claims evidentially and scientifically without always having to address values. Nonetheless, due to its value component, the HDA will never satisfy those who are engaged in a thoroughgoing philosophical naturalist project. Although in a profession like medicine, values may not be fully escapable in its foundational concepts, the HDA's two-component approach allows that values can sometimes be put aside in a purely naturalist assessment of diagnostic validity based on questioning the presence of evolutionary dysfunction. The general issue Thornton raises, by way of a discussion of Millikan's ambitious project to naturalize meaning in terms of biological function, is whether the HDA's naturalist "function" component can remain truly value free given potential Wittgenstein-inspired indeterminacy issues in the specification of functions.

Thornton lucidly addresses a challenge to the HDA's position on separate value and factual components posed by Fulford (1999). Fulford argued that the "dysfunction" component of the HDA is in fact normative because it requires the "failure" of a function, and "failure," he claimed, is irreducibly value laden. (Apparently for Fulford, the fact that the latest Brexit proposal "failed to pass" in Parliament or that an argument "fails to be valid" is a value judgment.) "Failure" is indeed often used to indicate lack of success in value-defined outcomes, but in the context of the HDA, "failure of function" is just a way of saying that a mechanism did not perform its biologically designed function. As Thornton observes, "If function and divergence from it can be analysed in descriptive terms then so can 'failure' of function: it is any divergence from function." Thus, in disputes over disorder status, there is at least a potential "place to stand" from which one can offer a factual/scientific/naturalist critique of a judgment that a condition is a disorder. In many instances, this is a much more potent tool of critique than clashing value judgments or radical antipsychiatric arguments that deny the obvious fact that there really are mental disorders.

A small point: Thornton mentions that Robert Kendell, perhaps the most respected psychiatrist other than Robert Spitzer on the topic of the concept of mental disorder, initially grappled with an all-factual account. He also flirted with the factual but hopeless "whatever psychiatrists treat" notion at one point. It is perhaps of some slight interest where his odyssey ended on this issue. In one of his last papers before his sudden death, he endorsed the HDA: "It also would be well worthwhile revising the basic *Diagnostic and Statistical Manual of Mental Disorders* (*DSM)* definition of mental disorder in light of Wakefield's (1992) cogent analysis of the concept. … Having struggled myself (Kendell 1975, 1986) to decide whether disease or disorder are better regarded as normative concepts based on value judgments or as value-free scientific terms, I am impressed by Wakefield's arguments that both elements are necessarily involved" (2002, 5).

Another small point: Thornton incorrectly suggests that my insistence on the term "disorder" comes from Boorse. However, it was in fact in opposition to Boorse's confusing use of "disease" that I adopted "disorder." "Disease" has a common meaning that excludes medical disorders such as injuries and poisonings—and which Boorse eventually abandoned. "Disorder" has actually been used for centuries for this reason.

I am not going to attempt in this reply to follow Thornton's fascinating journey through the contemporary philosophical landscape ranging from Millikan and Fodor to Wittgenstein and McDowell. Instead, I will provide just a few thoughts on some central issues raised by his observations.

Thornton points out some obvious parallels in the understanding of "function" between the work of Ruth Millikan on intentionality and my own on disorder. Because Millikan brilliantly uses biological function in an attempt to analyze mental content, and I use biological function to analyze medical disorder, and we both understand

biological functions broadly speaking in terms of natural selection, there is certainly a degree of affinity between our analyses. He then suggests that Wittgensteinian (or perhaps Kripkensteinian) doubts about content determinacy—due to the possibility of many different ways of generalizing a rule into the future from past behaviors—might also apply to disorder attributions and thus pose a challenge to function determinacy. I agree that this is an interesting issue, although I am doubtful that it leads to any deep problems specifically for the HDA for reasons I will explain.

First, I guess it is worth stating for the record that while I agree with Millikan that the intuitive notion of "natural/biological function" has a natural-selective evolutionary essence and can be elaborated in evolutionary terms, I do not believe that Millikan's program to leverage that understanding of "function" into a naturalist account of intentionality can possibly work for reasons related to Thornton's discussion. Meaning and content cannot be reduced to natural-selective history due to familiar indeterminacy considerations, namely, intentionality is more fine-grained and determinate than evolutionary history. In this regard, it seems to me that, although Thornton dismisses mere Quinian "gavagai" arguments as not relevant in his analysis (he is focusing on Wittgensteinian rule-paradoxical real divergences between rules and their "bent" analogs, which unlike the Quinian cases do actually diverge at some point and are not always and necessarily co-referential), in fact the gavagai-type examples seem to me to be conclusive counterexamples to Millikan's attempt to theorize meaning/content as naturally selected functions. The problem (analogous to the problem Quine pointed to in trying to use behavior or neurobiology to identify meaning) is that determinacy of meaning cannot be matched by determinacy of function. Simply and Quinianly put, there is a difference in meaning between "the heart's function is to pump blood cells throughout the body" and "the heart's function is to pump undetached parts of blood cells throughout the body," but there is no way that any theory in which functions are determined by natural selection can account for that difference because functions selected to accomplish necessarily co-occurring effects cannot be distinguished by natural selection history. Or, to take a favorite example of Millikan's, when the bee does its communicational dance to the hive's denizens after finding honey, there is no natural selection story that is going to allow a discrimination between the meanings "honey-yielding flowers at location X" and "undetached parts of honey-yielding flowers at location X." In other words, when push comes to shove, Millikan's view of meaning falls to the same old Quinian argument against content reduction that felled behaviorist and neurobiological reductions of content. So, to me, one does not require the additional machinery of Wittgenstein's critique of meaning to refute Millikan's project. Moreover, while Thornton emphasizes certain differences between Fodor and Millikan, Fodor's various theories of content similarly collapse before the elegant Quinian indeterminacy critique of content turned around by Searle to be an antireductionist argument, and Fodor's explicit attempts to address the problem don't work (Wakefield

2003). And, in keeping with this failure, Millikan and Fodor both admit that they don't have anything to say about consciousness (Wakefield 2018).

Obviously, I don't conclude from all this that therefore there is no such thing as content. Rather, I conclude, with Searle, that Quine's indeterminacy argument reduces to absurdity the notion that third-person accounts like Millikan's and Fodor's encompass all the evidence there is for content. The problem, of course, is that consciousness has been left out, but that is a long story that cannot be further addressed here.

Thornton's subtle analysis leads him to confront a problem facing not just the HDA but the concepts of function, dysfunction, and disorder within both evolutionary theory and medicine, namely, the problem of indeterminacy of function. Given an actual evolutionary history, there will be many ways of translating that into a function statement depending on how broadly or narrowly one interprets the selective "rule" being followed.

Now, although various philosophical problems have complex interrelationships, I don't think one can solve all these large general problems at once when addressing a special area. I can't solve "grue" and "underdetermination" and "Kripkenstein" when elaborating the concept of "disorder." Rather, I work within some standard assumptive systems. I grant that if these broader issues fall a certain way and it turns out that there are no filters to neutralize them down the line when doing concrete things like medical diagnosis, then there might be problems for psychiatric disorder judgments and medical diagnosis more generally, but this seems very unlikely to me. Evolutionary theory is an enormously successful scientific explanatory system, and whatever issues of indeterminacy arise for its notion of function presumably cannot undo that success, although exactly how that works needs to be understood. The HDA's evolutionary interpretations of "function" and of the intuitive notion of "biological design" take place within that framework, and thus whatever indeterminacy occurs in evolutionary theory is likely to occur within medical theory as well, but also medical theory can rely on whatever disambiguation techniques are found in the parent theories. This stance is similar to a point I recall being made by Christopher Boorse in response to those who argue that all scientific concepts are ultimately value laden so that "disorder" cannot be purely factual. He explained that if such universal value-ladenness turns out to be the case, then that would be a general limit on his claims and that he is only claiming in his "naturalist" theory that the concept of "disorder" has no *special* value loading over and above whatever value loading, if any, turns out to be routinely true of, say, physics and biology generally. Similarly, I am claiming that *given evolutionary theory's determination of functions* (however potential indeterminacies are dealt with at that level), disorder can be defined with adequate determinacy from there. If the theory of medical pathology piggybacks on the evolutionary theory of function, then it can rely on indeterminacy disambiguation within that home scientific discipline, and the claim is only that medical diagnosis does not introduce a new and seriously problematic level of indeterminacy. I see nothing in Thornton's paper that suggests that

disorder judgments do add indeterminacy problems over and above issues involved with "function" at the evolutionary theory level, but of course this impression could be proven wrong.

There are places to look for such indeterminacies. I would say with issues of trade-offs, side effects, pleiotropy, reactions to changing environments, balancing selection, and especially the precise scope of function versus dysfunction in terms of the breadth of performances that qualify as function versus dysfunction, issues of indeterminacy could arise. My inclination is to think that confident judgments of dysfunction and thus disorder would escape such problems by being failures of a clear core of function, but admittedly this intuition requires serious examination.

Thornton's exploration raises a series of profound questions that he does not attempt to answer and that I cannot undertake to answer here but that are well worthy of further analysis (I nibble at these questions in my replies to Garson and Lemoine in this volume). The questions might be put as follows: First, to what extent and in what ways is the concept of "natural function" as interpreted in evolutionary terms indeterminate? Second, how, if at all, are such indeterminacies resolved within evolutionary theorizing? Third, to what extent or in what ways, if any, do such indeterminacies yield corresponding indeterminacies in the concept of disorder that can impact meaningful real-world judgments of disorder versus nondisorder? Fourth, if our actual disorder and nondisorder judgments involve implicit resolutions of such indeterminacies, what is the logic or justification, if any, for how such indeterminacies are resolved, and in particular, does that resolution involve hidden normativist/value premises, thus building values into a deeper level of what appears on the surface to be a factual criterion? And fifth, if, as it appears, evolutionary theory is a highly successful scientific explanatory theory despite any such indeterminacies, is there any problem with the concept of disorder piggybacking on whatever function indeterminacy resolution process occurs in evolutionary theory to avoid independent function indeterminacy issues, or does it have a problem in moving from evolutionary function to dysfunction over and above whatever such issues occur within evolutionary theory itself?

References

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.

Fulford, K. W. M. 1989. *Moral Theory and Medical Practice*. Cambridge University Press.

Fulford, K. W. M. 1999. Nine variations and a coda on the theme of an evolutionary definition of dysfunction. *Journal of Abnormal Psychology* 108: 412–420.

Kendell, R. E. 2002. Five criteria for an improved taxonomy of mental disorders. In *Defining Psychopathology in the 21st Century,* J. E. Helzer and J. J. Hudziak (eds.), 3–18. American Psychiatric Publishing.

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. 2003. Fodor on inscrutability. *Mind and Language* 18: 524–537.

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. 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. 2018. *Freud and Philosophy of Mind: Vol. 1. Reconstructing the Argument for Unconscious Mental States*. Palgrave Macmillan.

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.