This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

11 Is the Dysfunction Component of the "Harmful Dysfunction Analysis" Stipulative?

Maël Lemoine

Introduction

The harmful dysfunction analysis (HDA) goes like this: for a condition to be a disorder, it is necessary and jointly sufficient to be harmful, according to a value judgment, and to be dysfunctional, according to a value-free appraisal. This is a simple, original, and powerful way to combine the "naturalist" and "normative" aspects of the concept of disease, and it is indeed a highly faithful account of what both laymen and psychiatrists mean by "disorder." However, in their "Conceptual Analysis versus Scientific Understanding," Murphy and Woolfolk have contested this point: "Wakefield's final position…demonstrates that his overall project represents a counterproductive attempt to stipulate conceptually the character and domain of scientific inquiry into psychopathology" (Murphy and Woolfolk 2001, 271). What he stipulates is, according to them, some sort of "folk psychology" consisting of our commonsense intuitions about mental disorders. I disagree. I think that there is indeed something stipulative in Wakefield's position about the concept of mental disorder but also something descriptive. Yet what is stipulative is not the general framework for the concept of mental disorder (what I will call the HAA: harmful abnormality analysis): it is rather the evolutionary concept of dysfunction and, most of all, the way in which Wakefield has tried to conflate two good ideas into one. As Bolton puts it, "Definition of mental disorder in evolutionary terms, whatever other virtues it may have, does not capture the usage of the term mental disorder in the diagnostic manuals" (Bolton 2008, xxv). I would like to add that it also fails to capture the lay public's usage of this term. Ultimately, I think that 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. The key problem here is in "stipulation": while a descriptive account must not stipulate, a scientific account has to.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

I. What Kind of Definition Does the HDA Aim at?

The HDA provides a definition of disorder. What kind of definition is it? Hempel's classic presentation of definitions contrasts the nominal, *stipulative* definition of an expression or concept, which cannot be true or false but has to be syntactically determined, univocal, and consistent (Hempel 1952, 12–14, 17–18, 18–19), with real definitions, which consist of true or false claims (7–8), whenever they conform (or not) to the given meaning of an expression or concept (*meaning analysis*) or to given facts (*empirical analysis*: 8). *Rational construction* or *logical analysis*—that is, *explication*—draws from both of them. On one hand, explication consists of a synthesis of choices between available senses of the word in question and new elements, to the effect that the term is given a much more precise meaning that was not originally contained in natural language. On the other hand, the synthesis is guided by an empirical purpose: generally, to provide some useful predictive and explanatory features (10–12). Explication differs from stipulation in that we stipulate each time we need to introduce a new concept in science (e.g., "tachyon" or "prion") and explicate a term each time we arrange and stabilize its existing meanings (e.g., the definitions of "probability_1" and "probability_2" in Carnap [1962]).

To illustrate, let me take a simple definition of disease:

(1) *disease* = state of an organism with a lesion.

It would be a *stipulative* definition if I took every existing usage of the term as unavailable or not relevant: "I (for myself) call 'disease' the state of an organism with a lesion" (the existence of so-called diseases with no lesions, or lesions without so-called diseases, does not speak against it in any way). It would be a *meaning analysis* if I intended to capture its usual meaning (or meanings) among physicians, laymen, or both: "we use the term and the concept of disease whenever speaking of the state of an organism with a lesion" (what I compare my definition to is the common usage of the term). It would be an *empirical analysis* if my definition formulated laws of nature of the kind: "every state of disease is associated with an organic lesion" (what I have in mind are actual cases, experiments, statistics, etc.). In the end, it would be an *explication* if it was a less vague and more empirically powerful concept than the one captured in the common meaning of the term (as was the case at the beginning of the nineteenth century for the definition of disease).

The HDA explicitly rules out stipulation. Speaking of Spitzer's analysis as well as of his own, Wakefield remarks that "such analyses do not stipulate how we should use mental disorder" (Wakefield 1999d, 1011). He is also clear about rejecting empirical claims. In his view, a definition of mental disorder initially requires "conceptual validity" (Wakefield 1992b, 232; 1993, 170), that is, a correct discrimination between disordered and nondisordered conditions. Only after that can operational criteria be sought and the issue of reliability addressed (Wakefield 1992b, 233; 1993, 163; 1997a, 634).[1]

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

Wakefield's suggestion of a "conceptual analysis," though, does not explicitly distinguish between meaning analysis and explication. "In a conceptual analysis, proposed accounts of a concept are tested against relatively uncontroversial and widely shared judgments about what does and does not fall under the concept. To the degree that the analysis explains these uncontroversial judgments, it is considered confirmed, and a sufficiently confirmed analysis may then be used as a guide in thinking about more controversial cases" (Wakefield 1992b, 233; Wakefield 1999a, 376). Those "widely shared judgments" about what is disordered and what is not are supposed to be "largely shared by professional and the lay public" (Wakefield 1992a, 374). Any hypothetical definition must be assessed according to those judgments, and counterintuitive consequences are supposed to rule it out (example in Wakefield 1993, 166). This approach seems to be inclining in favor of meaning analysis. Yet "formulating theories to explain a distinction we already make" (Wakefield 1999d, 1011) involves not only an analysis of the distinction but explication as well.

In my view, the distinction between meaning analysis and explication implicitly follows from the distinction between the three steps of the HDA (Wakefield 1999a, 375):

1. *disorder* means dysfunction;
2. *function* and *dysfunction* are "straightforward scientific, causal terms"; and
3. the only available scientific account of such a dysfunction is evolutionary.

Only the first two steps claim to be analytically determined from "a widely shared, intuitive medical and lay concept." The last one alone is supposed to be explicative and is expected to one day provide a means to "distinguish natural functions from other effects in a manner more precise than that afforded by commonsense intuitions."

In any case, the dysfunction clause in the definition of disorder is on no account supposed to be stipulative, whether it is analytic or explicative. Therefore, my questions are as follows:

1. Is it *description* or stipulation to define mental disorders as dysfunctional states and to define dysfunction as a scientific concept?
2. Is it *explication* or stipulation to define dysfunction (in mental disorders) as an evolutionary concept?

II. Is the Definition of "Mental Disorder" as Dysfunction Descriptive or Stipulative?

After a quick glance at the problem, it seems that defining a condition as a mental disorder may represent an instance of stipulation when considering the requirement that the condition must be judged harmful according to values; in contrast, the requirement of a genuine, value-free type of dysfunction prevents the definition of such a condition from being stipulative. Therefore, the harmful dysfunction (HD) definition of mental disorder would avoid stipulation thanks to the concept of dysfunction. However, this

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

statement requires the following qualifications: (1) a value-free concept of dysfunction does not necessarily provide a value-free definition of disorder, (2) some concepts of disorder are stipulative precisely because there are uncertainties about the existence of the dysfunction that supposedly defines them, and (3) defining mental disorders as dysfunctional states is not trivial and requires taking sides in theoretical controversies. My conclusion will be that the HDA does not avoid stipulation by introducing the concept of dysfunction in the definition of mental disorders.

2.1 A Stipulation Is a "Value-Free" Definition, Not the Definition of a "Value-Free" Concept

Stipulating often implies a commitment to some theory or, at least, some specified views on a subject. Here, we may neglect the possible, but improbable, cases of a stipulation made randomly and/or a stipulation made erroneously. For instance:

(2) "*mental disorder* = some kind of small, brown banana from the Caribbean"

is likely a random stipulation, and

(3) "*mental disorder* = the mental image of an untidy room"

is likely an erroneous stipulation. Putting aside such cases, stipulation always comes from some particular theoretical background. Therefore, a good strategy for providing a descriptive account of the meaning of "mental disorder" would be to provide a "value-free" definition. This is, I think, the purpose of the HDA.

First, we must consider that a value-free definition of a concept is not the same as a definition of a value-free concept. On one hand, we can stipulate (for normative motives) that some concepts are not normative. A definition could be biased in that way if, for instance, we wanted to say that

(4) "*depression* = a flaw in chemistry, not in character"

to avoid being judgmental about depressed people. On the other hand, there may be nonstipulative definitions of normative concepts. For instance, one could claim that

(5) "*dysfunction* = whatever prevents one from doing what is expected by a majority of people in a given culture"

is an analytic definition of a normative concept. Fulford, Cooper, and others have tried to provide such analytic definitions of disease, mental disorder, or dysfunction as normative concepts (see, e.g., Fulford 2001; Cooper 2002; Nordenfelt 2007). Although it might be contested that they are indeed value concepts, there is no contradiction between merely describing the usage of a given phrase and assuming it is value laden.

The HD definition of mental disorder may well be in the same situation as that in (4). Whether it is or not, it is clear that one does not avoid stipulation by defining mental disorder through a value-free concept of dysfunction.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

2.2 The Value-Ladenness of Some Concepts of Mental Disorder Also Comes from the Dysfunction Component of Mental Disorder

A value-free concept of dysfunction appears to permeate through to the concept of disorder itself. In other words, we cannot consider whatever state we like to be a mental disorder. By using a dysfunctional condition for the definition of mental disorder, *stipulative* definitions of various disorders would be dismissed. However, this inference requires further scrutiny.

First, I do not think that the reference to dysfunction in the definition of a given disorder implies stipulation in the sense of a commitment in favor of a given theory. Of course, given our present imperfect state of knowledge, many hypotheses are in competition, so the question is: how can the dysfunction component avoid stipulating any of them? According to Wakefield, the HD approach is "theory-neutral" but "inferential" (Wakefield 1993, 171). "Theory-neutral" means that "criteria … are framed in terms that are independent of any particular theory of the nature and genesis of mental disorder, such as psychoanalytic, cognitive, behavioral and biological theories" (Wakefield 1992a, 385; 1992b, 232). For instance,

(6) "*depression* = sadistic drives turned toward the self"

is not a theory-neutral definition of depression. Theory-neutral means that observational terms are not laden with any particular, *well-known* theory, not that they are not laden with any theory at all (Wakefield 1999c, 966: "avoid definitionally ruling out any of the major competing theories of etiology"). But the definition of a mental disorder cannot be atheoretical in the sense of being "non-inferential" in the sense that "criteria … are framed in terms that are entirely observational and do not depend for their application on inferences about internal, historical, evolutionary, or other unobservable features" (Wakefield 1993, 171). Definition (6) fails to be theory-neutral but is inferential, because it refers to so-called theoretical entities, that is, nonobserved processes (the "drives"). The concept of dysfunction the HDA refers to is inferential, because it has to refer to an unobserved dysfunction that explains what is observed: "diagnosing disorder is inherently theoretical in the sense that it goes beyond sheer symptoms to hypothesize the existence of a dysfunction, without necessarily specifying the exact nature of the dysfunction or its etiology" (Wakefield 1999c, 966). How can it be at the same time theory-neutral and inferential? Is not an inference to etiology theoretical per se? Indeed, it is; but Wakefield propounds the "black-box" (Wakefield 1997b, 658; 1999b, 471–472; 2001, 359–362) view of dysfunction, according to which it is permissible to speak of a dysfunction before knowing what it consists of exactly. A definition can thus be inferential, because it assumes the existence of a dysfunction, and theory-neutral, because it does not "explain the behavior in any substantial or full way" (Wakefield 1999c, 986). Thanks to this black-box view, the concept of dysfunction can avoid theoretical stipulation.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

Second, the concept of dysfunctional states is expected to stand against normative views of what is desirable and what is not. This means that it can oppose normative aspects of the choices made by authors of the diagnostic manuals. It is important not to conflate this kind of impartiality with reliability. There could be "reliable"—that is, shared and highly reproducible—clinical judgments based on precise, operational, and yet normative criteria (Wakefield 1992b, 233). For instance,

(7) "*antipsychiatric behavior* = denial, either in acts or in words, of the facts of psychiatric science, as assessed by the results of the tests of the Scientific Antipsychiatric Scale."

Third, the necessity to assess dysfunction in a suspected case of mental disorder also addresses the clinical level. It limits the operative role of the values of the clinician in clinical judgments. For instance,

(8) "*antipsychiatric behavior* = excessive denial of the facts of psychiatric science according to the feeling of the clinician"

and

(9) "*mental distress* = whatever state a subject may be in, which requires help from a mental practitioner (according to the practitioner)"

can be applied to a given situation without referring to anything other than the clinician's value judgment. Definition (9) may be reliable depending on the homogeneity of the clinical community (see Bolton 2008, 14–15), but it is value-laden nevertheless.

In helping to guard against both arbitrary categories and false-positive cases (Wakefield et al. 2010), the dysfunction component presents a further problem in the determination of what should properly be called a mental disorder. With Horwitz, Wakefield has emphasized the importance of the assessment of context in *The Loss of Sadness* (Horwitz and Wakefield 2007). The problem here is linked to the dysfunction component, not to the harm component. The difficulty lies in appraising whether there is indeed a dysfunction involved in some state (e.g., sadness) or whether the state is best understood as a normal response to a life event. This does not suggest that the general framework of the HD definition of mental disorder is stipulation, but it suggests that the dysfunction component is in part responsible for some stipulations about both definition and application of categories. This is, so to speak, the side effect of the black-box view.

2.3 The Concept of Mental Disorder Does Not Necessarily Imply a Concept of Dysfunction

Until now, it has not been proven that the HD definition of mental disorders is a stipulation but only that the presence of the dysfunction component does not immunize either the general definition of mental disorders or specific definitions of mental disorders against stipulation. I would not assert here that Wakefield stipulates a dysfunction

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

concept of mental disorders, but I would not consider defining mental disorder as dysfunction a trivial point to make either. Derek Bolton has rightly emphasized this:

> 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)

For the purposes of this chapter, the point to be made is just to emphasize the incompatibility between the dysfunction thesis and an analytic purpose. With Bolton, I think that Wakefield is correct in saying that the folk and scientific core concept of mental disorder requires a scientific component. But I do not think that it is merely *description* to assume that this scientific component is dysfunction.

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.

III. Is the Definition of "Dysfunction" as an Evolutionary Concept Explication or Stipulation?

Is the evolutionary concept of dysfunction some kind of elucidation of what we usually mean by dysfunction (as "probability_1" and "probability_2" are for the general concept of probability), or is it a brand-new scientific concept (as "prion," "tachyon," "money supply," or "energy")? Obviously, stipulation would be a legitimate approach in the field of philosophy of psychiatry as well as in science, but perhaps equally obviously, Wakefield means it to be an explication of the common term and not a stipulation of a new, specific concept of dysfunction. I would like to make a few points, though, in favor of a stipulative interpretation of the HDA: (1) there are other scientific definitions of dysfunction than the evolutionary definition contained within the HDA, and were the evolutionary approach to provide the best scientific account available, it could as well be stipulation; (2) even a theory-neutral concept of dysfunction can imply stipulation; (3) the obviousness of the concept of dysfunction does not mean

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

that its definition is explicative; and finally, (4) Boorse's naturalist account of dysfunction seems to fit an explicative purpose best.

3.1 The Best Scientific Account of Dysfunction May Be Stipulation All the Same

Some of us are inclined to consider abnormalities, as previously defined, to be statistical entities. Others view abnormalities in terms of biological defects, and yet another approach describes them as information-processing failures, social conflicts, and so on. Some of these positions equate abnormality with dysfunction, yet they do not all necessarily mean the same thing by "dysfunction" (for this line of arguments against the HDA, see, e.g., Murphy and Woolfolk 2001; Schramme 2010). Wakefield considers the question to be as follows: is a definition of abnormality as a biological dysfunction in the evolutionary sense the best one available or even the "most viable one" (Wakefield 1992b, 237)? He has made a strong case against several kinds of alternative definitions—for example, statistical definitions (Wakefield 1992a, 377–378; 1999a, 388–390, etc.), "disadvantage" definitions (Wakefield 1992a, 378), and "current causal role" definitions (Wakefield 2001), among others. He has also made a strong case for his definition of dysfunction against many different critiques. The more successful he is in defending his definition against his critics, the more his definition appears to be the best one available. This is not my point here, though. My point is more to question the implicit argument that states a scientific (i.e., evolutionary-based) definition of dysfunction would be the best one available, therefore making it impartial and nonstipulative.

As a matter of fact, impartiality has nothing to do with explication (as opposed to stipulation). Wakefield's account of abnormality as a dysfunction could well be the best one available, yet fails to elucidate the common meaning of "mental disorder." At the end of the nineteenth century, for instance,

(10) "*Diabetes mellitus* = pancreatic condition"

was indeed impartial in that it was the best definition available, based on the discoveries of Oscar Minkowski. Yet in no way would this have been an explication, for no one had yet thought of this as the definition of diabetes mellitus. What if, for instance, the evolutionary account of dysfunction was the best one on scientific grounds, yet was definitely not what scientists themselves usually meant by dysfunction? On the other hand, Wakefield's definition of dysfunction could also draw from no arbitrarily determined goals, yet be stipulation, precisely because a lot of effects we want to call functions might be value-laden (aside from any harm done by the condition). For instance,

(11) "*old age* = universal dysfunction of the telomerase in animals"

might happen to be impartial, yet we could oppose the claim that aging is the result of a dysfunction (see Wakefield 1999b, 468). Hence, an impartial definition of dysfunction is not a sufficient condition for an explicative definition of dysfunction.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

3.2 A Theory-Neutral Definition of Dysfunction Is Not Necessarily Explication

The fact that all the major theories of mental disorder (see, e.g., Bolton 2008, 15 sqq.) are *compatible* with a definition of disorder as a dysfunction in the evolutionary sense does not imply that disorder analytically means dysfunction in the evolutionary sense. For instance,

(12) "*light* = visible manifestation of the divine radiance known as uncreated light or Tabor's Light"

does not, in my view, rule out any claim from the corpuscular or wave theory of light and thus does not take sides. It is undoubtedly stipulative, though. Therefore, the theory-neutrality of the HDA does not make it immune to stipulation (not of course because it would be theological or nonscientific but rather because it is a new hypothesis and works on a different level of analysis with regard to the question of mental disorders).

3.3 A Self-Explicatory Concept of Dysfunction Is Not Necessarily Explicative

According to Wakefield, the evolutionary concept of dysfunction is not only the best available theory-neutral concept we possess but also the clearest: "natural selection is the only known means by which an effect can explain a naturally occurring mechanism that provides it" (Wakefield 1992a, 383; 1999b, 465). In other words, the evolutionary concept of dysfunction is explicative rather than stipulative in that it replaces an obscure and vague concept of dysfunction by an obvious or self-explicatory one.

Elsewhere, I have given a cursory account of what I mean by "self-explicatory" in biomedical science more generally (see Lemoine 2011). Let us assume here simply that scientific explanations can never be satisfactory except in a special kind of predicament where the scientist is bound to acknowledge that this particular explication takes everything important into account. In a scientific explanation, necessity comes free with some beliefs that make this predicament possible. For instance,

(13) "*Recessive genetic disease* = any disease that must affect 25% of the offspring of heterozygous parents as to this trait"

is an obvious definition, given the laws of Mendelian genetics. (In other senses, of course, it is not obvious at all, because it is not operational. First, no "must affect" can be observed; second, "25%" is probabilistic here.) On the other hand,

(14) "*mental disorder* = state of a person endowed with statistically rare (less than 2.5%) representations of some significant item of her environment"

and

(15) "*dysfunctional* = belonging to the first 2.5 percentiles of a Gaussian distribution of mean performances to a task"

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

are not obvious in the sense of self-explanatory, because they rest upon an arbitrary, obscurely determined threshold of normality or functionality.

I think that the evolutionary concept of dysfunction is indeed obvious and self-explanatory. But the fact that a concept is obvious within one particular predicament, say, the theory of evolution, does not prove that it is the only predicament possible in which an obvious concept of dysfunction is possible. Therefore, even if we could think of no other self-explanatory account of dysfunction, one cannot assume that this is the only one possible. Yet this would be a necessary condition for us to assume that this definition is explication.

3.4 Boorse's Account of Dysfunction Is Better Than Wakefield's as a Naturalist Explication of the Concept of Dysfunction

The challenge for a naturalist account of mental disorder is to elucidate it by shielding entirely the dysfunction component from values "partitioned off in the harmful part of the analysis" (Bolton 2008, 121, 231–233 on Gert and Culver's appraisal of the HDA). By assuming that *harm* and *dysfunction* are to be thought of as entirely separate things, the HDA stays at equal distance from two opposed lines of thought. One is the idea that dysfunctions are precisely the kind of mechanisms associated with the states we disvalue or those that an organism seems to disvalue itself. Valued or disvalued states do not change, but as we discover biological mechanisms and their links to valued and disvalued states, we change our views on their being functional or dysfunctional. In the end, we are expected to know what is functional and dysfunctional because we want to know what is linked to harmful states and how. The other line of thought is that what we conceive of as "harmful" is what we understand as dysfunctional in at least one respect. On this view, it would be precisely because we do not know every function and dysfunction, that the concept of harm is value-laden, and that the definition of mental disorder is too. Once function and dysfunction were properly understood, the concept of harm would be as value-free as the concept of function, and so would be the concept of mental disorder.

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. But 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. In the case of *mental* conditions, "this ignorance is part of the reason for the high degree of confusion and controversy concerning which conditions are really mental disorders" (Wakefield 1992a, 383). Confusion and controversy do not imply that dysfunction is a value-laden concept, given that there are scientific controversies (at least let us suppose that). However, it is clear that if we do not know of everything in a given mechanism (e.g., walking), everybody will agree with the idea that we can say it is malfunctioning all the same. How do we do that, if we must not use the concept of harm? An explicative, naturalist concept of

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

dysfunction has to explain this fact. Wakefield is correct, I think, in saying that "evolutionary design and human values do also often diverge" (Wakefield 2001, 352) and even in saying that "natural function is a scientific concept that cannot be reduced to values" (Wakefield 1992a, 376). But is it true that an evolutionary or "etiologic" account of *dysfunction* elucidates this independence of the concept of dysfunction best?

The general theoretical framework of Wakefield's analysis of dysfunction is "Hempel's challenge" on function (Wakefield 2001, 359): How can we discriminate among the natural effects of a biological mechanism so as to determine which one(s) is (are) its function(s) (Wakefield 1992a, 382)? A function, as Wakefield puts it, conforms to the "explanation criterion" (385). It is this special kind of effect that can explain the presence of its cause (Wakefield sometimes also talks of functions as the cause of natural effects). What is *dysfunction,* then? According to Wakefield, "dysfunction is the failure of a mechanism to perform its natural function" (383; Wakefield 1999b, 465: "failure of biologically designed functions"). Here, dysfunction is obviously an underlying cause, not its effects. We understand underlying dysfunction due to the absence of natural effects (what Wakefield calls "circumstantial evidence": Wakefield 1999b, 465; 1999c, 988).

In a strictly naturalist view, I can think of two ways to know that functional effects are absent. One is by comparing the effects of a mechanism in an individual to the effects of the same mechanism in the proper subgroup of the relevant species. This is a statistical definition of function, which we owe to Boorse (1977) and not to Wakefield. The other approach—and I presume it to be Wakefield's—is to *appraise* those observed effects that replace the natural ones. Some effects obviously run contrary to what could possibly be the natural function of a biological organism. This does not mean that every effect that cannot possibly explain the existence of a mechanism that is causing it is a dysfunction. For instance, the sound of the heart cannot explain the cardiac mechanism, but it is not a dysfunction. Rather, it means that a dysfunction is a failure of a mechanism to perform its natural function because this effect is somehow at odds with the presumed natural function of the mechanism.

I think that the only naturalist way to distinguish between nonnatural and harmful effects of a mechanism in an imperfect state of knowledge is to adopt Boorse's biostatistical views on dysfunction. Significantly, when considering the possibility of harmless dysfunctions, Wakefield appears to turn to this competing conception of dysfunction. For instance, he gives the imaginary example of "a dysfunction that slow[s] the aging process and lengthen[s] life" (Wakefield 1992a, 384; 2001, 352), but the question is how could this be imagined as a dysfunction except by comparison? The same goes for albinism, fused toes, *dextrocardia situs inversus*, and having one kidney. This is not a proof, but it is a sign. I am not saying that Boorse's views on function are compatible with Wakefield's examples on harmless dysfunction. I am suggesting that Wakefield's examples of harmless dysfunction are only compatible with Boorse's views on function.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

This leaves us with a challenge to the HDA. If it cannot give any example of a harmless condition, which would be thought of as dysfunctional without any comparison to a biological reference class, it has either to consider every dysfunctional state to be harmful in one way or to adopt Boorse's view on dysfunction, at least as a temporary stage, in providing a naturalist account of dysfunction. The latter is the least damaging to the HDA. But it means that the HDA works best to clarify what we currently mean by "dysfunction," therefore, that any other view on it is stipulative.

Conclusion

My first point was that the HD definition of mental disorder may rightly introduce the concept of dysfunction as the best way to define it, but this is not really a description of the common usage of the term. My second point was that in defining dysfunction as an evolutionary concept, Wakefield propounds a sound scientific hypothesis, but one that works poorly in elucidating the common meaning of the term. My conclusion is that there is a fundamental choice to make between a descriptive or explicative "conceptual analysis" and a stipulative, scientific contribution.

Because the HAA I suggested here as the most faithful description of the meaning of the term is probably not a very useful definition, I suggest that the HDA should be resolute in proclaiming stipulation instead and entering the scientific arena of competition between theories (compare with Wakefield 1999c, 965). In doing so, maybe Wakefield should abandon any belief that there exists an independent, precise content of the notion of mental disorder as belonging to folk psychology, as Murphy and Woolfolk have suggested. Or maybe he would have to acknowledge that there is not much to gain in a descriptive definition of mental disorder. In view of this, it would be possible to adopt a fully fledged naturalist conception of mental disorder as well as of dysfunction: here, I mean "naturalism" as the rejection of *any* legitimate independence of the folk concept of disease from science. I think that what prevents Wakefield from doing so is the strong strategic argument from plausibility: for instance, "Freud was sophisticated enough to realize that, to offer a persuasive theory of etiology, one must define a disorder in such a way that it can be identified by those who do not initially share one's theory" (1999c, 968). Besides, how could we consider mental disorder independently of commonsense views of it being harmful, if this is not possible even for a concept such as "dysfunction"? (This is a point one could reject in favor of Wakefield's position and against Murphy and Woolfolk.) In defense of naturalism, though, I think that this strong strategic constraint of folk plausibility on our conceptual definitions of mental disorder can be understood *extensively*, not necessarily *intensively*. I mean that as long as roughly the same patients are considered to be affected by mental disorders, the intensional content of the definition does not matter to the lay public, and the requirement of "conceptual validity" is respected. Besides, current boundaries are both fuzzy and

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

plastic, and they do not constitute such a stringent constraint on theoretical plausibility They are plastic because, thanks to science, the lay public has been convinced to view some conditions as disorders. This comes precisely from the fact that any mental disorder is stipulated; "because the actual identity of the essence is often unknown at the time that the term is defined, the definition uses a stipulated base set of known initial instances of the natural kind to establish the reference of the term" (Wakefield 1997b, 657). Moreover, "treatable conditions" in Bolton's sense (Bolton 2000, 149 sqq.; 2008, 191) are a broader class within which scientists might feel rather free to delineate classes without directly intervening in social conflicts about mental health. And this is the class philosophers actually have to address if they want to step in.

Note

1. "In making criteria more reliable, *DSM-III-R* has sacrificed some aspects of validity" (Wakefield 1992b, 241). See also Wakefield (1993, 161): "I suggest an alternative approach to diagnosis in which operational criteria for specific disorders are based on nonoperational functional definitions of mental disorder and of specific disorders." At last, Wakefield says, "A reference to dysfunction that is not translated into operationalized criteria leaves it entirely open to clinicians to make a global judgment of whether a dysfunction exists. This introduces a highly unreliable element into the criteria" (Wakefield, 1997a, 646). This last quotation addresses the "dysfunction" component of the harmful dysfunction analysis. This is the one I will focus on, leaving aside interesting questions about the stipulative or descriptive nature of the "harm" component.

References

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

Bolton, D. 2008. *What Is Mental Disorder? An Essay in Philosophy, Science, and Values*. *International Perspectives in Philosophy and Psychiatry*. Oxford University Press.

Boorse, C. 1977. Health as a theoretical concept. *Philosophy of Science* 44(4): 542–573.

Carnap, R. 1962. *Logical Foundations of Probability*. 2nd ed. University of Chicago Press.

Cooper, R. 2002. *Disease*. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33(2): 263–282.

Fulford, K. W. 2001. 'What is (mental) disease?': An open letter to Christopher Boorse. *Journal of Medical Ethics* 27(2): 80–85.

Hempel, C. G. 1952. *Fundamentals of Concept Formation in Empirical Science*. University of Chicago Press.

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

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a portion of the eBook at doi:10.7551/mitpress/9949.001.0001

Lemoine, M. 2011. *La désunité de la médecine. Essai sur les valeurs explicatives de la science*. Hermann.

Murphy, D., and R. L. Woolfolk. 2001. Conceptual analysis versus scientific understanding: An assessment of Wakefield's folk psychiatry. *Philosophy, Psychiatry, and Psychology* 7(4): 271–293.

Nordenfelt, L. 2007. *Rationality and Compulsion: Applying Action Theory to Psychiatry*. *International perspectives in philosophy and psychiatry*. Oxford University Press.

Schramme, T. 2010. Can we define mental disorder by using the criterion of mental dysfunction? *Theoretical Medicine and Bioethics* 31(1): 35–47.

Wakefield, J. C. 1992a. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47(3): 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(2): 232–247.

Wakefield, J. C. 1993. Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. *Journal of Abnormal Psychology* 102(1): 160–172.

Wakefield, J. C. 1997a. Diagnosing *DSM-IV*, Part I: *DSM-IV* and the concept of disorder. *Behaviour Research and Therapy* 35(7): 633–649.

Wakefield, J. C. 1997b. Diagnosing *DSM-IV*, Part 2: Eysenck (1986) and the essentialist fallacy. *Behaviour Research and Therapy* 35(7): 651–665.

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

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

Wakefield, J. C. 1999c. Philosophy of science and the progressiveness of the *DSM*'s theory-neutral nosology: Response to Follette and Houts, Part 1. *Behaviour Research and Therapy* 37(10): 963–999.

Wakefield, J. C. 1999d. The concept of disorder as a foundation for the *DSM*'s theory-neutral nosology: Response to Follette and Houts, Part 2. *Behaviour Research and Therapy* 37(10): 1001–1027.

Wakefield, J. C. 2001. Evolutionary history versus current causal role in the definition of disorder: Reply to McNally. *Behaviour Research and Therapy* 39(3): 347–366.

Wakefield, J. C., M. F. Schmitz, and J. C. Baer. 2010. Does the *DSM-IV* clinical significance criterion for major depression reduce false positives? Evidence from the National Comorbidity Survey replication. *American Journal of Psychiatry* 167(3): 298–304.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

This is a section of doi:10.7551/mitpress/9949.001.0001

Defining Mental Disorder

Jerome Wakefield and His Critics

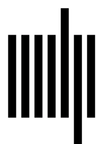
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl

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

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024

MIT Press Direct

© 2021 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC-BY-NC-ND license.

Subject to such license, all rights are reserved.

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

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268295/9780262362931_c001000.pdf by guest on 08 October 2024