PhiMiSci
Philosophy and the Mind Sciences

Mental disorder: An ability-based view

Sanja Dembić[a] (sanja.dembic@hu-berlin.de)

Abstract
What is it to have a mental disorder? The paper proposes an ability-based view of mental disorder. It argues that such a view is preferable to biological dysfunction views such as Wakefield's Harmful Dysfunction Analysis and Boorse's Biostatistical Theory. According to the proposed view, having a mental disorder is basically a matter of having a certain type of inability (or: an ability that is not sufficiently high): the inability to respond adequately to some of one's available reasons in some of one's reasons-sensitive attitudes or actions, where the threshold of inability is determined by one's being harmed. The relevant concepts of inability, reasons, and harm are sketched. The paper argues that the proposed view evades some problems of biological dysfunction views by remaining neutral on questions of causation and the evolution of the mind. Furthermore, it can capture better what is distinctively "mental" about mental disorder. On the proposed view, it is the rational relations among an individual's attitudes and actions that are "disordered" and the relevant norms in mental disorder are the norms of reasons. As further merits, the view can account for degrees of disorder, incorporate biological as well as social aspects, and elucidate the relations among disorders, symptoms, and their causes.

Keywords
Ability · Biological dysfunction · Defining mental disorder · Reasons · Mental disorder

This article is part of a special issue on "Models and mechanisms in philosophy of psychiatry," edited by Lena Kästner and Henrik Walter.

Introduction

What is it to have a mental disorder? In this paper, I propose an ability-based view of mental disorder and argue that such a view does not have certain problems that common biological ones do.[1] The goal of this paper is not to give a full-fledged new account, but to sketch an alternative that is worthwhile of further exploration. I proceed as follows: First, I discuss two major biological views of mental disorder.

[a]Humboldt-Universität zu Berlin.

[1]I use the term "biological" views of mental disorder to refer to those views which analyze the concept of mental disorder mainly in terms of biological dysfunctions.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630
 https://philosophymindscience.org ISSN: 2699-0369

Second, I delineate my own ability-based view. Third, I show how the view I propose evades certain problems of the two major biological views and point out some of its further merits.[2]

Common biological views of mental disorder refer to the concept of biological dysfunction to define "mental disorder" (Boorse, 1976b, 1977, 1997, 2014; Wakefield, 1992b, 1992a, 1999a, 1999b). Roughly, the idea is this: an individual has a mental disorder (if and) only if they are in a condition that results from a failure of some mental mechanism to perform at least one of its biological functions. The two major competing views are Wakefield's "harmful dysfunction analysis" (HDA) and Boorse's "biostatistical theory" (BST). They differ (1) with respect to their definition of "biological function" (Boorse, 1976a, 2002; Wakefield, 1992a, 1999a) and (2) with respect to whether they consider the presence of a biological dysfunction not only necessary, but also sufficient for having a mental disorder (in a theoretical sense of the term). Wakefield (1992b) argues that the presence of harm is necessary too. Boorse (1997) denies this for the presence of a mental disorder in a theoretical sense, but recognizes it for various practical senses of the term. So, the main difference between the HDA and the BST lies in (1) their definitions of "biological function". Boorse offers a goal-directed system definition of "biological function", whereas Wakefield offers an evolutionary one.

Biological dysfunction views of mental disorder face various problems. It has been argued that the presence of a biological dysfunction (in either sense) is neither necessary nor sufficient for having a mental disorder. In this paper, I refute some replies to these objections and add two further arguments against such views. First, I argue that the concept of biological dysfunction does not track the sense of "disorder" ascribers of mental disorder primarily care about when ascribing a mental disorder, because it doesn't account for what is specifically "mental" in "mental disorder". Second, I argue that a definition of "mental disorder" which remains neutral on the evolution of the mind and the causes of mental disorder is preferable to ones that rely on certain assumptions on either of those issues as the presented biological dysfunction views do.

On the view I propose, having a mental disorder is a matter of having a certain modal property, namely an *inability*. Roughly, the idea is this: an individual has a mental disorder if and only if they are *unable* (in the relevant sense) to respond adequately to some of their available *reasons* in some of their thinking, feeling, or acting, where the threshold of inability is determined by the individual's being *harmed*. I call this the "Rehability View" (RHA) which is a portmanteau of "REason, Harm, and ABILITY". I argue that the RHA tracks the sense of "disorder" ascribers of mental disorder primarily care about when ascribing a mental disorder. It also remains neutral on questions of causation and the evolution of the mind. I do not discuss the harm condition, because I am interested in a practically relevant

[2] I present my view in contrast to the two major biological ones (and not in a stand-alone way) to avoid a quite general objection: why do we need a *further* view when we already have at least two very elaborated views of mental disorder?

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

concept of mental disorder. In this respect, both the HDA and the BST recognize the relevance of harm.

A note on method: my aim is to offer an "explication" (or "ameliorative analysis") of the technical concept associated with the term "mental disorder" and *not* a conceptual analysis.[3] Roughly, the difference is this: Conceptual analysis is the (descriptive) project of making explicit the (ordinary or technical) concept associated with a certain term, that is *to capture all of its actual (ordinary or technical) uses*. An explication is the (potentially revisionary) project of "fashioning" a concept "that will do a certain job" (Millikan, 1984, p. 2). In other words, the goal of an explication is to formulate a concept that is fruitful to us for certain *purposes*. In order to achieve this goal it need not necessarily capture *all* of its actual (ordinary or technical) uses.

Why explication? Because mental disorder is a contested subject – despite widespread agreement about many phenomena, people in the scientific community disagree about whether certain phenomena do or should fall under the concept of mental disorder – and conceptual analysis cannot give us orientation in cases of disagreement. This is because it has to reflect all of our actual uses to be extensionally adequate.[4] Any conceptual analysis that does not capture all of our actual uses will be at least slightly revisionary, to wit in an arbitrary way.[5] By contrast, an explication is potentially revisionary in a way that is not arbitrary, because it is goal-oriented. Nevertheless, even an explication of "mental disorder" aims to capture as many phenomena which are deemed mental disorders within the scientific community as possible (otherwise it wouldn't be an explication of "our" technical concept associated with the term "mental disorder").

What are the purposes for which we, as members of a society, need the concept of mental disorder? Let me mention two: (a) for the purpose of scientific classification and (b) to help us settle practical or normative questions such as who should be eligible for psychiatric or psychotherapeutic treatment and who should be excused from moral or legal responsibility. In light of these purposes, I propose the following set of adequacy conditions which any fruitful definition of "mental disorder" will have to meet: to capture the distinction between (1) mental health and disorder, (2) mental and bodily disorder, (3) mental disorder and deviance from social, legal, or moral norms, as well as to clarify (4) what about having a mental

[3] I discuss Boorse's and Wakefield's methods in section 1.

[4] Someone might object that a conceptual analysis doesn't have to capture *all* of our actual uses, but only the *correct* ones. But at least with technical terms this simply begs the question, because in cases of disagreement, people precisely disagree about which use of the technical term "mental disorder" is correct.

[5] There are many slightly imperfect conceptual analyses. Which one should we accept? Without further argument, the choice seems arbitrary. If a further argument refers to some specific purpose, the purported conceptual analysis turns out to be an explication after all.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

disorder it is that may justify certain normative consequences (such as eligibility for treatment or excuse from responsibility).[6]

In sum, I aim to provide a definition that captures not necessarily all, but as many types of phenomena which are deemed mental disorders within the scientific community as possible and that captures the aforementioned distinctions which are fruitful to us for scientific and practical purposes. I argue that the RHA can capture (2) better than the HDA and the BST.

The paper is structured as follows. In section 1 I delineate the HDA and BST and discuss some of their major problems. Since the literature on both of these theories is too broad to be discussed in a single paper, I focus only on some of the problems which, as I will argue, are discussed but still unresolved. In section 2 I present the RHA, my own positive conception of mental disorder. In section 3 I argue that the RHA evades the aforementioned problems of the HDA and BST and point out some of its further merits. Section 4 concludes.

1 Biological dysfunction views

Wakefield's Harmful Dysfunction Analysis (HDA) of mental disorder which was originally presented in two papers (Wakefield, 1992b, 1992a) can be reconstructed as follows:

HDA An individual S has a mental disorder if and only if S is in a condition C which results from a failure of a mental mechanism to perform at least one of its proper biological functions and C is harmful by the standards of S's culture.

Wakefield specifies that "proper biological function" is to be understood in evolutionary terms, that is, in the sense of a naturally selected effect. Neander (1991, p. 174) offers the following explication of "proper biological function" (PBF):[7]

PBF It is the/a proper [biological] function of an item (X) of an organism (O) to do that which items of X's type did to contribute to the inclusive fitness of O's ancestors, and which caused the genotype, of which X is the phenotypic expression, to be selected by natural selection.

Some clarificatory remarks. A genotype is the set of genes an individual possesses. Genes can be expressed as traits by an individual. A phenotype is the set of traits an individual possesses. An individual's phenotype is not only determined by their

[6] Can all adequacy conditions be met by one definition? Perhaps, they cannot. But I treat the view that they cannot as a last resort. In this paper, I attempt to show that they can. It remains to be seen whether this project will be successful.

[7] An explication of "proper biological function" was first introduced by Wright (1973) and most prominently developed further by Millikan (1989) and Neander (1991). Wakefield's own characterization is imprecise, because it doesn't distinguish between types and tokens.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

genes but also by environmental influences. Natural selection operates on the phenotype – more specifically: on the variation of phenotypes. It is the phenotype that gives an individual advantages or disadvantages in the struggle for survival and reproduction. And it gives only *relative* advantages or disadvantages, namely relative to other variations in a population. A trait can be selected because of its effects only if having that effect counts as an adaptive variation in the population. However, what evolves is the genetic setup. Evolution is the change in a genotype due to the natural selection of a phenotype. So, two of the basic assumptions of natural selection are (1) that variation in a trait is possible, and (2) that a given trait can be inherited (in the genetic sense: that it can be passed on to the next generation via transmission of genes). In light of this, a trait can possess a proper biological function only if there were variations of that trait in a population and the trait is transmitted genetically. As a consequence, the concept of proper biological function is a highly "demanding" one in the sense that one needs to know a lot about evolution and genetics to justifiably ascribe a proper biological function to a certain trait. Intuitions cannot be taken at face value.

By contrast, Boorse defines "health" in terms of statistically normal physiological functioning and disease (pathological condition or disorder) as an impairment of it.[8] Boorse (2014, p. 684) formulates his Biostatistical Theory (BST) as follows:

BST

1. The *reference class* is a natural class of organisms of uniform functional design; specifically, i.e. an age group of a sex of a species.
2. A *normal function* of an internal part or process within members of a reference class is its statistically typical contribution to survival or reproduction.[9]
3. *Health* in a member of a reference class is *normal functional ability*: the readiness of each internal part to perform its normal functions on typical occasions with typical efficiency.
4. A *disease* or *pathological condition* is an internal state which impairs health, i.e., reduces one or more functional abilities below typical efficiency.

According to Boorse, the BST is an analysis of both, the concept of bodily as well as mental disorder. In mental disorders, the relevant mechanisms are mental ones.[10]

[8] Boorse uses "disease" and "pathological condition" as overarching terms for all deviances from health. I'll use "disorder" (synonymously), because it is more common in the mental health literature.

[9] Boorse identifies "internal parts or processes" as the bearers of functions. To keep it short, I'll use "mechanism" to denote the bearer of functions. Roughly, a mechanism is a system of causally interacting parts organized such that they are "responsible" for the phenomenon-to-be-explained (see Glennan, 2017; Krickel, 2018).

[10] See Boorse (1976a) for a more detailed view of how his theory applies to the realm of the mental. In short, Boorse (1976a, p. 67) defends the following view: "[A] mental disturbance gets classed

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

On Boorse's view, the bearer of disorders are organisms. The physiology of an organism is a hierarchy of goal-directed systems. Boorse (2011, p. 27) adopts a dispositional view of the goal-directedness of the relevant systems: a "physical system has the purely physical, nonintensional, property of being directed to a goal G when disposed to adjust its behavior, through some range of environmental variation, in ways needed to achieve G". Boorse contends that in physiology, the relevant highest-level goals of the organism as a whole are survival and reproduction.

Boorse's conception of "normal physiological function" (NPF) can be reconstructed as follows:

NPF A mechanism of an organism O has a *normal physiological function* F if and only if

1. it is statistically normal for the members of the class of organisms O to have a mechanism that causes or constitutes F *and*
2. F contributes causally to the individual survival or reproduction of the members of the class of organisms O.

For instance, the cardiovascular system in a human has the normal physiological function to pump blood through the organism on statistically normal occasions and with at least statistically normal efficiency, because doing so contributes to the individual survival or reproduction of humans and because this is what cardiovascular systems in humans statistically normally do.

Let me now turn to some problems of these views.

1.1 Problems of the HDA

The HDA faces many objections. I contend that the strength of some of them depends on whether the HDA is conceived of as a conceptual analysis or an explication of the concept associated with the term "mental disorder". How should we, as readers, conceive of it? Wakefield admits that he has been "sloppy" about it in some of his writings.[11] In Wakefield (2021, p. 282) he clarifies that he thinks of the HDA as a two-step approach: " 'Harmful dysfunction' is a conceptual analysis prior to the evolutionary interpretation of 'dysfunction,' and the evolutionary interpretation of 'function' is an essentialist theoretical move [...]."

as 'mental illness' when some accepted explanation of it refers not to the patient's physiology but to his feelings, beliefs, and experiences. The defining property of mental disease is mental causation."

For the purposes of this paper, it suffices to say that, according to Boorse, the mental health-relevant mechanisms are *mental* ones, regardless of how exactly those are to be understood, because nothing in my critical evaluation of the BST hinges on a specific account of mental mechanisms.

[11] For a discussion, see Faucher & Forest (2021, ch. 11 and ch. 12, 284).

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

So, on Wakefield's view, the first step is a conceptual analysis of "mental disorder" which, on his view, yields that "disorder" means "harmful dysfunction". The second step is an explication of "function" and "dysfunction" in terms of evolutionary theory. The rationale for this explication is a reference to the best explanation: evolutionary theory provides us with the best scientific theory of the nature of functions and dysfunctions. In the end, the HDA is at least partly an explication of "mental disorder".

Wakefield has good reasons to conceive of the HDA as an explication. If we conceived of the HDA as a conceptual analysis, it would be extensionally inadequate, because – as it has been argued in the literature – the presence of a biological dysfunction in terms of PBFs is not necessary for having a mental disorder. For instance, it is conceptually possible that there be (1) mental disorders of spandrels or (2) mental disorders which are themselves adaptations. Wakefield rejects these objections, but as I argue in the following, his replies are unsuccessful.

(a) Spandrels. According to Gould & Lewontin (1979), spandrels are adventitious side-effects of the development of certain functions which themselves never possessed any adaptive function. As Murphy & Woolfolk (2000, p. 243) point out, if there are spandrels of the mind (and failures thereof) that indicate mental disorders, but are not also failures of a PBF, then not all mental disorders are dysfunctions in the evolutionary sense. Wakefield (2000, p. 254) replies that nobody ever gave an *actual* example of a spandrel-inspired mental disorder and that merely pointing out the possibility of such a case does not prove anything. It merely asserts what has to be shown. However, if we conceive of the HDA as a conceptual analysis of "mental disorder", then Wakefield's reply is problematic. To show that the presence of a biological dysfunction is not *necessary* for having a mental disorder, pointing out a hypothetical class of counterexamples is enough. It suffices to show it *conceivable* that there be a mental disorder without a biological dysfunction.

(b) Adaptations. Murphy & Woolfolk (2000, p. 244) claim that some mental disorders such as depression may be adaptive mechanisms (for example, it might be fitness-enhancing to conserve energy and to elicit aid from others) and selected because of that (and thus, have a biological function). So, on their view, not all mental disorders necessarily involve biological dysfunctions. Wakefield (2000) replies that intuitions about mental disorders such as depression depend on the intensity of the condition and that this, in turn, correlates with our intuitions about whether they are adaptive and have a biological function. He claims that only moderate depressiveness may have been an adaptive response to a loss. But generally, such moderate cases are not considered dysfunctions or disorders. On his view, biological dysfunctions are attributed only to *extreme* cases where they do "not appear to be useful strategies by any stretch of the imagination" (Wakefield, 2000, p. 260). So, according to Wakefield, only some range of the continuum of depressiveness was selected for its beneficial effects and what falls under our concept of depressive disorder is clearly out of this range.

But again, if we conceive of the HDA as a conceptual analysis, then Wakefield's reply is problematic. Whether a certain genetic variant is selected for or against

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

depends not only on the phenotype but also on which other variants exist in a population. Imagine, in an analogous case, a population in which individuals have either very high levels of fear or no fear at all. Even though the chances of survival and reproduction may be low for individuals with high levels of fear, they might still be higher than the chances of survival and reproduction for individuals with no fear at all, since in evolutionary terms, it is better to be "safe than sorry". It is clearly more adaptive to be able to experience fear (in order to avoid dangerous situations) than to lack this ability completely. Hence, it may be the case that extreme cases of fear (or depressiveness) were selected for their effects and possess a biological function after all. Regardless of that, extreme fear is considered to be pathological.

Now, if we conceive of the HDA as an explication, then Wakefield can be revisionary about the imaginable counterexamples and argue that we *should* not think of spandrels and adaptations as disorders. But then the question is: why should we, as theoreticians, restrict the concept of mental disorder to biological dysfunctions in terms of PBFs to begin with? In the end, the biological dysfunction condition of the HDA is not a conceptual condition, but an empirical one, and as we have seen in the explication of PBF, a highly demanding one, because it requires heritability. But the relevance of the evolutionary perspective for the *concept* of mental disorder is not evident. As Murphy (2020) points out, medicine does not make such a restriction and, as Tsou (2021, p. 44) argues, it would be "pragmatically indefensible" if we, as members of a society, stopped considering depression or post-traumatic stress disorder (PTSD) as mental disorders if it turned out that they are not caused or constituted by biological dysfunctions.[12]

Wakefield could simply argue that medicine *should* make such a restriction and that we *should* stop considering depression or PTSD as mental disorders if it turned out that they are not caused or constituted by biological dysfunctions in terms of PBFs, because this is what our best scientific theory of the nature of functions and dysfunctions yields and our concept of mental disorder depends on the concept of dysfunction. However, as I'll argue in section 2 and section 3, I contend that when it comes to *mental* disorder, the relevant concept is a different one: the inability to respond adequately to one's available reasons.

In any case, it is worth noting that the HDA has two drawbacks. First, the status of the phenomena scientists consider mental disorders as mental disorders is *preliminary*, because it depends on our knowledge of the evolution of the mind, which so far is very limited. Second, the HDA *restricts* the causal explanations of mental disorders, because it depends on a highly demanding concept of biological

[12] Would it really be pragmatically indefensible? What if we put them in a different category, such as "socially accepted mental problems"? This is an interesting possibility and I thank an anonymous reviewer for raising this point. Tsou would have to clarify that at this point, he uses "mental disorder" in the widest sense of the term which encompasses all conditions that "depart from mental health". This is compatible with the view that there might be subcategories, for example "mental disorders" in a more narrow sense and "socially accepted mental problems" which do not fit that narrow definition, but are nonetheless departures from mental health.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

dysfunction. As long as we, as scientists, don't know enough about the evolution of the mind and the exact causes of mental disorders, a definition should stay neutral with respect to these issues to avoid unnecessary restrictions and a preliminary status of our classification. Even more, defining "mental disorder" is a conceptual project that "picks out" certain phenomena as a set; finding out about the causes of mental disorder (that is of the phenomena "picked out" by the definition of mental disorder) is an empirical project that explains the occurrences of those phenomena. It is not the job of a definition to tell us about the causes of mental disorder and the evolution of the mind, but the job of the empirical sciences. A definition of "mental disorder" which remains neutral on the evolution of the mind and the causes of mental disorders would be preferable, because it would enable us to keep these projects apart and remain open with respect to new empirical findings.

Most importantly, I contend that the HDA is missing something: it does not explicate the sense in which mental disorders are "mental" and does not relate the disorder concept to that specific concept of the mental. As a consequence, it does not track what ascribers of mental disorder primarily care about when talking about "disorder" in the psychiatric or psychotherapeutic context. To clarify this point: To evaluate whether an individual has a mental disorder, any ascriber of mental disorder should evaluate primarily the *rational relations* among their attitudes (and actions). Consider anxiety disorder. To evaluate whether an individual has an anxiety disorder, one has to evaluate whether their fear *makes sense* or is *reasonable in light of their own epistemic situation.* Do they have beliefs that give them sufficient reason to fear the situations they are, in fact, afraid of? If they do, they do not have an anxiety disorder (unless they don't have sufficient reason for their beliefs to begin with), even if their fear is objectively inadequate. Only if they don't, they might have one. But the HDA does not relate the disorder-status of mental disorders (such as anxiety disorder) to the concept of rationality, but to the concept of biological function of the relevant mental states (such as fear). However, to evaluate an individual's reasonableness of their fear, the ascriber is not committed to any specific view about the evolution of fear. Nothing hinges on whether fear is the product of evolution or an evolutionary by-product. Moreover, it seems that any rational creature can be the bearer of a mental disorder, whether it is the product of an evolution or not.

1.2 Problems of the BST

Boorse's (2011, p. 20) BST offers a conceptual analysis of the theoretical concepts of health and disease (in my terminology: disorder) in scientific medicine. Thus, its adequacy should be evaluated with respect to which conditions medicine actually considers to be healthy or disordered. The BST has been criticized extensively and Boorse (1997, 2014) offers a nearly comprehensive response to his critics. I focus on some problems which, as I will argue, are still unresolved.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

(1) Not necessary. One problem for the BST is that it does not capture universal disorders. More specifically, the BST yields that universal biological dysfunction is conceptually impossible. But this is false. Conceptually, it *is* possible that all tokens of a certain type of entity have a biological function, but do not function properly. So, a conception of biological function that does not capture this conceptual intuition does not qualify as an adequate analysis of it. Melander (1997, p. 57) gives the following example:

> If reticulum cancer were to become pandemic in the bovine population thereby making all or most bovine reticulums unable to break down cellulose, bovine reticulums would not typically or normally be able to break down cellulose. But contrary to the proposal, to break down cellulose would then still be a function of bovine reticulums.

Neander (1991, p. 182), too, argues that the BST yields an absurd consequence, because "if enough of us are stricken with disease (roughly, are dysfunctional) we cease to be diseased, which is nonsense". If we all go blind, blindness is still a dysfunction. On her view, it is conceptually possible that a biological dysfunction affects all members of a class of organisms. Spreading a disease does not make the condition less of a disease. Boorse (2002, p. 95) argues that vital biological dysfunctions, if universal, would simply extinguish the species. However, even if true, this would not make it impossible for an entire species to fall ill with a deathly disorder.

On Boorse's view, only less than vital universal disorders are a threat to his view, because *de facto* there are no vital universal disorders. To capture less than vital universal disorders, Boorse (2002, p. 95) proposes that we have "to use an extended time-slice of the species" to determine which mechanisms are normal for members of a species to have. The NPFs of a species are not determined only by the currently living members of that species, but also by a set of past members. So, for some mechanism to have a NPF, it will have to have had it for a sufficient period of time. How long is sufficient? According to Boorse (2002, p. 99),

> any time-slice shorter than a lifetime or two seems too short for the very idea of a species-typical functional design, since identifying many functions in maturation and reproduction requires a longitudinal view of an individual organism and its progeny.

Boorse points out that this is vague and that at some point vagueness is inevitable.

Interestingly, Boorse's proposal for less than vital universal disorders indicates that, in the end, our ascriptions of biological functions do not track statistical normality, but rather traits that have been beneficent to our ancestors. In a nutshell, it seems that identifying NPFs with respect to time-slices of a certain species might be a way to identify some of the PBFs that contributed to the inclusive fitness of that species. This makes sense, since one would expect that statistics follow function and not the other way around.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

There is another reason for believing that the presence of a biological dysfunction in terms of NPFs is not necessary for having a mental disorder. Tsou (2021, p. 31) points out that certain mental disorders "might turn out to be underwritten by biological mechanisms that behave in predictable ways, but fall within the (statistically) normal range of biological functioning".[13] This echoes the spandrel objection raised against the HDA: it seems that we simply do not know enough about the causes of mental disorders to evaluate whether the phenomena falling under our actual concept of mental disorder are caused by biological dysfunctions in terms of NPFs. In light of this, a definition of "mental disorder" that remains neutral with respect to the causes of mental disorders seems preferable.

(2) Not sufficient. Critics argue that the BST falsely yields that homosexuality is a mental disorder and that it must be refuted because of that (for example, Cooper, 2005, p. 17 and Heinz, 2014, p. 44). But this is only approximately true. The BST *would* yield that homosexuality is a mental disorder if, for instance, the following were true: In homosexual individuals, the mental mechanism responsible for heterosexual desire is not instantiated on statistically normal occasions C with at least the efficiency statistically normal for the members of the relevant reference class of O. Causing or constituting heterosexual desire, however, is the NPF of that mechanism, because it is statistically normal for members of the relevant reference class of O to have that mechanism *and* having heterosexual desires contributes causally to the individual survival or reproduction of the members of the relevant reference class of O.

The crucial (empirical) question is whether there actually is a mental mechanism that is responsible for heterosexual desire. If there were such a mechanism, the BST would yield that homosexuality is a mental disorder in the theoretical sense. Boorse himself does not consider this problematic, because (1) for empirical reasons the case is not clear and (2) even if empirical research indicated that there is such a mechanism, it would only indicate that homosexuality falls under a theoretical concept of disorder. But since normality does not entail desirability, this would have no practical significance: "We always have the right to ask, of normality, what is in it for us that we already desire." (Boorse, 1975, p. 63)

What should we make of this? Since Boorse is interested in an analysis of the theoretical concept of disorder and homosexuality is not considered a mental disorder in clinical psychology or psychiatry, the BST is inadequate as an analysis of our actual theoretical concept of mental disorder. Our actual view is not that we, as scientists, do not have sufficient empirical evidence to know whether homosexuality is a mental disorder, but that it is not; not even in a theoretical sense of the term. For people interested in an explication of the concept of mental disorder, this leaves the question open whether we *should* consider homosexuality a mental disorder. Here the answer is "no". Being homosexual does not give its bearer a *pro tanto* reason for seeking psychiatric or psychotherapeutic treatment (if available). The fact that someone is homosexual is not worthy of psychiatric or

[13] He refers to Maung (2016) and Stegenga (2018, ch. 4).

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

psychotherapeutic concern (though the fact that they are subject to stigmatization might be).

Another example showing that the presence of a biological dysfunction in terms of NPFs is not sufficient for having a mental disorder is *diminished jealousy*. According to the BST, an individual would have a mental disorder if there were a mechanism for jealousy instantiated in the individual on statistically normal occasions with less than statistically normal intensity. However, diminished jealousy is not a mental disorder and shouldn't be considered one, because it is not worthy of psychiatric or psychotherapeutic concern.

Finally, a point that applies to PBF and NPF views of biological dysfunction. The proper bearer of a biological function is a part, process, or mechanism of an organism. In any case, the bearer is a sub-personal entity, that is, something that is "responsible" (either causally or constitutively) for a mental phenomenon which we attribute to individuals as a whole, but not the mental phenomenon itself. Mental disorders, however, are ascribed to individuals as a whole. For example, we attribute both the mental phenomenon of fear as well as an anxiety disorder to individuals as a whole, but the PBF or NPF to produce fear is attributed to the fear mechanism (if there is any). This makes it possible that a dysfunctional mechanism gets compensated by some other mechanism so that it does not have any "person-level" effects. So, the presence of a biological dysfunction on a sub-personal level does not guarantee the presence of a mental disorder as a personal-level phenomenon.

2 The rehability view

In this section, I present the view I want to defend. I propose the following Rehability View (RHA) of mental disorder[14]:

RHA An individual S has a mental disorder if and only if S is *unable* to respond adequately to some of their available *reasons* in some of their reasons-sensitive attitudes or actions in view of their mental constitution and their life circumstances, where the threshold of inability is determined by S's being *harmed*.

The idea is this: When we ascribe a mental disorder to an individual, we basically ascribe to them that they "cannot" (in the relevant sense) respond adequately to some of their available reasons.[15] There are reasons for (or against) attitudes such as beliefs and emotions; as well as reasons for (or against) actions. To respond to

[14]For similar views, see Edwards (1981, p. 312), Gaete (2008, p. 331), Graham (2010, p. 117), and Nordenfelt (1987/1995). Ability-based views of mental health and disorder (Gaete, 2008; Nordenfelt, 1987/1995) are often deemed to be too broad. The view I propose aims to amend this shortcoming by making the relevant set of inabilities more precise and by drawing on some of the most recent literature on the concepts of ability, reasons, and harm. In doing so, it also aims to make the rationality claims in Edwards (1981) and Graham (2010) more precise.

[15]If you believe that "having" certain reasons already entails that they are in some sense "available" to you, then noting that they are "available" is redundant.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

one's available reasons for attitude A (or action φ) is, generally, to form attitude A (or to φ/intend to φ). To respond to one's available reasons against attitude A (or action φ) is, generally, to omit forming attitude A (or to omit φ-ing/intending to φ). Only if the individual is harmed, their condition is "clinically relevant", that is, worthy of psychiatric or psychotherapeutic concern. Only then are they in the clinically relevant sense "unable" to respond adequately to some of their available reasons.

A caveat: we should not confuse an inability to φ with an inability to *learn* to φ. The latter is a second-order ability: the ability to acquire the ability to φ. That S does not have the ability to φ does not imply that S does not have the second-order ability to learn to φ (or: the potential to φ). I may currently lack the ability to do seventy-five pull-ups, but this does not imply that I do not have the potential to do it. In addition, individuals may not only acquire abilities, but also lose them. I used to be able to speak French pretty well, but today: pas tellement.

To motivate the idea that mental disorders are intimately connected to inabilities, imagine an individual with

- an anxiety disorder and saying to them "Just relax!"
- a major depressive disorder and saying to them "Just cheer up!"
- a delusional disorder and saying to them "Just stop thinking that someone is following you!"
- an addictive disorder and saying to them "Just stop using!"

Not only are these responses rude and unhelpful, but they also seem to miss the crucial point. As I will argue, there is at least one relevant sense in which an individual with an anxiety disorder precisely "cannot" stop experiencing fear in certain situations; an individual with a major depressive disorder precisely "cannot" stop feeling depressed; an individual with a delusional disorder precisely "cannot" stop believing that somebody is following them; and an individual with an addictive disorder precisely "cannot" stop using drugs.[16]

If having the ability to φ is a necessary condition for being obliged to φ, then demanding of someone to φ when they precisely "cannot" φ in the relevant sense is, above all, inadequate. On this view, unless an individual "can" φ (in the relevant sense), we cannot demand of them to φ. The challenge for an ability-based view of mental disorder is to specify the exact sense in which an individual with a mental disorder "cannot" φ.

Let me describe the three core concepts of the RHA – ability, reasons, and harm – in more detail.

[16]See Dembić (2021) for a detailed account of addictive disorder in terms of inabilities.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

2.1 Abilities

According to the RHA, having a mental disorder necessarily involves a certain *inability*, to wit *in view of one's mental constitution and one's life circumstances,* where the *threshold* of inability is determined by S's being harmed. Let me clarify these points.[17]

First, having an ability is a modal property. In stating that S has the ability to φ, we state something about what S *can* do, and not primarily something about what S actually does. Of course, given that S φs, there seems to be a sense in which S "can" φ. In light of this, an individual's actual behavioral pattern may sometimes play a heuristic role: given that S φs, we have reason to believe that S has the ability to φ. However, the following does not hold: S's not φ-ing does not necessarily indicate that S is unable to φ. An individual can have the ability to φ without ever actually φing.

Second, we should understand locutions such as "S is unable to φ" or "S does not have the ability to φ" as follows: S does not have the ability to φ *to a sufficient degree*. This is because having an ability is typically not an all-or-nothing matter, but a matter of degree. I can dance better today than ten years ago (I had dance lessons), but Mikhail Baryshnikov is a better dancer than I am. We can compare an individual's abilities over time or compare different individuals with respect to their abilities (at the same time or over time). Theoretically, it is possible to arrange individuals along a scale of a certain ability in an ascending order. However, though abilities come in degrees, we often make categorical ascriptions: S "has" or "does not have" the ability to φ, period. And we often ask whether an individual has a certain ability, thereby expecting a yes-or-no answer. But that S does not have the ability to φ does not imply that S has the ability to φ to the degree of 0. For instance, if I were asked "Can you sing?" at a casting for an opera, it would be misleading of me to answer "yes". But, of course, this does not imply that I cannot sing *at all*. What I said or meant was that I cannot sing *well enough* to be part of the cast in an opera.

This specification leads to a further question: what fixes the degree above which an ability counts as "sufficient"? In other words, what determines the threshold on the scale of a certain ability, above which ascriptions of that ability apply categorically? One possible solution is to settle this by the speaker context – the context in which the ascription is made – or more precisely: by the standards that obtain in a given speaker context (Jaster, 2020, ch. 4).[18] In different practical contexts, different standards obtain. Which standards obtain, in turn, depends on the interests of the speakers or the *purposes* for which they make the ability-ascriptions in the first place. For instance, in the context of a casting for an opera the degree

[17] See Jaster (2020, ch. 1) for an informative overview of some general features of abilities. The following summarizes the main points.

[18] See Stalnaker (2014) on the concept of context.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

of singing that counts as "sufficient" will be higher than the one that obtains in the context of a karaoke bar.

For which purposes will an individual's abilities to respond adequately to their available reasons have to be good enough for them to count as *not* having a mental disorder? I contend that, most generally, the answer is this: to achieve a certain level of *well-being*. When we ascribe to an individual a mental disorder, we ascribe to them that some of their abilities to respond adequately to their available reasons is so low that they are *harmed* in some respect of their well-being. To spell out a substantive theory of well-being and harm that is relevant to the context of psychiatry is a task for psychiatric ethics and exceeds the scope of this paper. Nevertheless, I make some remarks on the concept of harm in subsection 2.3.

At this point, a critic might ask[19]: but what exactly is the threshold (especially since harm seems to come in degrees as well)? I contend that the threshold varies with the context and that it is not up to the theoretical philosopher to settle the exact threshold for the psychiatric context. On the view I propose, there simply is no "naturally given" threshold to which a theoretician could point. Just as mental disorder, abilities come in degrees (and thus, present a "continuum" in a certain sense). Any categorical distinction (ability/inability, mental health/disorder) is one that is *made within that continuum by speakers for certain purposes*. The threshold for psychiatric purposes (say, diagnosis and decisions concerning therapy) will have to be set by the psychiatric context. The exact threshold should not be set from the "armchair", that is, from a philosophical and purely theoretical point of view.

Third, abilities are always had "in view of" some facts. This traces back to Kratzer's modal semantics. According to Kratzer (1977), there is no absolute sense of "can". Rather, "S can φ" has to be understood as "S can, in view of F, φ", where F specifies a contextually selected set of facts in view of which the ability is said to be had. So, when evaluating whether S has the ability to φ, we first have to specify the facts which are relevant to the context in which we are interested in knowing that.

To illustrate this, consider a professional swimmer with a broken arm. Can she swim? In some sense, she can (she is a professional swimmer), but at least in one relevant sense, she cannot (she has a broken arm, after all). We can distinguish these senses by specifying the relevant facts *in view of* which we evaluate whether she can swim. In view of the fact that she is a professional swimmer and abstracting away of the fact that she has a broken arm it is true to claim that she can swim. But, in view of the fact that she has a broken arm, it is also true to claim that she cannot swim. In the second sense, we do not abstract away of the fact that she has a broken arm. In the literature, this distinction is typically expressed by saying that she has the "general" but not the "specific" ability to swim (Honoré, 1964 and Mele, 2003).

[19] I thank an anonymous reviewer for raising this point.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

Which facts are relevant when evaluating whether an individual has a mental disorder? I contend that in the psychiatric and psychotherapeutic context, ascribers of mental disorders are primarily interested in the individual's abilities in view of their *mental constitution*. Mental constitution is a vague concept: it comprises an individual's *relatively stable* attitudes, personality traits, attribution styles, and so on. To specify them in more detail is the job of psychology and the closely related sciences. When ascribing a mental disorder, we are not interested in the individual's abilities in view of every single one of their actual mental states at the time of diagnosis. To determine whether I have a major depressive disorder, for instance, we should not consider my brief depressiveness after reading "All Quiet on the Western Front". In light of that particular mental state, it might turn out that I have a depressive disorder, though I surely need not have one. In ascribing a mental disorder, we ascribe an inability in a more general sense.

Furthermore, some facts *external* to the individual are relevant too. Even abilities that appear to be fully "intrinsic" are only had in view of certain extrinsic facts. For instance, when we evaluate whether an individual has the ability to hit the bull's eye, we also hold the actual laws of nature fixed. I contend that the external circumstances that matter for mental disorder are all circumstances in the individual's life which they cannot easily change. I call these the individual's "life circumstances". Again, this is imprecise and will have to be specified by the relevant sciences.

The RHA can incorporate psychiatry-critical ideas such as that sociocultural structures, power relations, and structural barriers construct disability and mental disorder.[20] Recognizing that abilities relevant to mental disorder are always relative to certain life circumstances has important consequences for therapy, because, in principle, enabling an individual with a mental disorder could be achieved by other means than by "changing" their mental constitution, namely by changing their life circumstances. So, the RHA is not committed to the claim that to enable an individual with a mental disorder, it is necessary to change their mental constitution.

2.2 Reasons

According to the RHA, the abilities relevant to mental disorder are abilities to respond adequately to one's *available reasons*. Thus, mental disorder involves φ-ings (or aspects of φ-ings) which are *sensitive to reasons*. For some φ-ing (or aspect of φ-ing) to be "sensitive to reasons" means that with respect to that φ-ing (or aspect of φ-ing) it makes sense to ask the question "Why?" in a certain sense (Anscombe, 1957, p. 9). There are at least three different types of why-questions, only two of which reveal the "sensitivity to reasons" in the relevant sense. To illustrate them, compare the following examples:

[20] See Foucault (1965, 1973, 1977) and Tremain (2015).

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

1. *Evidential reasons*: "The last train to Berlin leaves before midnight." "Why should I believe that?" "That's what the schedule says."

2. *Reasons for action*: "Jen considers buying a gift for Berislav." "Why should she do that?" "To cheer him up."

3. *Mere causes*: "The candle went out." "Why did that happen?" "There was a draft."

In the literature, answers to the types of questions asked in (1) and (2) are typically called "normative" or "justifying" reasons. The answer in (3) is only superficially similar to (1) and (2). It is only for attitudes and actions that a why-question as in (1) or (2) can meaningfully be asked – or in other words: that normative reasons can be meaningfully asked for and offered. Mere events or processes such as a candle going out can be explained by referring to causes, but it does not make sense to ask for or to offer something like normative reasons for or against them. In light of this, it would be misleading to call the things refered to in (3) as "reasons", even though it is an answer to a certain type of why-question. To avoid misunderstandings, it is better to call them "mere causes".

A prevalent view of normative reasons as deployed in (1) and (2) is the following: a fact (or true proposition) gives us a normative reason when it *counts in favour* of (or against) our responding in some way, where the response is an attitude (of some type) or an action (of some type).[21] As such, normative reasons *support* some type of response (Kiesewetter, 2017). Normative reasons that support a response of some type are *pro tanto* reasons. If S has *sufficient reason* for a response (of some type), then S is justified to exhibit that response. If S has *decisive reason* for a response (of some type), then S should or is required to exhibit that response.

Actions are rationally evaluable with respect to the practical normative reasons rational creatures have for them, that is, with respect to the facts (or true propositions) that count in favour of some action φ of ours to be good or worthwhile of pursuit. We can deliberate on the *pro tanto* reasons we have for or against some action φ, weigh these reasons against one another, and settle the question on whether we have, all things considered, sufficient (or decisive) reason to φ. Attitudes such as beliefs or emotions are rationally evaluable (1) with respect to whether they "fit together", that is, whether they are consistent or coherent with one another and (2) with respect to the situations we are in, that is, with respect to the facts (or true propositions) that count in favour of the belief being true or the emotion being objectively adequate.

On the prevalent view, a fact gives us a normative reason when it counts in favour of our responding in some way, regardless of whether we have a belief

[21] See, after Kiesewetter (2017), Scanlon (1998, p. 17), Dancy (2000, p. 1), Velleman (2000), Gibbard (2003, pp. 188–189), Finlay (2006, p. 5), Thomson (2008, p. 127), Raz (2009, p. 18), Parfit (2011, p. 31), Broome (2013, p. 54).

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

related to that fact. But, of course, to play a role in an individual's reasoning or deliberation about what to do or what to believe, the normative reasons the individual has will have to be present or at least "available" to them in some sense (say, as dispositions in their memory). For instance, the fact that I have a nut allergy gives me a normative reason not to eat nuts, regardless of whether I am aware of that fact. Even more, I need not even believe that I have such an allergy. But, of course, if I do not believe that I have a nut allergy, I cannot (be expected to) respond to the normative reason that the respective fact gives me. This is why the relevant abilities in mental disorder are abilities to respond adequately to one's *available* reasons. Having knowledge of a normative reason is sufficient for it to be available to the individual knowing it.

Here is a worry concerning the claim that in mental disorder, the φ-ings (or aspects of φ-ings) involved are those that are sensitive to reasons. A critic might object that this view yields some obviously false verdicts, for instance, that *mood disorders* are not mental disorders. Why? Because on a prevalent view, moods such as depressiveness are not sensitive to reasons. Moods are often distinguished from emotions. Whereas it is relatively uncontroversial that emotions such as sadness are sensitive to reasons, it is controversial whether moods are. Emotions such as sadness can be objectively adequate or inadequate and they can "fit" to our beliefs or not. But with moods, the case appears to be different: when you wake up in the morning in a good mood, you typically do not think that you have a reason for being in such a good mood. You just happen to be in a good mood and it does not seem to be *about* anything. *Prima facie*, the RHA yields that mood disorders are not *mental* disorders.

However, I contend that this line of thinking involves an implausible view of (many) moods. I contend that moods such as depressiveness are sensitive to reasons. It makes sense to ask an individual for their reasons for or against being in a particular mood. The mere fact that we often do not know why we are in a particular mood does not imply that the question does not apply. Perhaps, the question does not apply to certain moods such as waking up in the morning in a good mood. But mostly, as Prinz (2004) argues, moods can be objectively adequate or inadequate in light of how life is going for the affected individual quite generally.[22] Because of that, they are just as sensitive to reasons as emotions are.

2.3 Harm

According to the RHA, having a mental disorder necessarily involves harm. Harm is typically contrasted with well-being. Typical examples of harm are: pain and

[22] Prinz (2004, p. 185) argues that the difference between emotions and moods is not whether they represent, but what they represent: "Sadness represents a particular loss, while depression represents a losing battle."

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

suffering.[23] An event *e* (or derivatively a property of, or a continuant participating in it) that causes some individual S some harm in some respect X – where X is a component of their well-being or other intrinsic good – is a "harmful event". An individual S who is caused some harm in some respect X by some event *e* is "harmed". For example, a car crash causes S to have a painful fracture which is a harm to them in some respect, because it is painful. We can say that the car crash is a "harmful event", that S was "harmed", and that the painful injury is a "harm". I contend that for S to be caused some harm in some respect X is for S to be caused to be worse off in respect X than before.[24] Since harm comes in degrees, I understand the locution "to be harmed" as "to be harmed to a sufficient degree".

Harm and well-being are not exhaustive. Some events are neither harmful nor beneficent to a certain individual. For instance, that I am working on my paper today is neither harmful nor beneficent to LeBron James in any respect whatsoever. Also, harm needs to be distinguished from a mere deprivation of good. An individual deprived of a good is not necessarily harmed. For instance, an individual who does not win the lottery is deprived of a good, but they are not necessarily harmed by that in any respect. However, it is possible that an individual who is prevented from receiving a good is harmed in some cases, namely when it leaves them in a state which is bad to begin with. Thus, an individual is not only harmed when caused to be worse off in some respect X than before, but also when being prevented from receiving a good in some respect X and therein being left in a bad state in respect X.[25]

Some event can either be intrinsically or instrumentally good/bad for an individual. An event is intrinsically good for an individual if it has value for them in itself and it is instrumentally good if it has value for them for the sake of something else.[26] For instance, dancing is intrinsically good for an individual if it has value for them in itself. Having a fever may be instrumentally good for them if it has value for her for the sake of something else, say, for the sake of avoiding a difficult meeting at work.

As a consequence, something may at the same time constitute an intrinsic harm for an individual and be instrumentally beneficent to them. For instance, given that vomiting is painful or unpleasant, vomiting is an intrinsic harm for an individual. However, given that it prevents them from further pain by poisoning, it is, at the same time, instrumentally beneficent to them.

[23] What about masochism? Masochism is defined as the practice of seeking pleasure or gratification by means of pain. This does not show that pain is not intrinsically bad, but rather, that something intrinsically bad can be instrumentally good.

[24] This view is similar to the so-called historical view of harm (Rabenberg, 2014; Shiffrin, 2012). See Feinberg (1986), Parfit (1989), and Klocksiem (2012) for a counterfactual comparative view of harm, see Bradley (2012) and Rabenberg (2014) for a critique.

[25] See Rabenberg (2014, p. 19).

[26] This is compatible with the view that something is valuable only if and because it is valued by someone.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

An individual can be *pro tanto* or *all things considered* harmed by an event. For instance, a visit to a dentist is typically *pro tanto* harmful for a patient, because it typically involves causing them an unpleasant or painful experience. But the same visit may be harmless or even beneficent for them on balance or all things considered, because it ultimately prevents them from further unpleasant or painful experiences.[27] Furthermore, some type of event can be harmless or harmful in a short-term, but harmful or harmless in the long-term. Smoking *one* cigarette probably never killed anyone, but smoking on a regular basis is clearly harmful to smokers in the long-term.

It is uncontroversial that in having a mental disorder, the affected individual's condition can be instrumentally beneficent to them. Also, they need not be all things considered harmed by their condition. However, for the condition to fall under the concept of mental disorder, it needs either to constitute an intrinsic harm or to be instrumentally harmful to S in some relevant respect.

An individual S can be harmed by their condition C at least in three different ways:

1. C is intrinsically bad for S in some respect X *or*
2. C causes S to be worse off than before in some respect X *or*
3. C prevents S from receiving a good in some respect X and therein leaves them in a bad state in respect X, *where X is some component of S's well-being.*[28]

A comprehensive view of mental disorder will have to be supplemented with a substantive theory about the components of human well-being that are relevant to the psychiatric and psychotherapeutic context. (In doing so, any proponent of the view that harm is necessary to mental disorder would also have to deal with the objection that some mental disorders, such as mania, apparently come with heightened well-being.) As already stated in subsection 2.1, providing such a theory exceeds the scope of this paper.

At this point, it becomes clear that the concept of mental disorder is not free from normative considerations. We, as rational creatures and members of a society, "expect" people to be able to cope with certain problems of their everyday living, in the sense that we hold them to certain standards, not simply in the sense that they will typically do so. So, whether a condition falls under the concept of mental disorder depends, in part, on our normative expectations. Sometimes we criticize the normative expectations that *de facto* obtain in a sociocultural context. For instance, we may believe that "too many" individuals in a given sociocultural

[27] See Bradley (2012, p. 393). Sometimes the distinction is drawn between *prima facie* harm and harm all things considered, see Klocksiem (2012). But to call it "*prima facie* harm" is misleading, because it suggests that something only *appears* to be harmful, but actually is not. However, there is a sense in which, for instance, chemotherapy *is pro tanto* harmful, although it can be all things considered beneficent.

[28] This is a version of Rabenberg (2014).

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

context count as having a mental disorder. In such cases, we also tend to believe that our expectations obtaining in that context are too high and we do so, because we believe that we have *reasons* for this kind of criticism. Thus, any comprehensive view of mental disorder will not only have to answer the empirical question "What normative expectations do we *de facto* have in a given sociocultural context X?" but also the normative one "What normative expectations should we have in a given sociocultural context X?".

2.4 Example: Anxiety disorder

To make the RHA more tangible, let me illustrate it with an example: anxiety disorder. In light of the RHA, to evaluate whether an individual has an anxiety disorder, we first have to ask: does the individual have the ability to respond adequately to their available reasons against their fear in view of their mental constitution and their life circumstances? If the answer is "yes", their fear is non-pathological; if the answer is "no", their fear is pathological just in case the individual is also harmed by the resulting condition.

More formally, a view of anxiety disorder in terms of the RHA could look roughly like this:

RHA$_{\text{anxiety}}$ S has an anxiety disorder (if and) only if

1. S experiences fear in situations c_1, c_2, ..., c_n *and*
2. S is *unable* to not experience fear in situations c_1, c_2, ..., c_n in view of
   a. S's mental constitution (including the fact that S has sufficient available reasons against experiencing fear in situations c_1, c_2, ..., c_n) *and*
   b. S's life circumstances, where the threshold of inability is determined by S's being *harmed*.

According to the RHA, an anxiety disorder is a *mental* disorder, because a certain ability to *respond adequately to one's available reasons* is impaired or underdeveloped. The relevant adequacy of mental conditions is measured by the standards of reasons (and not primarily by biological, social, or moral standards). *Harm* due to a (low) level of *ability* to respond adequately to one's available reasons is what makes the condition count as a mental *disorder*.

It is worth noting that the RHA does not locate mental pathology in the veracity of the content of mental states, but in the individual's *epistemic situation*, that is, in the relations among their mental states. The RHA spells the individual's "epistemic situation" out in terms of their ability to respond adequatly to their available reasons.

Let me clarify this point. To ascribe an anxiety disorder to S, we do not have to evaluate whether S's fear is objectively adequate, that is, whether S is actually

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

in danger when in fear. Experiencing objectively inadequate fear is neither necessary nor sufficient for having an anxiety disorder. It is not necessary because, conceptually, it is possible for S to have an anxiety disorder *and* S's fear to be objectively adequate, namely in cases in which S is actually in danger but does not *know* or *justifiably believe* it. One could easily imagine cases in which S suddenly experiences intense fear, truly believes that they are in danger, but has no available reason whatsoever to believe that they are in danger. In such cases, S's fear would be objectively adequate, but still pathological. Why? Because S's fear is inadequate in light of their epistemic situation.

Experiencing objectively inadequate fear is also not sufficient for having an anxiety disorder. It is not sufficient because it is possible for S *not* to have an anxiety disorder even though S's fear is objectively inadequate, namely in cases in which S justifiably believes that they are in danger. Again, one could easily imagine cases in which S has every available reason to believe they are in danger even though, in fact, they are not. There can be, after all, misleading evidence. Suppose, for instance, that S sees a tiger mock-up. This causes them to believe that they see a tiger and that there is a tiger. Since S also believes that tigers are dangerous, S's fear would be subjectively rational. S's belief that there is a tiger gives them sufficient apparent reason for their fear. Nevertheless, S's fear is objectively inadequate, because, in fact, they are not in danger. In such cases, S's fear would not be pathological. Why? Because S's fear is adequate in light of their epistemic situation.

3 Why adopt the rehability view?

In this section, I argue that the RHA does not have the aformentioned problems of the HDA and the BST and that it has some further merits.

3.1 The rehability view vs. biological dysfunction views

First, contra the HDA, the RHA has no problem with disorders of *spandrels* or *adaptations*. Furthermore, contra the BST, the RHA is compatible with the idea that there might be mental disorders that are caused or constituted by biological mechanisms that fall within the (statistically) *normal* range of *biological functioning*. This is because the RHA is not committed to any particular causal story of mental disorder. According to the RHA, having a mental disorder is a matter of having a certain inability in view of one's mental constitution and one's life circumstances. To have an inability is simply to exhibit a certain *modal* pattern. The RHA is compatible with different views on how to spell out the individual's mental constitution that underlies this pattern. An inability in view of one's mental constitution and one's life circumstances might be due to an impairment (a "broken" or dysfunctional biological mechanism) or due to an underdeveloped biological mechanism, but it might also be due to the fact that the individual's life circumstances

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

are such that they are too great a burden to bear. To illustrate the last point, think of a person who has an addiction when at war, but who has no trouble stopping use when back home. The RHA can capture the fact that it is not always required to change one's mental constitution to get rid of a mental disorder, but that sometimes it suffices to change one's life circumstances.[29] In sum, the RHA puts no restrictions on the causal explanations of mental disorders.

Second, according to the RHA, the disorder status of the phenomena we consider mental disorders does not depend on our knowledge of the *evolution of the mind*, because the RHA is not dependent on any evolutionary concept. Because of that, the disorder status of the relevant phenomena is not (in that sense) preliminary.

Third, the RHA can capture *universal disorders*. Humans, in general, are able to respond adequately to their available reasons in most of their reason-sensitive attitudes and actions. (Though they are also, in general, every now and then unreasonable.) Even individuals with mental disorders are, in general, able to respond adequately to their available reasons in *most* of their reasons-sensitive attitudes or actions. It's just in *some* that they are not. Their inability is "local", not global. To adopt Davidson's (1982, p. 169) words, just like irrationality, mental disorder "is a failure within the house of reason". However, nothing in the RHA hinges on how many humans are unable to respond adequately to their available reasons in some of their reason-sensitive attitudes and actions. In fact, it is compatible with the view that most or even all humans have such a local inability. Hence, it can capture universal mental disorders.

Fourth, *homosexuality*. One might worry that the RHA falsely yields that homosexuality is a mental disorder. A critic might argue as follows: To be homosexual is to have sexual desires only for people of one's own sex (or gender). Sexual desire is an attitude which is sensitive to reasons, because a certain question "Why?" applies to it. Now, consider an individual who lives in a society which prescribes the death penalty for homosexuality. It seems that the fact that there is a death penalty for homosexuality gives that individual a reason against their homosexual desire. But we shouldn't conclude that the individual has a mental disorder if they are not able to change their sexual desire in light of that reason.

The trouble with this objection is that the fact that there is a death penalty for homosexuality is no reason against one's homosexual desire, since it is a reason "of the wrong kind".[30] I do not have a sexual desire for a certain individual *because* the external circumstances are favourable. Considerations showing that it is good or bad for me to have a certain sexual desire in certain circumstances are like considerations showing that it is good or bad to believe something: they do not

[29] This does not imply that mental pathology can be constituted by life circumstances alone. According to the RHA, the relevant inability is always one in view of one's *mental constitution* plus life circumstances.

[30] See Hieronymi (2005, 2013) and Gertken & Kiesewetter (2017) for a discussion of the wrong kind of reasons.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

render the belief rationally intelligible to the believer. The sexual desire is not about the external circumstances, it is about a certain individual. Reasons for or against a certain sexual desire are given by the qualities of "the object" of desire: whether the sexually desired individual is worthy of sexual desire. Of course, the fact that there is a death penalty for homosexuality might be a reason for an individual not to *act* on their homosexual desire, but it is not a reason against the homosexual desire itself.

I contend that the RHA correctly yields that homosexuality is not a mental disorder, because it is simply not the case that objects of homosexual desires are unworthy of such desires because they are of the same sex. Nothing about the sameness of sex justifies such a verdict. This view rests on a substantive claim about values. But it seems to me correct to interpret potential conflicting views about the status of homosexuality as a mental disorder as conflicts about values (and not, as biological dysfunction views do, as conflicts which can be solved by empirical findings).

Fifth, the RHA does not capture *diminished jealousy* as a mental disorder, because it is not harmful. The view yields the correct verdict, because diminished jealousy is not worthy of psychiatric or psychotherapeutic concern.

Sixth, the RHA locates the mental disorder on the correct level of description: on the *personal level*. The bearers of mental disorders are individuals as a whole and not some of their parts or mechanisms. This is captured by the RHA, because the bearer of abilities and harm are also individuals as a whole.

Seventh, the RHA tracks what ascribers of mental disorder primarily care about when they talk about "disorder" in the psychiatric or psychotherapeutic context, because it ties the concept of disorder specifically to the concept of the mental which, in turn, is understood in terms of the sensitivity of attitudes or actions to (available) reasons. According to the RHA, the crucial question is basically whether an individual is able to think, feel, and act reasonably in light of their epistemic situation. The relevant standards of deviance are the norms of reasons and not the "norms" of biological functions (which are "norms" in a very different sense). Thus, the RHA can capture what is distinctive about *mental* disorders.

3.2 Further merits

In the following, let me point out three further merits of the RHA. (1) The RHA can capture that mental disorders come in *degrees*. (2) The RHA is conceptually unifying in that it leaves room for both biological and social aspects to play a role in specifying the concept of mental disorder. (3) The RHA can illuminate questions about symptoms, disorders, and their causes.

First, degrees. Mental disorders can be more or less severe. This can be captured by the RHA, because both abilities and harm come in degrees as well. Consider again the example of an anxiety disorder. In light of the RHA, it is easy to

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

determine the dimensions which are relevant to assess the severity of an anxiety disorder. The severity seems depend on:

1. the frequency, intensity, and duration of the experience of fear
2. the degree of (in)ability to not experience fear
3. the degree of harm

Generally, the higher (1) and (3) and the lower (2), the more severe the anxiety disorder will be. Furthermore, the RHA can acccount for temporary mental disorders (think of a brief psychotic episode). Mental disorders can be more or less stable depending on how stable the inability is. As with the threshold, the question of how long an inability needs to last to constitute a mental disorder will have to be set by the psychiatric context.

Second, biological and social aspects. According to the RHA, to have a mental disorder is to have a certain inability in view of:

- one's own *mental constitution* (relatively stable attitudes)
- and one's own *life circumstances* (relatively stable external circumstances)

The RHA doesn't exclude the relevance of *biological dysfunctions*, because it is possible that the individual's mental constitution supervenes on, is identical with, or is realized by a certain mechanism which doesn't perform some of its biological functions. Whether that is the case is an empirical question. If there is a biological dysfunction that corresponds to a certain type of mental disorder, we can specify the inability relevant to it by including the biological dysfunction in the "in view of"-part of the description.

The RHA doesn't exclude the relevance of *social aspects* neither, because an individual's relatively stable life circumstances include facts about their social environment as well. They include the society in which the individual lives in, the stable relationships and social roles they have, and so forth. If there is a stable set of social facts that corresponds to a certain type of mental disorder, we can specify the inability relevant to it by including those social facts in the "in view of"-part of the description.

Third, symptoms, disorders, and their causes. The RHA can make sense of the fact that in mental disorder neither the symptoms nor the causes of the disorder need to be mental and it can elucidate the relations among symptoms, disorders, and their causes. Symptoms are indicators of or evidence for a disorder. The RHA contends that having a mental disorder involves being in a certain mental condition and to have a certain modal property, namely the inability to (omit) φ(-ing). What the symptoms indicate is the individual's mental condition and some of their modal properties. In light of the RHA, it is easy to see how there can be symptoms without a disorder. For example, that an individual uses a certain substance may

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

be a symptom of an addictive disorder. But it need not. That an individual uses a certain substance does not imply that they are unable to omit it. To evaluate whether they have an addictive disorder, we have to evaluate their corresponding modal pattern. (For that, we may need an epistemology of abilities.) When we ask about the causes of mental disorder, we ask about the causes of an individual's mental condition and how they came to have a certain modal pattern.

4 Conclusion

In this paper, I proposed an ability-based view of mental disorder, according to which having a mental disorder is basically a matter of having a certain type of inability (or: an ability that is not sufficiently high): the *inability* to respond adequately to some of one's available *reasons* in some of one's reasons-sensitive attitudes or actions, where the threshold of inability is determined by one's being *harmed* (as relevant in the psychiatric and psychotherapeutic context). I argued that the proposed view evades some of the problems of two prominent biological dysfunction views of mental disorder, the HDA and the BST. Most notably, I argued that the RHA can account for what is specifically "mental" about mental disorders. It is the rational relations among an individual's attitudes and actions that is "disordered" and the relevant norms in mental disorder are the norms of reasons. Furthermore, I argued that the RHA can account conceptually for both, social as well as biological aspects of mental disorder. I conclude that the RHA is *conceptually unifying* while remaining *highly flexible* due to the fact that abilities are always had in view of certain facts and those facts can be of various types. As such, it presents a theoretical alternative that might be worthy of further exploration.

Acknowledgments
I gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG) within the Centre for Advanced Studies in the Humanities "Human Abilities", grant number 409272951.

References

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.

Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs*, *5*(1), 49–68. https://doi.org/10.1515/9781400853564.3

Boorse, C. (1976a). What a theory of mental health should be. *Journal for the Theory of Social Behaviour*, *6*(1), 61–84. https://doi.org/10.1111/j.1468-5914.1976.tb00359.x

Boorse, C. (1976b). Wright on functions. *Philosophical Review*, *85*(1), 70–86. https://doi.org/10.2307/2184255

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*(4), 542–573. https://doi.org/10.1086/288768

Boorse, C. (1997). A rebuttal on health. In J. M. Humber & R. F. Almeder (Eds.), *What is disease?* (pp. 1–134). Humana Press.

Boorse, C. (2002). A rebuttal on functions. In A. Ariew, R. C. Cummins, & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology* (pp. 63–112). Oxford University Press.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

Boorse, C. (2011). Concepts of health and disease. In F. Gifford (Ed.), *Philosophy of medicine* (pp. 13–64). Elsevier.

Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, *39*(6), 683–724. https://doi.org/10.1093/jmp/jhu035

Bradley, B. (2012). Doing away with harm. *Philosophy and Phenomenological Research*, *85*(2), 390–412. https://doi.org/10.1111/j.1933-1592.2012.00615.x

Broome, J. (2013). *Rationality through reasoning*. Wile-Blackwell.

Cooper, R. (2005). *Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders*. Springer.

Dancy, J. (2000). *Practical reality*. Oxford University Press. https://doi.org/10.1111/j.1933-1592.2003.tb00298.x

Davidson, D. (1982). Rational animals. *Dialectica*, *36*(4), 317–328. https://doi.org/10.1111/j.1746-8361.1982.tb01546.x

Dembić, S. (2021). Defining addictive disorder - abilities reconsidered. *Philosophers' Imprint*, *21*(24), 1–23.

Edwards, R. B. (1981). Mental health as rational autonomy. *Journal of Medicine and Philosophy*, *6*(3), 309–322. https://doi.org/10.1093/jmp/6.3.309

Faucher, L., & Forest, D. (2021). *Defining mental disorder: Jerome Wakefield and his critics*. The MIT Press.

Feinberg, J. (1986). Wrongful life and the counterfactual element in harming. *Social Philosophy and Policy*, *4*(1), 145–178. https://doi.org/10.1017/s0265052500000467

Finlay, S. (2006). The reasons that matter. *Australasian Journal of Philosophy*, *84*(1), 1–20. https://doi.org/10.1080/00048400600571661

Foucault, M. (1965). *Madness and civilization: A history of insanity in the age of reason*. Tavistock. https://doi.org/10.4324/9780203164693

Foucault, M. (1973). *The birth of the clinic: An archaeology of medical perception*. Pantheon Books. https://doi.org/10.4324/9780203715109

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Pantheon Books. https://doi.org/10.1515/9780822390169-058

Gaete, A. (2008). The concept of mental disorder: A proposal. *Philosophy, Psychiatry, and Psychology*, *15*(4), 327–339. https://doi.org/10.1353/ppp.0.0210

Gertken, J., & Kiesewetter, B. (2017). The right and the wrong kind of reasons. *Philosophy Compass*, *12*(5), 1–14. https://doi.org/10.1111/phc3.12412

Gibbard, A. (2003). *Thinking how to live*. Harvard University Press. https://doi.org/10.4159/9780674037588

Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptionist programme. *Proceedings of the Royal Society B: Biological Sciences*, *205*(1161), 252–270. https://doi.org/10.1098/rspb.1979.0086

Graham, G. (2010). *The disordered mind: An introduction to philosophy of mind and mental illness*. Routledge. https://doi.org/10.4324/9780203857861

Heinz, A. (2014). *Der Begriff der psychischen Krankheit*. Suhrkamp.

Hieronymi, P. (2005). The wrong kind of reason. *Journal of Philosophy*, *102*(9), 437–457. https://doi.org/10.5840/jphil2005102933

Hieronymi, P. (2013). The use of reasons in thought (and the use of earmarks in arguments). *Ethics*, *124*(1), 114–127. https://doi.org/10.1086/671402

Honoré, A. M. (1964). Can and can't. *Mind*, *73*(292), 463–479. https://doi.org/10.1093/mind/LXXIII.292.463

Jaster, R. (2020). *Agents' abilities*. De Gruyter. https://doi.org/10.1515/9783110650464

Kiesewetter, B. (2017). *The normativity of rationality*. Oxford University Press.

Klocksiem, J. (2012). A defense of the counterfactual comparative account of harm. *American Philosophical Quarterly*, *49*(4), 285–300. https://doi.org/10.1111/papq.12031

Kratzer, A. (1977). What ‚must' and ‚can' must and can mean. *Linguistics and Philosophy*, *1*(3), 337–355. https://doi.org/10.1007/BF00353453

Krickel, B. (2018). *The mechanical world - the metaphysical commitments of the new mechanistic approach*. Springer International Publishing.

Maung, H. H. (2016). Diagnosis and causal explanation in psychiatry. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *60*, 15–24. https://doi.org/10.1016/j.shpsc.2016.09.003

Melander, P. (1997). *Analyzing functions: An essay on a fundamental notion in biology*. Almqvist & Wiksell.

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630

Mele, A. R. (2003). Agents' abilities. *Noûs*, *37*(3), 447–470. https://doi.org/10.1111/1468-0068.00446

Millikan, R. G. (1984). *Language, thought and other biological categories: New foundations for realism.* The MIT Press.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, *56*, 288–302. https://doi.org/10.1086/289488

Murphy, D. (2020). Concepts of disease and health. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/health-disease/

Murphy, D., & Woolfolk, R. L. (2000). Conceptual analysis versus scientific understanding: An assessment of Wakefield's folk psychiatry. *Philosophy, Psychiatry, and Psychology*, *7*(4), 271–293.

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, *58*(2), 168–184. https://doi.org/10.1086/289610

Nordenfelt, L. (1995). *On the nature of health: An action-theoretic approach* (2nd revised and enlarged edition). Springer. https://doi.org/10.1007/978-94-011-0241-4 (Original work published 1987)

Parfit, D. (1989). *Reasons and persons.* Oxford University Press. https://doi.org/10.4324/9780429488450

Parfit, D. (2011). *On what matters* (Vol. 1). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199572809.001.0001

Prinz, J. J. (2004). *Gut reactions: A perceptual theory of the emotions.* Oxford University Press.

Rabenberg, M. (2014). Harm. *Ethics & Social Philosophy*, *8*(3), 1–32. https://doi.org/10.26556/jesp.v8i3.84

Raz, J. (2009). Reasons: Explanatory and normative. In C. Sandis (Ed.), *New essays on the explanation of action* (pp. 184–202). Palgrave-Macmillan. https://doi.org/10.2139/ssrn.999869

Scanlon, T. M. (1998). *What we owe to each other.* Harvard University Press. https://doi.org/10.1111/j.1933-1592.2003.tb00249.x

Shiffrin, S. (2012). Harm and its moral significance. *Legal Theory*, *18*(3), 357–398. https://doi.org/10.1017/S1352325212000080

Stalnaker, R. C. (2014). *Context.* Oxford University Press.

Stegenga, J. (2018). *Medical nihilism.* Oxford University Press.

Thomson, J. J. (2008). *Normativity.* Open Court.

Tremain, S. (2015). *Foucault and the government of disability.* University of Michigan Press.

Tsou, J. Y. (2021). *Philosophy of psychiatry.* Cambridge University Press.

Velleman, D. (2000). *The possibility of practical reason.* Oxford University Press.

Wakefield, J. C. (1992a). Disorder as harmful dysfunction: A conceptual critique of DSM-III-r's definition of mental disorder. *Psychological Review*, *99*(2), 232–247. https://doi.org/10.1037/0033-295X.99.2.232

Wakefield, J. C. (1992b). The concept of mental disorder. On the boundary between biological facts and social values. *American Psychologist*, *47*(3), 373–388. https://doi.org/10.1037//0003-066x.47.3.373

Wakefield, J. C. (1999a). Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology*, *108*, 374–399. https://doi.org/10.1037//0021-843x.108.3.374

Wakefield, J. C. (1999b). Mental disorder as a black box essentialist concept. *Journal of Abnormal Psychology*, *108*, 465–472. https://doi.org/10.1037//0021-843x.108.3.465

Wakefield, J. C. (2000). Spandrels, vestigial organs, and such: Reply to Murphy and Woolfolk's "The harmful dysfunction analysis of mental disorder." *Philosophy, Psychiatry, and Psychology*, *7*(4), 253–269. https://doi.org/10.1353/ppp.2000.0040

Wakefield, J. C. (2021). Is the harmful dysfunction analysis descriptive or stipulative, and is the HDA or BST the better naturalist account of dysfunction? Reply to Maël Lemoine. In L. Faucher & D. Forest (Eds.), *Defining mental disorder. Jerome Wakefield and his critics* (pp. 367–424). The MIT Press.

Wright, L. (1973). Functions. *Philosophical Review*, *82*(2), 139–168. https://doi.org/10.2307/2183766

Dembić, S. (2023). Mental disorder: An ability-based view. *Philosophy and the Mind Sciences*, *4*, 2. https://doi.org/10.33735/phimisci.2023.9630