

Oxford Handbooks Online

Naturalist Accounts of Mental Disorder

Elselijn Kingma

The Oxford Handbook of Philosophy and Psychiatry

Edited by K.W.M. Fulford, Martin Davies, Richard G.T. Gipps, George Graham, John Z. Sadler, Giovanni Stanghellini, and Tim Thornton

Print Publication Date: Jun
2013

Online Publication Date: Sep
2013

Subject: Psychology, History and Systems in Psychology,
Clinical Psychology

DOI: 10.1093/oxfordhb/9780199579563.013.0025

Abstract and Keywords

This chapter examines naturalistic accounts of mental disorder: accounts that define disorder as biological dysfunction. There are three such accounts: an eliminativist account (Szasz); a forward-looking or goal-contribution account (Boorse) and a backward-looking or evolutionary account (Wakefield). I argue first, and contra Szasz, that biological functions can be attributed at a mental level. But our mental architecture might simultaneously support many different ways of attributing function claims, which might undermine a strong naturalism about mental disorder. Second, I argue that Boorse's forward-looking account of disorder as dysfunction is not value-free. Third, I argue that Wakefield's backward-looking account does not accurately map onto our disorder judgments or medical purposes. I conclude that whilst Boorse's forward-looking account of disorder as dysfunction tracks medical usage better than Wakefield's evolutionary account, a truly value-free account of mental disorder cannot be sustained

Keywords: naturalism, disorder, Boorse, Wakefield, Szasz, function, dysfunction, evolution, goals, mental

Whether one is believed to have a mental disorder or not has consequences: it can give access to special and/or medical treatment as well as other social, economic, and emotional benefits, but it can also result in significant harms or risks of harm such as stigma, social exclusion, and infringement of rights. Less often mentioned, but perhaps more importantly, whether a person is thought to have a mental disorder affects how she and others view, interpret, respond to—and thereby partially form—who she is and what she does (Hacking 1995, 1998).

It seems crucial, then, to get our judgment on whether someone has a mental disorder right. This chapter shall examine part of that judgment: it shall ask what mental disorder *is*. There are at least three possible versions of that question. First, what makes a disorder a *mental* rather than a *physical* or *somatic* disorder?; second, what makes a *particular* mental disorder different from *another*?; third, what makes a condition an instance of *disorder*?¹ This essay focuses on the third question.²

The Demarcation Problem in Psychiatry

The question of what distinguishes disorder from health, the *demarcation problem*, sparks considerable controversy and is most hotly debated in the context of psychiatry. But it does (p. 364) not belong there exclusively: obesity, deafness, and high blood pressure have all posed demarcation problems in somatic medicine. Still, there are good reasons for worrying especially about mental health. First, mental disorders seem more controversial and elusive than somatic ones; most of us, including doctors and scientists, have a clearer idea of broken legs than broken personalities, and a better grasp on high blood pressure than low moods. Second, psychiatry seems more easily “polluted” by values and social considerations than somatic medicine—indeed the

idea that psychiatry is a mere “tool for social and political control,” as opposed to a scientifically or biomedically founded enterprise that has the interest of its patients at heart, is an enduring critique of psychiatry (see e.g., Foucault 1961; Laing 1959; Szasz 1960, 1972).

The history of psychiatry provides a wealth of examples that support this latter view. Masturbation, hysteria, and nymphomania were widely considered to be disorders in the nineteenth century, for example, as was homosexuality.³ A more outlandish example is “drapetomania,” the disorder supposedly exhibited by slaves who ran away from their master (Cartwright 1851; Engelhardt 1974). Viewed from our present perspective, these misdiagnoses result not just from an inferior scientific understanding of the natural world, but seem to indicate a conception or understanding of mental disorder that embodies contemporary—and morally dubious—social and evaluative standards rather than scientific ones: gender norms, sexual morality, and racism.

The worry that mental disorder constitutes social rather than biomedical deviance is not confined to the distant past either; our present-day expansions and increases in the prevalence of certain psychiatric diagnoses, such as depression and attention-deficit/hyperactivity disorder (ADHD), are still being challenged as *merely* reflecting social norms (about how long children should sit still, for example, or about how happy we should be) rather than scientific facts about *actual* disorder (see, e.g., Hawthorne 2007; Horwitz and Wakefield 2007). Indeed the definition of mental disorder was a heavily contested topic in the construction of the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), when precisely such worries played a role in the process that ultimately resulted in the exclusion of homosexuality from the DSM (Bayer 1987). The literature on the demarcation problem very much exists against the background of these particular worries, which also explain and motivate its two main and opposing positions: naturalism and normativism.⁴

Naturalism and Normativism

According to naturalism, the concepts of health and disorder are predominantly driven by objective natural categories, that is, categories that exist independent from our values and interests: *biological function* and *dysfunction*.⁵ Whilst these categories may interact with (p. 365) values, social considerations, and/or social norms to generate more complex judgments about what conditions qualify for particular social and medical treatment,⁶ and whilst values may also play a role in our identification of these categories (Murphy 2009), naturalists maintain that these values or social norms do not determine what disorder/dysfunction is; nature does. On the naturalist view, then, the history of psychiatry is purely one of mistaken science and illegitimate social influence. Indeed, to criticize past or present boundary shifts in our application of mental disorder as (merely) reflecting values or social norms is to assume already that a category of “real disorder” exists that is independent from those values and social norms.

Normativism, by contrast, must give an alternative account of the history of psychiatry; according to normativism, the demarcation between health and disorder is not a matter of nature alone.⁷ Normativists therefore tend to present past and present psychiatry not as illegitimately influenced by values, but as illustrative of the way in which shifting values and social norms have driven changes in mental disorder.⁸ Indeed, if normativism is correct, there is no non-evaluative notion of “real” disorder that a shifting diagnostic pattern of ADHD is a deviation from.

In the rest of this chapter I focus on naturalist accounts of mental disorder and examine its three most influential positions: Szasz’s rejection of mental disorder, Boorse’s biostatistical account of disease, and Wakefield’s evolutionary account of disorder. I conclude that the statistical concept of dysfunction offered by Boorse tracks medical usage more closely than the evolutionary account offered by Wakefield. But neither succeeds in offering a completely value-free account of disorder.

Szasz and Mental Functional Architecture

The modern debate on mental disorder has been deeply influenced by Thomas Szasz (1960), whose “The Myth of Mental Illness” put pressure on both the concept of mental disorder and psychiatry as a whole. Here is a version of Szasz’s argument against mental disorder:

P1. (Naturalist premise): what constitutes disorder is a dysfunction or lesion at a structural, cellular, or

molecular level.

P2. (Empirical premise): “mental disorders” present without such a physical lesion.⁹

C. (Conclusion): mental disorders do not exist.

(p. 366) Szasz argued, not that the things we call mental illnesses do not exist, but that they are not *disorders*—naturalistically conceived of by him as dysfunctions. Instead they are problems in living and/or departures from “psychosocial, ethical, and legal” norms. To apply the term “disorder” to these problems is to use the term metaphorically, which Szasz considered both illegitimate and obfuscating. Illegitimate because it is employed to excuse inexcusable behavior and justify unjustifiable measures, such as forced treatment and isolation. Obfuscating because “the concept functions as a disguise; for instead of calling attention to conflicting human needs, aspirations, and values, the notion of mental illness provides an amoral and impersonal ‘thing’ (an ‘illness’) as an explanation for problems in living” (Szasz 1960, p. 116).

Szasz’s criticisms map neatly onto the naturalist/normativist divide: he defends a *naturalist* concept of dysfunction as the only legitimate concept of disorder, then criticizes psychiatry for illegitimately employing a *normativist* conception of disorder as deviation from social norms. Subsequent normativists have mostly argued that Szasz was right to think that mental disorder is a non-naturalist concept, but wrong to think this makes it illegitimate. Naturalists, by contrast, agree with Szasz’s contention that dysfunction is the only legitimate construal of disorder, but tend to disagree with his views on mental disorder; where Szasz thought that naturalism undermined psychiatry, most naturalists attempt to *support* and *legitimize* (at least part of) psychiatry. For this latter project to succeed Szasz’s argument that the mind does not instantiate dysfunction has to be answered—and such an answer is provided by David Papineau (1994).

Against Szasz: Minds over brains

One way to argue that Szasz is wrong about mental disorders is to question his second premise: “mental illnesses present without a structural, chemical or molecular-level lesion.” It is conceivable that this premise is wrong, and that unbeknownst to us every instance of mental disorder is the result of some structural or physiochemical deviation in the brain/body. I believe, with Szasz, that this is unlikely, but settling that question is not relevant here: we can grant Szasz his second premise and *still* conclude, on the basis of his first premise, that his argument against the existence of mental disorders is unsound.

Szasz’s first premise supposes that biological dysfunctions can only manifest in physical lesions. But contemporary philosophy of mind gives us good reason to think that there could be natural mental dysfunctions that are not reducible to abnormalities in the structural and physiochemical make-up of the body/brain. Papineau helpfully uses the software/hardware analogy to illustrate this: a bug in a computer program need not be a problem in the hardware of your computer, because the physical substrate of your computer can run a multitude of software programs, including dysfunctional ones. Similarly a structurally and functionally intact *brain*—that is to say all neurons working as they should, and no neuronal or structural lesions are present—can *run* (instantiate) a variety of mental patterns, including dysfunctional ones (Papineau 1994).

By driving a wedge between the functional integrity of the brain and the functional integrity of the mind, Szasz’s challenge to the existence of mental disorders/dysfunctions can be answered. Subsequent naturalists have thus been freed to discuss the mind and its functions (p. 367) without reference to the brain, and can thus attempt to construct a naturalist account of mental disorder as mental dysfunction that could support psychiatry.

Mental dysfunction and functional architecture

Once we can posit the mind as a legitimate object of inquiry, we can proceed to uncover its functions. But that is only possible if the mind possesses a functional architecture. Now in somatic medicine, conceiving of the organism as having a functional architecture has been a successful approach. But when it comes to the mind, our understanding of mental functional architecture is far less developed, and the scientific project of carving the mind into traits that have functions falls far short of the success that its somatic counterparts enjoy. Compare, for example, physiological descriptions of the working of the ear with functional description of the *mental* component of auditory processing. Ear physiology includes detailed accounts of cells and processes that have particular functions, which have mostly achieved the status of scientific fact. In accounts of auditory processing, by contrast,

multiple and mutually exclusive models remain in contention.

Still, in absence of agreement on the functional structure of the mind, we can ask what the truths about mental functional architecture would have to be if we are to end up with a suitable account of disorder as dysfunction. I contend there are at least two such truths. First, the mind must be modular, with different modules having distinct functions. Second, this modular structure must not exist merely at a broad level, but at a level of considerable detail. Both, I suggest, are problematic, and let me start with the second point.

Although we can easily agree upon certain broad-level functions of the mind, such as “tracking the environment” or “social interaction,” a considerably more detailed account of mental functions is needed to support an account of mental illness as mental dysfunction. Take, for example, “tracking the environment.” Most of us are pretty bad at that; we suffer not only from a range of well-described systematic biases such as self-serving bias and attribution error, but are also subject to optical illusions and attentional deficits. An accurate account of mental function and dysfunction needs to be sensitive to this: it should explain, or at least accommodate, that sometimes our inability to track the environment is not a dysfunction, even though in other circumstances, in the case of hallucinations, for example, it is.

This is particularly important for psychiatry, because conceiving of mental functions broadly, as “successful social interaction” or “tracking the environment” means that any lack of social functioning or false belief becomes a disorder. And that would violate both our ideas about mental disorder and the core expectation that motivated the search for a naturalist account of disorder in the first place. This is the expectation that such account of disorder should help distinguish between false beliefs *per se*—of which we have many—and false beliefs that are, or indicate, mental dysfunctions; between impaired interaction that results from different values and ideas—all too common—and impaired interaction that is mental disorder; between problems that beset a healthy mind, and those that indicate something wrong. A broad description of mental function does not allow for that distinction, therefore a very detailed account is necessary if an account of mental disorder as dysfunction is to succeed.

(p. 368) Levels of detail aside, we should also question whether the mind has a modular functional structure at all. This is not an unpopular proposition: the modular theory of mind supports a considerable research program in cognitive psychology. This presents the mind as consisting of distinct but cooperating, hierarchically and sequentially organized “specific modules” that have functions, and non-specific or general modules that support other modules.¹⁰ Such a modular mind would be a good candidate for supporting an account of disorder as dysfunction. But it is not *obvious* that our minds should have such a structure. It may be the case that our minds operate in diffuse ways that do not lend themselves to easy carving into traits and functions, or that do not lend themselves to carving at all.¹¹ To use a well-trodden metaphor, minds may not have joints—and if they do, they may have so many that there are multiple possible yet mutually incompatible ways of carving them.

If such a pluralist stance to mental organization was justified, and multiple functional descriptions of the mind are true,¹² then that would pose problems for a naturalist account of mental disorder: multiple true functional descriptions of the mind means multiple true demarcations between function and dysfunction, which means multiple true accounts of mental disorder. And this (I argue in the next section) undermines naturalism about mental disorder.

Boorse and Requirements of Naturalism

The most widely known and well-discussed naturalist account of disorder is Christopher Boorse’s “biostatistical theory” (BST):

1. The *reference class* is a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species.
2. A *normal function* of a part or process within members of the reference class is a statistically typical contribution by it to their individual survival and reproduction.
3. A *disease* is a type of internal state that is either an impairment of normal functional ability, i.e., a reduction of one or more functional abilities below typical efficiency, or a limitation on functional ability caused by environmental agents.¹³
4. *Health* is the absence of disease.¹⁴

In other words, health is normal function, where normal function is the statistically typical contribution to survival and reproduction in a reference class. Disorder, on this account, is dysfunction, which is an adverse departure from normal function.

(p. 369) To illustrate, let us suppose that visual perception can be understood as the mind's interpreting of visual information to facilitate its tracking of environmental happenings. This is a statistically typical contribution the mind makes to survival and reproduction. The normal execution of this function is determined by what is statistically typical in my reference class. Therefore if I interpret visual information as accurately as most humans of my sex and age, I am healthy. Optical illusions cause me to interpret visual information in a manner that fails to track environmental happenings in the world correctly, but since that is statistically normal in my reference class, this is not a disorder. Hallucinations, by contrast, are interpretive failures to track the environment that are *not* statistically typical. They are thus an adverse departure from normal function, and a (form of) disorder on Boorse's account.

Boorse's account of disorder has attracted criticism proportional to its notoriety,¹⁵ and his account of biological *function* is subject to a separate and additional set of objections in the literature on philosophy of biology.¹⁶ In this chapter I want to focus on just one, so far unanswered, objection to Boorse, which does not question the accuracy of the BST, but contends that even *if* Boorse's account were to accurately depict disorder, it does not succeed in being completely interest-independent or value free.

Against naturalism: Reference classes

The BST performs statistical abstractions to determine what is normal, and therefore healthy, functional performance in humans. But normal functional performance varies tremendously within the human species. What counts as normal mental function in a two-year-old would not be normal in a twenty-year-old, for example, and normal hormone levels vary considerably amongst sexes, over a lifetime, and even during the day. The BST therefore employs *reference classes*, in which it calibrates statistics separately to determine what the normal function is. These reference classes divide the population by sex, age, and (for certain purposes only) ethnicity.

Reference classes are at the core of an objection against Boorse's account, that I defend in more detail elsewhere (Kingma 2007). Imagine different reference classes than the ones proposed by Boorse. If we allow such different reference classes, then different accounts of function, and of health and disorder, would emerge. For example, if people with depression or ADHD formed a separate reference class on the BST, then depression or ADHD would be statistically normal in these groups. These would then no longer be a disorder according to the BST. This move could be repeated for many conditions such that they become "healthy": all we need to do is create a BST-type account of function that admits a separate reference class for the condition in question. Note that these different accounts of functions would be just as value-free as the BST presented by Boorse, provided that the reference classes they admit are grounded in natural or value-free categories.

Presumably Boorse wants to maintain that the BST provides the correct analysis of function and health, and that alternative concepts generated by alternative reference classes do not. If so, he needs to provide a non-circular justification for why the reference classes he proposes are admissible, and alternative reference classes are ruled out. (p. 370) In other words, Boorse needs to justify, without prior reference to health and disorder, why out of all possible ways of groupings humans, only age, sex, and perhaps race are the groupings that underpin an account of health and disorder. And since Boorse is committed to arguing that the BST is *value-free*—i.e., that a preference for the BST over alternatives represents a *value-free* choice or objective representation of reality and not a value-laden preference—this justification for admitting certain reference classes only should be value-free too.

Such a non-circular, value-free justification cannot be provided: restricting reference classes to healthy groups of humans would be circular, and other proposals either fail to generate the correct results or appeal to values (Kingma 2007). We therefore lack a reason to think the reference classes Boorse admits, and thereby the BST, are value-free or interest-independent. Instead they are likely to reflect prior, and possibly value-laden, assumptions about which groups are normal and healthy—assumptions that are deeply embedded in this account of disorder.

The example of the "XST" illustrates the problem that the BST is in. Suppose the XST is an account of disorder that closely resembles the BST, with one exception: sexual orientation is a reference class on the XST. According to the BST homosexuality is a disorder, because non-heterosexual attraction is a worse-than-typical contribution to

reproduction.¹⁷ On the XST, however, homosexuality is normal in the reference class of people with homosexual orientation, and therefore healthy. In the absence of an objective justification of the kind that Boorse should but cannot provide, i.e., an objective justification for favoring the BST over the XST (or vice versa), we lack a way to settle the demarcation problem regarding homosexuality. And since XST-type accounts can be generated for many if not all “disorders,” the BST—even if it appears successful at delineating our concept of disorder—lacks the resources, let alone a value-free method, to withstand challenge posed by such accounts. The BST is therefore neither a value-free account, nor is it able to provide a value-free justification for our actions or a value-free method for settling controversy.

Naturalists’ burden of justification

The earlier argument has wider implications for naturalist accounts of disorder: it increases their burden of justification. Naturalists’ central claim is that disorder is not a reflection of social values or norms, but a feature of the natural world. Following Boorse, naturalists have attempted to defend this view by demonstrating that health and disorder can be defined in value-free terms, most notable in terms of *dysfunction*. But the earlier argument suggests that this is not enough; giving a definition or account of the concepts “health” and “disorder” in value-free terms does not suffice for proving that these concepts are completely value-free. As the example of the XST demonstrates, accounts stated in value-free terms can still embody deeply held social values. It follows that if naturalists are to defend their main claim, they have to meet an additional challenge: demonstrate that there is some value-free justification for employing the particular naturalistic concept they define, *rather than* another one.

(p. 371) One might object that that is too demanding. We cannot be expected to give a value-free justification for why we employ concepts in certain roles rather than other ones, and it is quite plausible that our ultimate reason for employing the concept “disorder” in its present role is a social or evaluative one: disordered conditions are worthy of attention as a group because they are painful, disabling, and so on.¹⁸ The challenge for naturalists is therefore slightly narrower: *given* that we are interested in certain conditions that are disabling, painful, and so on, naturalists have to provide a value-free justification for favoring one possible naturalist extension of these conditions over another. In other words, *given* that pneumonia, cancer, and arthritis are disorders, naturalists have to justify why the category “disorder” that they belong to is described by—say—the BST rather than the XST. It is this justification that Boorse cannot provide. Without it, accounts of disorder in value-free terms cannot on their own deflect the worry that the concept “disorder” might reflect underlying evaluative categorical judgments. Naturalists have considerable extra work to do.

Naturalism and pluralism

At this point, it is worth recalling a suggestion made in the “Mental dysfunction and functional architecture” section, which is that the structure of the mind may be such that a pluralist stance toward mental functional architecture may be justified. In that case—*whatever* the right account of function turns out to be—we end up with a problem very much similar to the one the BST faces: any particular account of disorder as dysfunction, whilst describing disorder in value-free terms, has to choose and employ a particular way of carving up functional architecture, rather than another one, to determine what are disorders. Such an account will therefore reflect the values that drove this choice, and—just like the BST—be unable to withstand challenges that can be framed in terms of employing a different method of carving up the mind. Pluralism, in this case, would put the view that disorder is a value-free concept on shaky grounds.

At the same time we should also not overstate the force of this argument. If naturalism were to modify its commitments and restate its aims, much of the problems outlined would disappear. Naturalism could admit that disorder was a value-laden category in the limited sense I have outlined, and restrict its aims to identifying and defining this category in clear factual terms so that questions can be settled by first-order appeal to facts without having to *explicitly* agree or disagree about values. This much weaker and pragmatic version of naturalism may be defensible, and it could possibly help us settle some disputes, conflicts, and boundary cases by allowing us to move straight from the facts about a condition to a judgment about its disordered status, without further appeal to values or norms. As the example of the XST illustrates, however, it would lack the resources to answer many challenges or resolve all problems.

(p. 372) Wakefield and Evolutionary Accounts of Disorder

The main naturalist alternative to Boorse's BST is Jerome Wakefield's "harmful dysfunction analysis" (HDA) of disorder. This account also defines disorder in terms of *dysfunction*, but gives an account of function that is explicitly grounded in evolutionary theory.¹⁹ Such an evolutionary account of disorder has two *prima facie* advantages. First, it might provide a response to the challenge posited in the previous section. For where Boorse had to construct unjustified and arbitrary reference classes to extract a particular account of function out of several possible closely related ones, an evolutionary account of function is based in a more substantive metaphysical claim: given a contingent evolutionary history, there exists a specific way in which biological systems should function. Thus *if* disorder is indeed (harmful) dysfunction, evolutionary history provides a non-arbitrary and naturalistic extension of the concept "dysfunction."²⁰ Second, the other apparent advantage of an evolutionary account of function is that it enjoys considerable support in the literature of philosophy of biology.²¹

Cashing out evolutionary dysfunction

The core idea behind the evolutionary account of disorder is that evolution, as a design-like process, can create and explain biological norms of functioning that are non-evaluative. But that idea needs development, for not every effect or trait produced by evolution is useful or has a function. Congenital disorders, for example, are just as much the product of evolution as the absence of them; disease agents are products of evolution too; and detailed evolutionary explanations have been posited for a variety of acquired diseases and/or disease susceptibilities.²² What the evolutionary disorder naturalist therefore needs is a distinction between those products of evolution that are functioning as designed, and other results of evolution that count as dysfunctions and disorders.

One way to generate this distinction is to contrast evolution by *natural selection* with other forms of evolutionary processes, such as drift.²³ On this proposal, functions are effects (p. 373) of traits that are a result of evolution by *natural selection*. But this fails to get at the desired distinction: first, *traits* that have spread in the population as the result of natural selection can include disorders. For example, sickle-cell anemia, a homozygous trait, has spread in the populations in which it exists as a consequence of natural selection for malaria-resistant heterozygous sickle-cell trait.²⁴ Second, *effects* that have spread in the population as a result of natural selection for underlying traits can also include diseases. Atherosclerosis, for example, is understood as an effect of the very good immune system that may have evolved by natural selection when humans lived in crowded cities in very unsanitary conditions.

A better interpretation of an evolutionary account of dysfunction/disorder does therefore not appeal to the traits or effects that are the *result* of evolution by natural selection but those that were the *drivers* of evolution by natural selection.²⁵ In the earlier examples, sickle-cell anemia may be the result of natural selection, but only malaria resistance drove natural selection by contributing to the inclusive fitness of ancestors. Therefore only malaria resistance is a function of heterozygous sickle-cell trait. Homozygous sickle-cell anemia, in contrast, has no function. Similarly atherosclerosis—although a result of natural selection—did not drive its selection because it did not contribute to the inclusive fitness of ancestors, and therefore has no function. Superior bacterial immunity, by contrast, does have a function, and is indeed what drove the natural selection of an aggressive immune system.

In keeping with this, Wakefield states: "The natural function [...] is not just any benefit or effect provided by a mechanism but a benefit or effect that explains, through evolutionary theory, why the mechanisms exists or has the form that it does."²⁶ This statement is too strict, however; in philosophy of biology a distinction is routinely drawn between exaptations and adaptations. A trait is an *adaptation* when it has an effect that enters into the explanation of the structure or form of that trait. Turtles' flippers, for example, have probably gotten their present shape because of their fitness-enhancing swimming effects—and are therefore adaptations for swimming. But turtles also use their flippers to bury eggs. This effect does not explain how flippers got their present form, although it may explain why flippers continue to exist. Flippers are therefore an *exaptation* for burying eggs (Gould 1991; Gould and Vrba 1982). Whether exaptations are legitimately called functions is "perhaps to be settled by one's taste for neologisms" (Allen 2009). But in the context of defining function for the purposes of defining disorder it is best to take a permissive account of selected effects that includes exaptations.²⁷ We might, then, replace Wakefield's original account with something like Neander's: "It is the/a proper function of an item (X) of an organism (O) to do that which items of X's (p. 374) 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."²⁸

One problem for evolutionary accounts of functions and disorder is that they depend heavily on the details of our evolutionary history. And we rarely, if ever, have access to those details. This is a problem at two levels. First, it means that if the etiological account is the correct account of disorder, it might not be of much use to the naturalist or society more broadly. For though this means that what health and disorder are is a matter of fact, we will never be in a position to access all those facts; we are lucky to have access to any.²⁹ The second problem is directly relevant to our present task: if our lack of epistemic access to selective histories makes it difficult to determine whether a condition is a disorder *if* etiological accounts are correct, how do we judge that this is the right approach to disorder in the first place?³⁰ This problem will crop up repeatedly in the rest of this chapter, but I will attempt to proceed in a way that is least affected by this.

Against Wakefield: “Selected disorders”

Like Boorse’s BST, Wakefield’s evolutionary account of disorder has been widely discussed.³¹ In this section I wish to focus on two main objections to his account, one of which Wakefield can absorb—at a price—and one that I believe he cannot.

The first possible problem for Wakefield is that of “*selected disorders*.” These are selected effects or strategies that have very negative effects in our present society. Possible examples include forms of antisocial behavior: rape, a violent disposition, or dependent or attention-seeking behavior. All of these may have been beneficial in selective terms: serial rape can be a good strategy for increasing one’s reproductive output, for example, and violence, dependence, or attention-seeking may all increase one’s access to resources that in turn increase fitness.³² But if it is true that these conditions have been selected, then they are not a disorder according to Wakefield—and this countervenes our current way of thinking about these conditions.

In response, Wakefield sticks to his guns: if “attention seeking” is an effect of a trait that drove the selection of that trait, then it cannot be a disorder—no matter how much we dislike it.³³ Indeed, he argues, there is historical precedent for this. When we discovered fever (p. 375) was both functional and naturally selected we stopped seeing it as a disorder (Wakefield 2000, p. 259).

Whether Wakefield’s response to these “selected disorders” is tenable depends, of course, on what conditions turn out to fall into the category outlined (which we cannot know without knowledge of the relevant evolutionary facts). But it also depends on our willingness to revise the DSM and our treatment attitudes in light of such facts. It seems possible that we would adjust our opinion of such conditions, as Wakefield predicts, but that is not inevitable; we could also continue to treat them as disorders, which would put pressure on the idea that defining a naturalistic concept of disorder has practical relevance (see also Bolton 2008).

Against Wakefield: “Non-selected effects”

The second and more damaging problem for Wakefield is that of *non-selected effects*. As we saw earlier, not all of our mental traits have been *selected* for effects that they themselves perform; in some cases the presence of a trait is explained by the effects of a different trait. For example, the presence of blue eyes is not explained by an effect of blue eyes, but by the increased ability of lighter skin to absorb ultraviolet B radiation (which helps with vitamin D production). This can happen because the trait “blue eyes” is *linked* to the trait “light skin.” Such linkage can happen in various ways: traits can be genetically linked, when their genes appear close together on chromosomes; they can be pleiotropically linked, when they result from genes that give rise to or are involved in the development of multiple traits; or they can be developmentally linked, when physical and other constraints on human development are such that the development or evolution of one trait cannot happen without giving rise to or changing another.³⁴

Such linked traits pose a serious problem for Wakefield’s account of disorder. As discussed, the selected effect account of function to which he is committed is very strict: it *only* ascribes a function to effect *e* of trait *t* if *e* was causally responsible for the selection of *t*.³⁵ Accordingly, traits that are not selected for their own effect, but are selected because of their linkage to other successful traits, do not have a function on Wakefield’s account.

An example will illustrate this problem, and how Wakefield proposes to deal with it. Take the human ability to learn how to read. If this ability were a function, the effects of reading would have to explain why we find this ability in

humans. But that is not borne out by the data. Humans developed script very recently even in non-evolutionary terms—a couple of thousand years BCE—and only in a few human subgroups in the Middle East, Mesoamerica, and perhaps China. Even within those groups only a small proportion of the population has been exposed to script until quite recently. Thus if the effects of reading had driven the natural selection of our ability to learn how to read, we would expect to find that ability (p. 376) only in small proportions of the aforementioned populations. Since, however, the ability to learn how to read is found universally around the globe, some effect other than reading must explain its selection.³⁶ Reading, then, is not a function on Wakefield's account. And this means that a problem in our ability to read, such as dyslexia, would not be a disorder.³⁷

Wakefield responds to this example by acknowledging that reading cannot be a function, but denying that this would stop dyslexia from being a disorder (Wakefield (1999a, pp. 382–383, 1999b, 2000)). When our ability to learn to read is compromised, he argues, it is because a mechanism underlying this ability is not working as it should. And since that mechanism must have been selected for something, it has a function that is compromised. Strictly speaking, then, dyslexia is not a dysfunction of our ability to learn to read, but a dysfunction of whatever effect it was that *did* drive the selection of our ability to learn how to read.³⁸

Problems for Wakefield's response

There are two problems with Wakefield's response.³⁹ First it depends heavily on the mistaken assumption that when the effect of a trait is selectively explained by an effect elsewhere, a failure to produce the former effect *must* indicate a dysfunction of the latter. But even though the selection of blue eyes is explained by the effects of fair skin, it is entirely possible for something to happen to the blueness of my eyes without my skin being affected. We therefore have no reason to accept this assumption. In fact the multiple ways in which traits can be linked, the developmental complexity of human organisms, and the length of the human lifespan all give active reason to reject it. The combination of these factors means that it is not just possible, but in fact overwhelmingly likely, that all manner of things—either later in life or during development—could affect one out of a pair of genetically or developmentally linked traits without affecting the other.

Second Wakefield's response seems terribly ad hoc. Of course it *may* be the case that our ability to read is produced by a mechanism that was selected for a particular effect, and that dyslexia indicates a breakdown of that that mechanism. But it is just as plausible that that mechanism is itself a by-product of the selection of a different, linked trait, and therefore (p. 377) lacking in function. Or that both the normal ability to learn to read and dyslexia are on a spectrum of normal variation in non-selected effects produced by a functioning underlying mechanism. And then there is a fourth option, which is that dyslexia itself is explained by its linkage to an adaptive trait, such as superior visual-spatial ability. In that case, dyslexia is indicative not of dysfunction, but of superior function. All four explanations are conjectures, of course, and I am not committed to any particular one of them. But Wakefield is: out of these four—on the face of it equally plausible and equally adaptive explanations for dyslexia—only one would make dyslexia a disorder. Wakefield, therefore, seems to be making a risky bet.⁴⁰

This second problem is more damaging than might seem at first. Remember that the discussion here is not really about dyslexia—which is just an example—but about whether Wakefield can provide a successful account of *all* mental disorders. Wakefield can bet against the odds in one case, dyslexia, and either win or lose. But if very many of our mental capacities are like reading—that is, effects of traits that do not themselves explain why those traits were selected, and that are therefore not functions—Wakefield's position starts to look more precarious. Now he is betting in a series, and at least some of those bets he must lose. Whatever the truth on dyslexia, then, Wakefield's account of disorder will end with a substantial group of conditions that it cannot label as disorders, even though they are according to our folk concepts and the DSM.

Here is one reason to suppose that more rather than fewer of our mental capacities will be like reading. Recall a point made at the beginning of the chapter, which is that the mind, if it is to allow for an account of disorder as dysfunction, must consist of distinct modules with distinct functions. That—I argued—was contentious on the face of it. But if a selected effect account is to bear out that mental modules have functions, an even more demanding picture emerges: modules should not only be distinct and have distinct effects, but those effects should also—in every single case—have been the drivers of the selection of those very modules. In other words, every single mental module or capacity must have been “visible to natural selection” via its own effect rather than through any of the other possibilities discussed: pleiotropic, developmental or genetic linkage, or other effects of that module.

Given the developmental complexity of our mind, that seems extremely unlikely. The moment that we allow, however, that mental modules are genetically, developmentally or pleiotropically linked, and that the selection of some traits and effects may be explained by effects elsewhere, Wakefield's account is already in a position where at least some apparent mental dysfunctions turn out not to be dysfunctions at all. Wakefield's account of disorder, it turns out, is very strongly revisionist.

Wakefield: More revisionist than he realizes

What should be done with the observation that Wakefield's account is so very revisionist? Bolton (2008) thinks it invalidates his account. Wakefield (1999a, 1999b, 2000) insists that it is not a large problem, and that where necessary we will adjust our ideas in the light of evolutionary facts. But Murphy and Woolfolk (2000b) point out—and I fully agree—that (p. 378) Wakefield is not really owning up to the degree of revisionism that his account is committed to. Wakefield has yet to acknowledge explicitly that his dyslexia bet is a slim one, or that he is committed to maintaining that dyslexia, schizophrenia, and a host of other disorders would not be disorders if it turns out that they are not, in fact, failures of a selected effects—which is a distinct possibility.

Wakefield in his later work (2000, pp. 244–245) explicitly commits to the narrow selected effect account of function that I have outlined, which means he is subject to the problems that accompany it. But the bulk of his defense remains devoted to insisting that his account would not be terribly revisionist because the evolutionary facts will bear out our commonsense judgments—which is implausible. What Wakefield is not defending is that his account of disorder would be correct even if these facts go radically against the grain of common sense judgments. And that is the far more likely situation.⁴¹

We cannot settle this stalemate without access to the relevant evolutionary facts. But we can say with confidence that there is something deeply inconsistent about Wakefield's claim that his account won't be so terribly revisionist. In making this claim Wakefield assumes adaptive explanation for traits we like, such as reading, but not for ones we do not like, such as dyslexia. If one is favorably disposed toward adaptive explanations in general, however—as Wakefield appears to be—then one should be willing at least in principle to expect them equally in instances of health and in instances of disorder, unless one has good reason not to. And Wakefield has not given us such good reasons.

What we do have, of course, are bad reasons to assume adaptive explanations for traits we like, where we do not assume them for traits we dislike. One such bad reason is the prior judgment that the traits we do not like are disorders. This justification is occasionally offered by Wakefield (1999, 2000), but it is a bad one because it makes Wakefield's core claim circular; Wakefield believes that a condition's being a *dysfunction* should determine our judgment that it is a disorder. That point becomes rather hollow if our judgment that something is a disorder becomes our main, if not only, reason for thinking that condition is a dysfunction.⁴²

A second bad reason is the usefulness or complexity of a trait, as the example of reading illustrates. It seems almost impossible that our ability to learn to read would not be designed: it is unique to humans, complicated, widespread, and incredibly useful, so how could it be a fluke? But it is a mistake to think that something is either selected, or a fluke. There is a good explanation for why we seem so well adapted to reading, but this explanation points not from script to our traits, but rather from our traits to script: the usefulness of reading does not explain why we developed the ability to read, but rather our ability to learn to read explains the discovery—or rather creation—of script and why it is useful to us. Therefore the fact that our traits seem beautifully adapted to what they do should not tempt us into thinking that they were selected for what they do.⁴³ Rather, in some cases, a trait doing things well is what explains why it was and continues to be put to the use to which it is suited.

(p. 379) In summary, then, we have only bad reasons to accept Wakefield's claim about the lack of revisionism implied by his account, and good reasons to reject it on both theoretical and empirical grounds, even if we are not in possession of *all* the empirical facts. It is overwhelmingly likely that we have an abundance of traits that fulfill important roles for us and in our culture, particularly in the mental realm, but whose effects may not be what drove their selection. These traits therefore lack functions and, by consequence, the ability to dysfunction.

And this leads us to the final and most damaging aspect of the non-selected effects objection, which has not been raised so far. I take it that most if not all of our physical, and the vast majority of our mental traits, fall within the

domain of health and disorder. That is to say, they are either disordered, and if not, they are healthy.⁴⁴ But an evolutionary account of disorder can never bear this out. As argued earlier, on such an account only those effects with the right evolutionary causal role can function and dysfunction. All other effects—that is, all effects of traits that did not drive the selection of those traits, and all effects produced by traits that were not selected for their own effects—neither have a function nor can they dysfunction. A consequence of this is that the revisionism of Wakefield's account lies not just in its labeling certain common-sense disorders as non-disorders. It also lies in the creation of an account of health and disorder that applies only to a subset of our traits. For whatever the precise evolutionary facts, Wakefield's account of disorder places a substantial portion of our physiological and mental traits out of the realm of health and disorder altogether. And that seems a clear violation of one core conceptual element of the health and disorder dichotomy, which is that all of our physical and mental traits seem to fall within them.

Naturalist Accounts of Disorder: Conclusions

Naturalism about disorder consists of two claims that should be kept separate. First, the claim that disorder can be defined at least in part in terms of biological dysfunction. Second, the claim that disorder and dysfunction are value-free concepts.

With respect to the former claim, I have examined two dominant accounts of dysfunction in philosophy of medicine and philosophy of biology: the causal role account defended by Boorse and the etiological account defended by Wakefield. A careful examination of the latter reveals it is highly revisionist in two ways. First, it relabels many conditions we now consider disorders as non-disorders. Second, and more problematically, it places many conditions—both normal and uncommon—out of the realm of disorder altogether. That latter revision is particularly hard to stomach as it seems to be a departure not merely from the extension of our current concepts of health and disorder, but from their very structure. In terms of tracking our actual concept of disorder, then, something like the forward-looking statistical causal role account offered by Boorse appears superior to Wakefield's.

(p. 380) One could decide to stick with Wakefield's account and accept its revisionism nevertheless. But I strongly suggest that this is undesirable. Medicine and psychiatry are pragmatic disciplines; they are primarily interested in the effects of our bodies and minds that are useful to us in the here and now, not in how those effects were selected under different circumstances in the past. Being able to learn how to read, for example, is an important function in modern life. But whether it is even included in the realm of functions, health and medicine, in Wakefield's account, depends entirely and crucially on the precise details of its causal role within a contingent evolutionary history. That does not do justice to how we want our concepts of health and disorder, and the disciplines that use them, to operate. If we are going to accept Wakefield's highly revisionist proposals we need a good defense of his revisionism first. Pretending that it does not exist or is marginal, as Wakefield does now, is not good enough.

With respect to the second claim I argued that it should be rejected; even if health and disorder can be defined in terms of function and dysfunction, there is no reason to think that this makes these concepts value-free. The reason is that a concept definable in value-free terms can still embody the values that caused us to classify the natural world using this concept in the first place, rather than an alternative one. This argument only gains serious traction, however, if the natural world is such that alternatives are available. I have given several reasons to think that this is the case. First, the core role of reference classes in Boorse's account of health means that naturalist alternatives can readily be constructed. Second, any kind of pluralism about detailed descriptions of mental functional architecture would afford alternative classifications of mental disorder, and thereby undermine the second naturalist claim.

I conclude that our concepts of health and disorder are not value-free. Naturalism in that sense fails. But a weaker form of naturalism can survive. This weaker naturalism offers an account of disorder in value-free terms but does not claim that this makes the concept of disorder value-free. Such an account might be helpful in settling some controversy and in clarifying our thinking. But it will be based in the forward-looking model created by Boorse, not the backward-looking one by Wakefield.

References

- Agich, G. J. (1983). Disease and value: a rejection of the value-neutrality thesis. *Theoretical Medicine*, 4, 27–41.
- Allen, C. (2009). Teleological notions in biology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition). [Online.] Available at: <>.
- Ananth, M. (2008). *In Defense of an Evolutionary Concept of Health: Nature, Norms and Human Biology*. Aldershot: Ashgate.
- Ariew, A., Cummins, R., and Perlman, M. (2002). *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press.
- Bayer, R. (1987). *Homosexuality and American Psychiatry: The Politics of Diagnosis*. Princeton, NJ: Princeton University Press.
- Bechtel, W. and Abrahamsen, A. (1990). *Connectionism and the Mind: And Introduction to Parallel Processing in Networks*. Cambridge, MA: Blackwell Publishing.
- Boorse, C. (1975). On the distinction between disease and illness. *Philosophy and Public Affairs*, 5, 49–68.
- Boorse, C. (1976a). Wright on functions. *Philosophical Review*, 85, 70–86.
- Boorse, C. (1976b). What a theory of mental health should be. *Journal for the Theory of Social Behaviour*, 6, 61–84.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44, 542–73.
- Boorse, C. (1987). Concepts of health. In D. Van de Veer and T. Regan (Eds), *Health Care Ethics: An Introduction*, pp. 359–93. Philadelphia, PA: Temple University Press.
- Boorse, C. (1997). A rebuttal on health. In J. M. Humber and R. F. Almeder (Eds), *What is Disease?*, pp. 1–134. Totowa, NJ: Humana Press.
- Boorse, C. (2002). A rebuttal on functions. In A. Ariew, R. Cummins, and M. Perlman (Eds), *Functions: New Essays in the Philosophy of Psychology and Biology*, pp. 7–32. Oxford: Oxford University Press.
- Boorse, C. (2011). Concepts of health. In F. Gifford (Ed.), *Philosophy of Medicine*, pp. 13–64. Oxford: Elsevier.
- Bolton, D. (2008). *What is Mental Disorder? An Essay in Philosophy, Science and Values*. Oxford: Oxford University Press.
- Brandon, R. N. (2005). The difference between selection and drift: a reply to Millstein. *Biology and Philosophy*, 20, 153–70.
- Brandon, R. N. (2006). The principle of drift: biology's first law. *Journal of Philosophy*, 103, 319–35.
- Cartwright, S. A. (1851). Report on the diseases and physical peculiarities of the negro race. *The New Orleans Medical and Surgical Journal*, 7, 707–9.
- Clouser, K. D., Culver, C. M. and Gert, B. (1981). Malady: a new treatment of disease. *The Hastings Center Report*, 11, 29–37.
- Clouser, K. D., Culver, C. M. and Gert, B. (1997). Malady. In J. M. Humber and R. F. Almeder (Eds), *What is Disease?*, pp. 175–219. Totowa, NJ: Humana Press.
- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 33, 263–82.
- Cooper, R. (2005). *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*. Dordrecht: Springer.

Naturalist Accounts of Mental Disorder

- Cosmides, L. and Tooby, J. (1987). From evolution to behavior: evolutionary psychology as the missing link. In J. Dupre (Ed.), *The Latest on The Best: Essays on Evolution and Optimality*, pp. 277–306. Cambridge, MA: MIT Press.
- Cosmides, L. and Tooby, J. (1999). Toward an evolutionary taxonomy of treatable disorders. *Journal of Abnormal Psychology*, 108, 453–64.
- Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.
- Engelhardt, H. T. (1974). The disease of masturbation: values and the concept of disease. *Bulletin of the History of Medicine*, 48, 234–48.
- Engelhardt, H. T. (1975). The concepts of health and disease. In H. T. Engelhardt and S. F. Spicker (Eds), *Evaluation and Explanation in the Biomedical Sciences*, pp. 125–41. Dordrecht: Reidel.
- Engelhardt, H. T. (1976). Ideology and etiology. *Journal of Medical Philosophy*, 1, 256–68.
- Engelhardt, H. T. (1986). *The Foundations of Bioethics*. New York, NY: Oxford University Press.
- Ereshefsky, M. (2001). *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge: Cambridge University Press.
- Foucault, M. (1961). *Madness and Civilisation: A History of Insanity in the Age of Reason* (R. Howard, Trans.). London: Tavistock.
- Fulford, K. W. M. (1999). Nine variations and a coda on the theme of an evolutionary definition of dysfunction. *Journal of Abnormal Psychology*, 108, 412–20.
- Garson, J. (2010). Connectionism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition). [Online.] Available at: <>.
- Gluckman, P., Beedle, A., and Hanson, M. (2009). *Principles of Evolutionary Medicine*. New York, NY: Oxford University Press.
- Godfrey-Smith, P. (1993). Functions: Consensus without unity. *Pacific Philosophical Quarterly*, 74, 196–208.
- Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, 28, 344–62.
- Goosens, W. (1980). Values, health and medicine. *Philosophy of Science*, 47, 100–15.
- Gould, S. J. (1991). Exaptation: a crucial tool for an evolutionary psychology. *Journal of Social Issues*, 47, 43–65.
- Gould, S. J. and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B*, 205, 581–9.
- Gould, S. J. and Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, 8, 4–15.
- Griffiths, P. E. (1993). Functional analysis and proper functions. *The British Journal for the Philosophy of Science*, 44, 409–22.
- Guerrero, J. D. (2010). On a naturalist theory of health: a critique. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, 41, 272–8.
- Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, and A. J. Premack (Eds), *Causal Cognition: A Multi Disciplinary Debate*, pp. 351–83. Oxford: Clarendon Press.
- Hacking, I. (1998). *Mad Travelers: Reflections on the Reality of Transient Mental Illnesses*. Charlottesville, VA: University Press of Virginia.
- Hausman, D. (2011). Is an overdose of paracetamol bad for one's health? *British Journal for the Philosophy of Science*, 62, 657–68.

Naturalist Accounts of Mental Disorder

- Hawthorne, S. (2007). ADHD drugs: values that drive the debates and decisions. *Medicine, Health Care and Philosophy*, 10, 129–40.
- Horwitz, A. V. and Wakefield, J. C. (2007). *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford: Oxford University Press.
- Houston, A. I. (2009). San Marco and evolutionary biology. *Biology and Philosophy*, 24, 215–30.
- Kass, L. R. (1975). Regarding the end of medicine and the pursuit of health. *The Public Interest*, 40, 11–42.
- Kendell, R. (1975). The concept of disease and its implications for psychiatry. *British Journal of Psychiatry*, 127, 305–15.
- Kingma, E. (2007). What is it to be healthy? *Analysis*, 67, 128–33.
- Kingma, E. (2010). Paracetamol, poison and polio: why Boorse's account of function fails to distinguish health and disease. *British Journal for the Philosophy of Science*, 61, 241–64.
- Kingma, E. (2012). Health and disease: Social constructivism as a combination of naturalism and normativism. In H. Carel and R. Cooper (Eds), *Health, Illness and Disease: Philosophical Essays*, pp. 37–56. Durham: Acumen Publishing.
- Kirmayer, L. J. and Young, A. (1999). Culture and context in the evolutionary concept of mental disorder. *Journal of Abnormal Psychology*, 108, 446–52.
- Kitcher, P. (1982). Genes. *British Journal for the Philosophy of Science*, 33, 337–59.
- Kitcher, P. (1984). Species. *Philosophy of Science*, 51, 308–33.
- Kitcher, P. (1996). *The Lives to Come: The Genetic Revolution and Human Possibilities*. New York, NY: Touchstone.
- Klein, D. F. (1999). Harmful dysfunction, disorder, disease, illness, and evolution. *Journal of Abnormal Psychology*, 108, 421–9.
- Laing, R. D. (1959). *The Divided Self*. London: Tavistock.
- Lewens, T. (2004). *Organisms and artifacts*. Cambridge, MA: MIT Press.
- Lilienfeld, S. O. and Marino, L. (1995). Mental disorder as a Roschian concept: a critique of Wakefield's "harmful dysfunction" analysis. *Journal of Abnormal Psychology*, 104, 411–20.
- Lilienfeld, S. O. and Marino, L. (1999). Essentialism revisited: Evolutionary theory and the concept of mental disorder. *Journal of Abnormal Psychology*, 108, 400–11.
- Margolis, J. (1976). The concept of disease. *The Journal of Medicine and Philosophy*, 1, 238–55.
- Matthen, M. and Ariew, A. (2002). Two ways of thinking about natural selection. *Journal of Philosophy*, 49, 55–83.
- McKenna, J., Trevathan, W., and Smith, E. O. (2008). *Evolutionary Medicine and Health: New Perspectives*. Oxford: Oxford University Press.
- McNally, R. J. (1994). *Panic Disorder: A Conceptual Analysis*. New York, NY: Guilford.
- Millikan, R. G. (1984). *Language, Truth and Other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, 56, 288–302.
- Murphy, D. (2006). *Psychiatry in the Scientific Imagine*. Cambridge, MA: MIT Press.
- Murphy, D. (2009). Concepts of disease and health. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2009 edition). [Online.] Available at: <>.

- Murphy, D. and Stich, S. (2000). Darwin in the madhouse: evolutionary psychology and the classification of mental disorders. In P. Carruthers and A. Chamberlain (Eds), *Evolution and the Human Mind: Modularity, Language and Meta-Cognition*, pp. 62–92. Cambridge: Cambridge University Press.
- Murphy, D. and Woolfolk, R. L. (2000a). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, 7, 241–52.
- Murphy, D. and Woolfolk, R. L. (2000b). Conceptual analysis versus scientific understanding: an assessment of Wakefield's folk psychiatry. *Philosophy, Psychiatry, & Psychology*, 7, 271–93.
- Neander, K. (1991a). Functions as selected effects: the conceptual analyst's defense. *Philosophy of Science*, 58, 168–84.
- Neander, K. (1991b). The teleological notion of 'function'. *Australasian Journal of Philosophy*, 69, 454–68.
- Neander, K. (1995). Misrepresenting and malfunctioning. *Philosophical Studies*, 79, 109–41.
- Nordenfelt, L. (1987). *On the Nature of Health: An Action-Theoretic Approach*. Dordrecht: Reidel.
- Nordenfelt, L. (2001). *Health, Science and Ordinary Language*. Amsterdam: Rodopi.
- Nordenfelt, L. (2007). The concepts of health and illness revisited. *Medicine, Health Care, and Philosophy*, 10, 5–10.
- Papineau, D. (1994). Mental disorder, illness and biological dysfunction. In A. Griffiths (Ed.), *Philosophy, Psychology and Psychiatry. Royal Institute of Philosophy Supplement 37*, pp. 73–84. Cambridge University Press.
- Perlman, M. (2004). The modern philosophical resurrection of teleology. *The Monist*, 87, 3–51.
- Reznek, L. (1987). *The Nature of Disease*. London: Routledge.
- Richters, J. E. and Hinshaw, S. P. (1999). The abduction of disorder in psychiatry. *Journal of Abnormal Psychology*, 108, 438–45.
- Sadler, J. Z. (1999). Horsefeathers: a commentary on 'evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology*, 108, 433–7.
- Sadler, J. Z. and Agich, G. J. (1995). Diseases, functions, values, and psychiatric classification. *Philosophy, Psychiatry, & Psychology*, 2, 219–31.
- Scadding, J. G. (1988). Health and disease: what can medicine do for philosophy? *Journal of Medical Ethics*, 14, 118–24.
- Scadding, J. G. (1990). The semantic problem of psychiatry. *Psychological Medicine*, 20, 243–8.
- Schramme, T. (2007). A qualified defence of a naturalist theory of health. *Medicine, Health Care, and Philosophy*, 10, 11–17.
- Simons, J. (2007). Beyond naturalism and normativism: reconceiving the 'disease' debate. *Philosophical Papers*, 36, 343–70.
- Sober, E. (1980). Evolution, population thinking and essentialism. *Philosophy of Science*, 47, 350–83.
- Sober, E. (2004). *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press.
- Szasz, T. S. (1960). The myth of mental illness. *American Psychologist*, 15, 113–18.
- Szasz, T. S. (1972). *The Myth of Mental Illness*. London: Paladin.
- Tooby, J. and Cosmides, L. (1990a). On the universality of human nature and the uniqueness of the individual: the role of genetics and adaptation. *Journal of Personality*, 58, 17–67.

Naturalist Accounts of Mental Disorder

Tooby, J. and Cosmides, L. (1990b). The past explain the present: emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, 11, 375–424.

Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, and J. Tooby (Eds), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pp. 19–136. Oxford: 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, 232–47.

Wakefield, J. C. (1992b). The concept of medical disorder: on the boundary between biological facts and social values. *American Psychologist*, 47, 373–88.

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

Wakefield, J. C. (1999a). Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology*, 108, 374–99.

Wakefield, J. C. (1999b). Mental disorder as a black box essentialist concept. *Journal of Abnormal Psychology*, 108, 465–72.

Wakefield, J. C. (2000). Spandrels, vestigial organs, and such: reply to Murphy & Woolfolk's "the harmful dysfunction analysis of mental disorder." *Philosophy, Psychiatry, & Psychology*, 7, 253–69.

Walsh, D., Lewens, T., and Ariew, A. (2002). Trials of life, natural selection, and random drift. *Philosophy of Science*, 69, 452–73.

Whitbeck, C. (1978). Four basic concepts of medical science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1, 210–22.

Williams, G. and Nesse, R. M. (1996). *Why We Get Sick: The New Science of Darwinian Medicine*. New York, NY: Vintage Books.

Notes:

(1) Because the contrast is between *disorder* and *health*, I take the former to denote all departures from health. That is not entirely consistent with either medical or ordinary language use, but in keeping with an established tradition in the literature (e.g., Boorse 1975; Cooper 2002). Boorse and Cooper use the term "disease," Clouser et al. (1981, 1997) speak of "malady," Wakefield (1992b) of "disorder," and Boorse in his later work (1997, 2011) prefers "pathological condition." All intend the same inclusive denotation.

(2) See Murphy (2006) for a recent treatment that touches on all three questions.

(3) Boorse (2011, pp. 13–14) lists many examples and sources.

(4) The terms "naturalism" and "normativism" are most widely used, but Murphy (2009), following Kitcher (1996), contrasts "objectivism" and "constructivism."

(5) See Boorse (1975, 1976b, 1977, 1987, 1997, 2011), Kass (1975), Kendell (1975), Scadding (1988, 1990), Schramme (2007), Szasz (1960) and most recently Ananth (2008). The disease-as-dysfunction approach is defended in most detail by Boorse (1977, 1997) and Wakefield (1992b).

(6) Boorse (1975, pp. 54–55, 60, 1977, p. 544, 1997, pp. 11, 12–13, 55, 95–99). See also Wakefield (1992a).

(7) See Agich (1983), Clouser et al. (1981, 1997), Cooper (2002), Engelhardt (1976, 1986), Goossens (1980), Margolis (1976), Nordenfelt (1987, 2001, 2007), Reznek (1987), and Whitbeck (1978). See Simons (2007) for an overview and comparison of different normativist positions, which are far more heterogeneous than is generally presented.

- (8) This position is certainly found in, e.g., Foucault (1961), but it is also implicit in, e.g., Cooper's (2002) definition of disease.
- (9) Some mental disorders, such as neuropsychiatric conditions, do present with physical lesions of course—and presumably these would have to be recognized as disorders by Szasz. It should be remembered that he wrote his essay in the 1960s, on the basis of contemporary knowledge.
- (10) See Cosmides and Tooby (1987), Murphy and Stich (2000), and Tooby and Cosmides (1990a, 1990b, 1992) for more detail.
- (11) Such a conception of the mind may be supported by another research program in cognitive psychology: connectionism. See, e.g., Bechtel and Abrahamsen (1990) and Garson (2010).
- (12) And pluralism already is a well-respected position in philosophy of biology when it comes to both species (Dupré 1993; Ereshefsky 2001; Kitcher 1984) and functions (e.g., Godfrey-Smith 1993; Griffiths 1993; Perlman, 2004).
- (13) The clause on environmental agents is controversial. It appeared in Boorse's (1977, p. 567) original proposal, but in 2011 (p. 27) following discussion in 1997 he proposed to drop it. For an objection to Boorse that is based on environmental agents see Kingma (2010); Hausman (2011) responds.
- (14) Boorse (1997, pp. 7–8). See also Boorse (1977, pp. 562, 567).
- (15) Boorse (1997) provides a near-complete treatment of objections to that date. He summarizes several more in 2011 (pp. 29–32). For subsequent objections see Kingma (2010) and Guerrero (2010).
- (16) See Lewens (2004) for an overview, and Boorse (1976a, 2002) for a response to some objections.
- (17) Although Boorse grants that homosexuality is a pathology on the BST, he maintains that it is not a bad thing and therefore should not be treated (1975, p. 63, 1997, p. 99). See also footnote 22.
- (18) This latter observation is often employed as an argument by normativists in order to show that health and disease must be value-laden categories (see, e.g., Cooper 2002, p. 271, 2005, p. 22; Engelhardt 1975, pp. 127, 136, 1976, p. 226; Goosens 1980, p. 106; Margolis 1976, p. 242; Nordenfelt 2007a). I argue elsewhere that the normativist conclusion does not follow (Kingma 2012; see also Schramme 2007).
- (19) There is another difference between Wakefield and Boorse: Wakefield defines disorder as *harmful* dysfunction, whereas Boorse defines disease as dysfunction per se. But since Boorse agrees with Wakefield that only harmful dysfunctions are the kind of conditions that we should treat, this difference is merely semantic, and has no practical or theoretical implications (Boorse 1975, p. 54–55, 60, 1977, p. 544, 1997, p. 11, 12–13, 55, 95–99).
- (20) Though it may still lack the resources to define the boundary between normal function and dysfunction (Schwartz 2007). See also Lilienfeld and Mario (1999, p. 404) and, for a response, Wakefield (1999).
- (21) See Godfrey-Smith (1993, 1994), Griffiths (1993), Millikan (1984, 1989), and Neander (1991a, 1991b, 1995) for a development and defense of evolutionary accounts of biological function, and Lewens (2004) and Ariew et al. (2002) for overviews.
- (22) See, e.g., Williams and Nesse (1996), Gluckman et al. (2009), McKenna et al. (2008).
- (23) Simply put, drift happens when the outcomes of selection are different from what one would expect based on the fitness of the traits involved. This can happen, amongst others, because of a founder effect in a small population; a freak accident that killed all fitter types; or the inherent variability involved in probabilistic sampling. See, e.g., Brandon (2005, 2006), Matthen and Ariew (2002), and Walsh et al. (2002).
- (24) See also Wakefield (1999, p. 389).
- (25) Sober (1980, 2004) calls the former “selection of” and the latter “selection for.”

- (26) Wakefield (1992a, p. 236). For similar definitions see Wakefield (1992b, pp. 382, 384, 1995).
- (27) Wakefield (1999, pp. 380–381) in response to Lilienfeld and Marino (1995, p. 412) and McNally (1994). See also Murphy and Woolfolk (2000a). On the same pages Wakefield makes clear that “maintenance selection” should also be subsumed under natural selection on his account. In maintenance selection a trait does not increase in frequency in the population, but is maintained at present frequencies by natural selection—for example because new mutations are being selected against. See, e.g., Godfrey-Smith (1994) and Griffiths (1993).
- (28) Neander (1991a, p. 174). See also Godfrey-Smith (1993, 1994), Griffiths (1993), Millikan (1984, 1989), and Neander (1991b, 1995).
- (29) Bolton (2008) argues that if an evolutionary account of disorder were correct, it would put the DSM in an epistemic situation barely different from the one it is at present.
- (30) This problem is very evident in discussions between Murphy and Woolfolk (2000a, 2000b) and Wakefield (2000) that very much revolve around competing “just-so” stories about evolutionary histories.
- (31) See, e.g., Cosmides and Tooby (1999), Fulford (1999), Klein (1999), Kirmayer and Young (1999), Lilienfeld and Marino (1995, 1999), Richters and Hinshaw (1999), Sadler (1999), Sadler and Agich (1995).
- (32) These examples and others can be found in Bolton (2008), Cooper (2002), Lilienfeld and Marino (1995), Murphy and Stich (2000), Murphy and Woolfolk (2000a), and Richters and Hinshaw (1999) all of whom discuss a version of the problem of “selected disorders.”
- (33) See, e.g., Wakefield (1999a, 1999b, pp. 466–467, 2000). See also Murphy and Stich (2000) who attempt a detailed classification of the different kind of problems that can beset an “evolve mind,” including a further distinction between selected traits that are disliked and presently maladaptive, and selected traits that are disliked but not presently maladaptive.
- (34) All of this is an oversimplification of course; the development of organisms, the relations between traits and genes, and indeed the very notion of a gene itself, are all highly complex (see, e.g., Kitcher 1982).
- (35) Even on the permissible interpretation outlined here that includes exaptations and recent selection. See also footnote 27.
- (36) See also Gould (1991) and Gould and Lewontin (1979).
- (37) Lilienfeld and Marino (1995) first raised this as an objection to Wakefield.
- (38) Wakefield (1999b, p. 466, 2000, pp. 255–256). Wakefield does not think that all disabilities to read are dysfunctions; only those that are due to a hypothesized underlying dysfunction. Illiteracy due to lack of schooling, for example, is not a dysfunction on his account. To avoid potential confusion I have been talking about the *ability* to learn to read.
- (39) Note that Wakefield in his responses talks about spandrels rather than non-selected effects. Spandrels are the decorated triangles between the skeletal support-structures in churches, which Gould and Lewontin (1979) famously claim are the necessary side effect of churches having roofs. I prefer to speak of non-selected effects for two reasons. First, because it is broader and more in line with the literature in philosophy of biology: not every non-selected effect is a spandrel, even if every spandrel is a non-selected effect. In fact it is the very spandrel analogy that leads Wakefield to suppose that a failure in the non-selected effect must indicate a failure in the selected mechanism producing it (Wakefield 2000, pp. 244–245). This may be true for actual spandrels, if they *have* to exist if a church has a roof, but I argue in the upcoming paragraph that this is not the case for every non-selected effect. Second, Gould and Lewontin were wrong about spandrels, even if they weren’t wrong about adaptationism (e.g., Houston 2009).
- (40) See also Murphy and Woolfolk (2000a).
- (41) See Murphy and Woolfolk (2000b) for a more detailed argument that Wakefield seems to hold evolutionary science hostage to folk concepts of mental disorder and psychiatric practice, and seems not truly committed to the

idea that science may prove our folk concepts to be radically wrong—which science does regularly.

(42) See also Murphy and Woolfolk (2000b) and Bolton (2008).

(43) Famously illustrated by Gould and Lewontin (1979).

(44) Indeed this is the very assumption that underlies most of the interest in and literature on division between health and disease. See also footnote 1.

Eselijn Kingma

Eselijn Kingma, Department of History and Philosophy of Science, Cambridge University, Cambridge, UK

