Introduction

Denis Forest and Luc Faucher

Jerome Wakefield's work is at the center of the contemporary debate as to the nature of mental illness (and the related question of psychiatry's scope and limits), a decades-old debate in both scientific and philosophical literature. His key proposal, the "harmful dysfunction analysis" of mental disorders (HDA thereafter), has been discussed at great length by scientists and philosophers alike. In psychology, discussions of Wakefield's proposal abound in special issues of journals (see, e.g., *Journal of Abnormal Psychology* [1999] and *World Psychiatry* [2007]), but although philosophers have commented on and criticized Wakefield's position on many occasions (see, e.g., Nordenfelt 2003; Bolton 2008; Gold and Kirmayer 2007; Murphy and Woolfolk 2000; Murphy 2006), no book or special issue of a major philosophy journal has been dedicated to the task of offering a survey of these critiques.

With this volume, we propose to remedy that situation, and for the occasion, we have gathered together some of today's most eminent and up-and-coming philosophers of psychiatry to discuss Wakefield's position as well as its theoretical implications and empirical consequences. We hope that the resulting collection of chapters—with extensive replies from Wakefield himself—may be of interest to researchers and students in several related fields ranging from clinical psychiatry to social work, as well as philosophy of mind and philosophy of psychiatry.

HDA: A Presentation

HDA is the claim that "a disorder is a harmful failure of some internal mechanism(s) to perform a naturally selected ('designed') function" (Wakefield 2000, 253). This notion was originally presented by Wakefield in two papers published during the same year (Wakefield 1992a, 1992b). At first sight, each of these papers is quite different: one is a general presentation of HDA, contrasting it with rival conceptions of mental disorders. The other is a critique of the definition of mental disorders as "unexpectable distress or disability" that is used in the revised third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-III-R*) (published in 1987). In fact, these two articles

offer two different perspectives on the implications of HDA: one is more philosophically oriented and deals with foundational issues; the other is more of a dialogue with medical research and practice and deals with the empirical consequences of theoretical choices, a type of research that the majority of Wakefield's subsequent publications can be grouped into (e.g., the two books coauthored with Allan H. Horwitz; Horwitz and Wakefield 2007, 2012). Since 1992, Wakefield has vindicated his thesis on many occasions, without revising it significantly. Critiques of HDA have tended to focus keenly on the terms "dysfunction" and "harmful," but "analysis" is no less important to understand the nature of his project. HDA is offered as a definition of what a mental disorder is, but it is also the outcome of the application of a method, the method of *conceptual analysis*, and it would be an error to separate the two.

Wakefield characterizes conceptual analysis in the following manner: "In a conceptual analysis, proposed accounts of a concept are tested against relatively uncontroversial and widely shared judgments about what does and does not fall under the concept. To the degree that the analysis explains these uncontroversial judgments, it is considered confirmed, and a sufficiently confirmed analysis may then be used as a guide in thinking about more controversial cases" (Wakefield 1992b, 233). Conceptual analysis is a tool that allows one to judge the merits of competing accounts of what a mental disorder is, HDA being one of the latter. These merits can be evaluated using two criteria. One is that a proper analysis of the concept allows us to correctly specify its extension. The characteristic tone of many of Wakefield's publications derives from the critical use of this method: (1) if analysis A of the concept of mental disorder (C) were sound, then condition X would not be a disorder and condition Y would, (2) but it is uncontroversial that X is recognized as a disorder and that Y is not; (3) accordingly, A is not an adequate analysis of C. For instance, if post-traumatic stress disorder (PTSD) is commonly recognized as a disorder and is quite expected in the context of trauma, then the previously mentioned *DSM-III-R*'s definition of a mental disorder as "a mental condition that causes distress and disability and that is not a statistically expectable response" is not correct (Wakefield 1992b, 233). The other task that an analysis of a given concept must complete is explaining consensus by making explicit what *grounds* the common intuitions of professionals and laymen. An analysis of the concept of mental disorder has to be able to tell us what deserves to be called a mental disorder, to set a standard for the proper use of the concept. As such, the ambition of HDA is not simply to be in harmony with a consensual view.

In the two 1992 articles, Wakefield contrasts the *concept* of mental disorder and a *theory* of disorder (Wakefield 1992a, 374; Wakefield 1992b, 232). The concept defines the proper domain of psychiatry (analyzing it is answering the question, What are mental disorders?), while a theory of mental disorders offers a general strategy for the explanation of such disorders (its purpose is to answer the question, Where do mental disorders come from?). This distinction, as pointed out by Wakefield, is crucial to the

DSM's project: an atheoretical classification of mental disorders, a goal that is only possible if clinicians with divergent theoretical commitments can agree on criteria that enable the diagnosis of mental illnesses in a converging manner. But is it possible to completely disentangle an *analysis* of the concept of mental disorders like HDA and a *theory* of mental disorders? HDA, with its reference to evolutionary biology, natural selection, and design, is obviously theory-laden, and it is in principle possible both to reject (or to ignore) Darwinism and to grasp the usual distinctions between disorders and nondisorders. To address this difficulty, the solution proposed by Wakefield is to argue that the HDA is composed of two distinct claims (Wakefield 1999, 374–375): the first more general claim is that disorders are dysfunctions of mental mechanisms with negative (harmful) consequences. This claim is supposed to articulate what grounds experts' and laymen's shared judgments of what counts (or what does not count) as a disorder in psychiatry. It is not linked to any specific construal of "function," and it would be HDA in its strictest sense. The second claim concerns the meaning of function and dysfunction: the dysfunction of a mental mechanism within the framework proposed by Wakefield is its "failure to perform a natural function for which it was designed by evolution" (Wakefield 1992a, 373). This second claim is derived from the idea that any ascription of mental disorder to an individual involves a factual component and that, to date, our best understanding of biological facts and of biological mechanisms (of which psychological mechanisms are a subtype) comes from evolutionary biology. Thus, HDA is theory-laden because it relies on our best knowledge of natural facts in its understanding of what a dysfunction is.

Once we understand what conceptual analysis is, the next logical question is, "What it is *for*?" The main issue for Wakefield is what he calls "conceptual validity," that is, "discriminating disorder from non-disorder" (Wakefield 1992b, 232), and conceptual validity is what can be achieved through a proper use of conceptual analysis. Of course, one can see that the demarcation between disorder and nondisorder matters for practical reasons—it determines in principle who should be cared for by mental health professionals and who should get reimbursement for treatment. It also matters from an institutional point of view—"'mental disorder' demarcates the special responsibilities of mental health professionals from those of other professionals, such as criminal justice lawyers, teachers, and social welfare workers" (Wakefield 1992a, 373). But conceptual validity matters first of all because we need a theoretical concept of mental disorder to justify the existence of psychiatry as a field of scientific knowledge. The question of justification has a close relationship to the question of boundaries. The question of the boundaries is more important to psychiatry than to other subfields of medicine, because the use of the concept of disorder in this field is surrounded by controversy and suspicion. On one hand, there is the "nihilism" of antipsychiatry, wherein there is no such thing as mental disorders (interestingly, this extremely skeptical view is the first that Wakefield addresses in 1992a). On the other hand, there is what has been called

the "medicalization of society," the "process by which nonmedical problems become defined and treated as medical problems" (Conrad 2007, 4). This is a well-known phenomenon in the field of mental health, and it is of much interest for Wakefield as is demonstrated by book titles such as *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder* (Horwitz and Wakefield 2007) or *All We Have to Fear: Psychiatry's Transformation of Natural Anxieties into Mental Disorders* (Horwitz and Wakefield 2012). The rejection of psychiatry as a whole, as well as its overextension, can be understood as two consequences of a same set of difficulties that reinforce each other. According to Wakefield, both derive (at least partly) from a lack of proper understanding of what mental disorders are. In particular, defining mental disorders exclusively through their unhappy consequences (they cause harm and often distress), or by the use of evaluative notions (something being "wrong" with a given individual), has the undesirable consequence of blurring the distinction between genuine disorders and problems in living, unusual, or disapproved behaviors. However, "grieving a lost spouse involves considerable suffering and being in a bad marriage is a problem of living but neither is a disorder" (Wakefield 1992a, 374). It would also be unsatisfying to claim that drapetomania is not a disorder *for us*, just because we do not share the values and beliefs of nineteenth-century advocates of slavery: this would stop us from saying that drapetomania is simply (and has always been) an erroneous category. Similarly, sociology can describe the medicalization of ordinary life, but it is not in a position to justify normative judgments about medical practice. For this reason, it is the specific task of philosophy to adequately address the question of conceptual validity. To define mental disorders as what is taken care of by psychiatry would leave us unable to consider false positives (as in the case of the Rosenhan experiment; Rosenhan 1973), as well as genuine disorders that do not receive proper medical attention.

It is the risk of diluting psychiatry (through relativism, overextension, and lack of legitimacy) that explains why Wakefield holds that the reference to an internal dysfunction, independent from values and social norms, is needed in our analysis of the concept of mental disorder. In his 1992 publications, Wakefield insists on the many disadvantages of the use of the term "dysfunction": (a) the word "dysfunction" in itself is vague and could be taken as a mere synonym of disorder; (b) contrary to the term "harm," dysfunction refers to something that is not directly observable and can only be inferred—we *record* signs of distress, but we *postulate* internal dysfunctions; (c) in the context of psychiatry, speaking of dysfunction obliges us to specify *what* is dysfunctional—cancer is obviously not a *mental* disorder, although organic diseases are both the result of the dysfunction of biological mechanisms and a source of distress and disability, just like mental disorders (Wakefield 1992a, 384); and (d) if we consider that a dysfunction is the basis of a given disorder, this obliges us to specify the norms of functioning that justify our judgment. If we can't provide such a justification, it will

always be possible to suspect that what we call a dysfunction is just the product of our negative evaluation of a given context or case.

However, Wakefield holds that these difficulties can be overcome and that the evolutionary view of mental functioning, in particular, is there to solve at least two of these problems.

One of these problems is (c), the question of what is dysfunctional. Thomas Szasz (1974) famously argued that there was no such thing as mental disorder because of an imperfect analogy: organic lesions impair bodily functioning, causing disorders with a causal history that we can describe in medical terms, but in the realm of the mental, there are no organs or lesions to be observed; we are just left with behaviors that are judged abnormal or deviant. He concluded that speaking of a mental "disorder" can only be a metaphorical means of expression hiding some hidden agenda. However, according to an evolutionary view of the mental, mental mechanisms can be inferred from their effects; they can be conceived of as efficient, adaptive tools like other biological mechanisms; and in given circumstances, they may be unable to perform their proper function just like any other evolved features of organisms. Fear responses to a dangerous environment, for instance, can be no less adaptive than any contribution of a bodily part of the organism to its well-being. If so, the breakdown of a mental mechanism is responsible for aberrant fear responses and can cause behaviors that can be considered maladaptive.

The other problem wherein the solution is offered by an evolutionary perspective is (d), the question of the norms of functioning: the proper function of evolved mental mechanisms is what they have been designed to do by evolution, which is independent of our values and preferences. In the case of drapetomania, we agree that there is no mental disorder—not simply because we reject the beliefs and values of advocates of slavery but primarily because to explain the behavior of the fleeing slave, we only need ordinary folk psychology and do not need to postulate anything abnormal within the mind of a slave. In choosing an evolutionary background to define the function of mental mechanisms, Wakefield intends to solve the problem of the normative dimension of the concept of function: "not working as designed" is proposed as a naturalized version of "not working as it should." In this evolutionary approach to the mental, Wakefield is in agreement both with the research program of evolutionary psychology and with the philosophical account of functions suggested by Larry Wright (1973), whom he explicitly references (see Wakefield 1992a). This account was later more fully developed by philosophers of biology such as Karen Neander (1991), and is known as the *etiological* view of functions, where F is a function of a component of type C in an organism O, if by doing F, former tokens of C contributed to the reproductive success of the ancestors of O.

According to HDA, however, the dysfunction of a psychological mechanism is a necessary but nonsufficient condition for the attribution of a mental disorder. To be

considered a disorder, the dysfunction must also be *harmful*. In 1992, the harm component of HDA did not receive as much attention as the dysfunction component, but it is equally important, and within the second part of his analysis of the concept, Wakefield distances himself from purely naturalistic accounts of disorder (e.g., Boorse 1976, 1977).

The importance of harm derives from two types of considerations. Considerations of the first type are related to practical aspects of general medicine. According to Wakefield, medicine is not concerned with dysfunction per se but with *significant* dysfunction, that is, dysfunction-producing effects that have some clinical salience (Wakefield 2014). As the breakdown of an internal mechanism or an anatomical anomaly resulting from an atypical developmental path may have no significant impact on the overall functioning of a given individual, we shall only speak of a disorder when the breakdown or the anomaly is detrimental to this individual in terms of well-being and ability.

Considerations of the second type are specific to psychiatry. When it comes to mental functioning and behavior, according to Wakefield, what is detrimental cannot be judged without a context wherein the resulting behavior is valued or disvalued according to established norms. This is why HDA predicts cases where in an individual A, the failure of a mental mechanism to perform a natural function for which it was designed by evolution is not a source of harm or is even advantageous. This possibility results from the difference between the environment in which the effect of the mechanism has been selected because it was advantageous to A's ancestors and the present environment wherein this same effect is no longer adaptive. As the notion of harm is said to have "an intrinsic value component" (Horwitz and Wakefield 2007, 217), the concept of mental disorder, according to HDA, cannot be a purely scientific concept.

In the original presentation of HDA, Wakefield (1992a) pays special attention to the discrepancy between past and present environments. But, perhaps because of the uneasy relations between HDA and cultural psychiatry (Gold and Kirmayer 2007), the roles of harm and values are also reconsidered in a different perspective, above all in more recent writings where Wakefield deals with the issue of cultural relativity. The key question, then, is no longer that of a historical modification of norms of behavior; rather, the question is the context sensitivity of responses to the environment and the distinction between normal and pathological responses. Wakefield recognizes that culture may shape behaviors in such a way that responses to the environment that would be inappropriate in one context may be unproblematic in another. Yet that does not imply that the project of a demarcation between normal and abnormal responses is a chimera. It only means that, for instance, in the case of the distinction between sadness and depression, culture may define what types of loss for which sadness is a normal response (it defines what is, in general, *valuable* to possess) and that we should take this sensitivity to the environment into account to draw the line correctly between disorder and nondisorder (Horwitz and Wakefield 2007).

Worries about HDA

Before introducing the chapters of the present volume, which offer new critical perspectives on HDA, we shall present a brief survey of the debate that has been generated by Wakefield's view of mental disorders since 1992. We will not, however, be able to address every question, such as the relation between the descriptive and the normative dimensions of HDA (Kirmayer and Young 1999) or the role given by HDA to shared intuitions about mental disorders.

One of Wakefield's key claims is that conceptual analysis matters for medical practice, especially because we need a solution to the conceptual validity/demarcation problem. On one hand, Wakefield holds that only a valid definition of what mental disorders are is able to ground our classificatory judgments. On the other hand, symptom-based definitions of mental illnesses too often lead to an unjustified medicalization of normal conditions. Defining mental disorders, then, is of primary importance for the psychiatric community. Yet as we can see from the ongoing debate in the literature, this view is controversial. One issue is the possibility of analyzing the concept of mental disorder in terms of necessary and sufficient conditions that would be identical for all types of syndromes, from schizophrenia to personality disorders (see, in this volume, the contributions of Leen De Vreese and Peter Zachar in chapters 5 and 7, respectively). If there are only family resemblances between kinds of mental disorders, then the quest for an overarching definition (including, through the dysfunction clause, a similar etiology), which would allow us to solve the demarcation problem in a great majority of cases, may prove futile. Another issue is the relevance of theoretical definitions (like that offered by HDA) to medical decisions. Some have argued that when Wakefield is discussing the *DSM*'s criteria and unsubstantiated ascriptions of disorder, he is using, in fact, "folk concepts" and commonsense intuition about what proportionate or appropriate responses to the environment can be in given circumstances and that he does not rely on an evolutionary psychological theory of mental mechanisms (Bolton 2008, 143–145). This retreat could at least partly be explained by our ignorance of the limits of normal variation (Schwartz 2007) and the way we usually infer the existence of a disorder. For instance, in his reply to Lilienfeld and Marino, Wakefield holds that the symptoms "caused by design failures" are "so extreme that they do not significantly overlap with normal functioning" (Wakefield 1999, 387). He adds that "there is a naturally selected range of the sensitivity of fear-response mechanisms, but the spontaneous terrors of panic disorder are not part of that range" (387). We could ask ourselves, however, if "design failures" and "naturally selected range" of a mechanism's sensitivity add anything but adjectival nuance from an evolutionary biology-inspired lexicon of familiar medical categories (disorder and clinical heterogeneity within populations). In this case, deciding which responses are pathological is not based on evolutionary considerations but only on the fact that some clinical phenomena are both statistically

rare and harmful. In such contexts, drawing a line between disorder and nondisorder is *compatible* with an evolutionary view of mental functioning, but it does not *depend* on it and it is *not inferred* from a prior knowledge of natural function and design.

Other issues are linked to the dysfunction component of HDA. First, Wakefield's view of the design of the human mind is close to that of evolutionary psychology, which has been criticized on several grounds—in particular, for not meeting the methodological requirements of evolutionary biology (Richardson 2010; see Faucher, chapter 3, this volume). Second, the etiological account of functions is only one among several alternatives, and as a consequence, it is conceivable that one can redefine mental disorders with a different, nonhistorical background. One option is the analysis of causal roles suggested by Robert Cummins (1975). According to Cummins, a function F of a component C in a system S is a contribution of C to the explanation of a given capacity of S. Although it has often been said that this view of functions reflects the use of functional talk in physiology and neuroscience, Cummins himself has linked his view of causal analysis to explanatory practices in psychology (Cummins 1985). Moreover, it is not impossible to derive an account of dysfunctions from this view of functions (Godfrey-Smith 1993), and regarding psychiatry, it has been vindicated as an alternative to HDA, for instance, by McNally (2001) and Murphy (2006; Murphy, chapter 13, this volume). Furthermore, the view of functions and dysfunctions within the biostatistical theory of health offered by Christopher Boorse, which he applies to psychiatry (Boorse 1976), can be understood as a combination of Cummins-style functions with a biological background via the reference to survival and reproduction (Forest and Le Bidan 2016).

Even if we keep Wakefield's evolutionary framework, his historical view of functions, disorders, or problems in living may not come from the dysfunction of an evolutionary mechanism. One alternative is a variant of the idea of evolutionary mismatch evoked in Wakefield (1992a). Instead of a "design failure" that is not harmful, because of the difference between the environment of evolutionary adaptedness (EEA) where the functional effect has been selected and present conditions, we would observe the reverse association (i.e., harmful consequences of normal functioning): in this latter case, the mechanism is working as designed, but its selected effect is no longer beneficial in the given circumstances. If we use Wakefield's example (Wakefield 1992a, 384), high levels of aggression may become grossly inadequate in certain conditions or life. In other cases, evolved mechanisms may be triggered by the "wrong" kind of stimuli in contemporary environments, stimuli for which they have not been designed to respond. Either we should still count the outcome in these kinds of cases as an instance of a genuine mental disorder (on the basis of its negative consequences) and give up HDA, or we should revise psychiatry manuals and shorten the list of mental disorders. Another type of scenario corresponds to what has been called by Nesse (2002) "evolved defenses": evolved defenses may cause pain or discomfort, but they are beneficial nonetheless. Some conditions usually labeled "disorders" would be frequent

and heritable because they are, in fact, adaptive; even accompanied by distress, they would not, strictly speaking, be detrimental. Such a hypothesis has been vindicated in the literature, for instance, in the case of depression (for a review, see Faucher 2016). Introducing evolutionary considerations in psychiatry, then, may challenge the traditional understanding of conditions such as depression and block the ascription of an underlying dysfunction instead of supporting it, as is the case within the framework of HDA. It is also worth pointing out that both hypotheses in terms of mismatch and in terms of evolved defenses are concerned with the explanation of the persistence of disorders within human populations rather than with the explanation of individual disorders—a question that is not directly addressed by HDA.

One of Wakefield's key claims with HDA is that we should combine (rather than oppose) two ways of understanding mental disorders: one being biological and objective, the other being social and perspectival, with the idea of harmful consequences that are necessary to the ascription of a mental disorder and open to variation in different contexts of evaluation. As we have seen, the "harm" component of HDA has not received as much attention as the dysfunction component in the original presentation of the theory (Wakefield 1992a). Since then, *how* this process of valuation is supposed to take place, *whose* values are (and should be) taken into account, and *how* conflicts about values are managed have never been completely clarified (Poland 2003). Moreover, choosing a middle ground between naturalism and social constructivism exposes HDA to attacks from both sides. Some researchers may question the role of values and look for purely causal explanations of mental disorders (see Gerrans, chapter 19, this volume). Others criticize HDA either because of a division between facts and values, the natural and the social, that they judge illusory (Bolton 2008), or because of what they perceive as an inadequate vision of the role of social factors in Wakefield's proposal (Kirmayer and Young 1999). In particular, HDA makes a distinction between the (natural) basis of a disorder and the (social) evaluation of the consequences of the underlying dysfunction. Yet social and cultural factors may intervene in the causal chain leading to a disorder (Kendler 2005), and even if we pay attention to the distinction between broad (nonspecific) and narrow (specific) etiology (Wakefield 2014), it seems difficult to restrict the role of society to the evaluation of a preexisting condition whose existence depends solely on the failure of an internal, mental organ to do what it has been designed to do.

Last, HDA was conceived in a context where scientific psychiatry was exemplified by the *DSM*. A key architect of the *DSM* project (like Leo Spitzer) has welcomed HDA as a positive contribution that would help future editions of the *DSM* "make revisions in the diagnostic criteria more valid as true indicators of disorder" (Spitzer 1999, 430). However, in the psychiatric community, there is a growing dissatisfaction with the whole project of an "atheoretical" classification of mental disorders (Demazeux and Singy 2015), and the past decade has been marked by the emergence of the Research Domain Criteria (known as the RDoC) of the National Institute of Mental Health,

which has been explicitly presented as an alternative to the *DSM* (at least in psychiatric research context) where mental disorders (more precisely, their symptoms) are linked to their genetics, molecular and neural basis (Insel and Cuthbert 2010; Faucher and Goyer 2015). Wakefield himself has expressed strong reservations as to the RDoC methodology (Wakefield 2014): according to him, the RDoC project as it stands is unable to deal with the key issue of *conceptual validity*, as the description of brain circuits can be linked equally well to disorders or nondisorders; it would be only at the psychological level that we can make the distinctions that allow us to delineate the proper domain of psychiatry. However, the theoretical landscape is quite different today from what it was in 1992, and one wonders if the definition of mental disorders has to be completely divorced from ongoing psychiatric and scientific research.

The Content of the Volume

The present volume is organized in four sections, each reflecting an aspect of Wakefield's analysis of health as "harmful dysfunction." Sections comprise chapters reflecting on HDA's methodology (mostly, conceptual analysis), on its goal (the demarcation between disordered and nondisordered states), or on the elements of the *analysans* proposed by Wakefield ("dysfunction" and "harm"). Each chapter is followed by a reply (which sometimes is also followed by a supplementary reply) from Jerome Wakefield.

Part I: On Conceptual Analysis

In chapter 1, "*DSM* in the Light of HDA (and Conversely)," Steeves Demazeux challenges on historical grounds a claim made by Jerome Wakefield in his defense of his theory—namely, that the HDA is in complete agreement with the spirit of modern psychiatry in general, especially with the conception of mental disorders relied upon by Spitzer and his colleagues in the conception of the *DSM-III*. In fact, Demazeux makes two separate but related claims. The first claim concerns Spitzer's views: in his early, seminal papers, where the criteria of "distress and disability" are essential to the identification of mental disorders, Spitzer does not appear to give a prominent role to the criterion of dysfunction, as it is claimed by Wakefield. The second claim concerns the relationship between HDA and the whole *DSM* project. With his symptom-based approach, the *DSM* could not be easily reconciled with the HDA approach, which involves a very specific type of etiology—a dysfunction of an evolved psychological mechanism. Thus, Wakefield's claim that the *DSM* is contradictory rather reflects his own mischaracterization of the *DSM*'s ambitions rather than being the result of the *DSM*'s failure to be faithful to its own characterization of mental disorder. This chapter is offered not as a rebuttal of HDA but as an attempt to more precisely contextualize the emergence of HDA within the context of psychiatry during the 1980s and 1990s.

In chapter 3, "Facts, Facts, Facts: HD Analysis Goes Factual," Luc Faucher takes on a different task by challenging Wakefield to go "factual" all the way, without reservation, in terms of his theory. Faucher identifies two domains where going factual might prove to be worthy of the effort. First, he reminds us that Wakefield thinks of conceptual analysis as a form of empirical investigation into the structure of our concepts. As the X-Phi (experimental philosophy) movement has shown, conceptual analysis is not devoid of biases, and for this reason, it is better to use various techniques to reveal the content of our concepts. As Faucher observed, Wakefield has already started using some of these techniques and claims that the results of his experiments support his version of HDA. Taking home some of the lessons gleaned from discussions about the methodological limitations of actual X-Phi experiments, as well as identifying some limitations inherent to Wakefield's experiments, Faucher invites Wakefield to more extensively test his theory (with a wider variety of questions and on a variety of groups), in addition to using different methods. As it has been shown via some preliminary studies' results using different methodologies, the concept of mental disorder held by different people might be much more diversified and sensitive to context than Wakefield had originally posited. Second, Faucher considers what has been seen by Wakefield as an "epistemological problem" (i.e., a problem that does not question the validity of his conceptual analysis but only its capacity to be applied in certain contexts): the problem of establishing precisely what is the proper function of a particular mental mechanism. According to Faucher, this problem might indeed demonstrate the limits of Wakefield's analysis. If the dysfunction portion of Wakefield's analysis is supposed to be a prophylactic against the excess of normative theory of mental disorders such as Szasz's, it is important to be able to establish what constitutes the proper function of the mechanisms that are thought to be dysfunctional. If one is not able to do so, there is a risk that values and social norms will sneak back in through the postulation of mental mechanisms that do not exist. Faucher argues that this is precisely the problem in Wakefield's analysis: for many important mental "faculties" or "capacities" (faculties or capacities that play a central role in the explanation of some paradigmatic mental disorders), it might not be possible to establish their proper function, which would leave psychiatry without a scientific image of the properly working mind to which it could refer to, in order to ground its judgments of dysfunction.

Leen De Vreese, in "Against the Disorder/Nondisorder Dichotomy" (chapter 5), argues that "disease" is a multifaceted concept that cannot be captured by HDA (or by any single definition). According to De Vreese, we need a pluralistic approach that would capture the different ways we use the notion of disease, rather than an approach that would aim to capture our intuitions about it. De Vreese analyzes the motivations behind Wakefield's conceptual analysis and observes that Wakefield seems to be moved by contradictory objectives: either to *describe* our intuitions or to *present a revised version*

of them (to correct them where they err). Yet De Vreese also shows that whichever objective Wakefield's conceptual analysis is pursuing, it will encounter problems, which will ultimately make HDA difficult to use in practice to demarcate disorders from nondisorders. In and of themselves, these are arguments against the usefulness of conceptual analysis and a reason for the development of new methods to study our different uses (and meanings) of the concept in practice.

In chapter 7, Harold Kincaid begins "Doing without 'Disorder' in the Study of Psychopathology" by identifying what he takes to be three of Wakefield's major contributions. The contributions include (1) maintaining a "healthy" skepticism concerning psychiatric classification, (2) supplying reasons for the belief in a nonarbitrary distinction between disorders and nondisorders, and (3) assessing specific psychiatric categories (depression, phobias, etc.) to determine whether or not they capture (only) disordered conditions (rather than problems in living). Kincaid's main point is that these contributions do not necessitate a conceptual analysis of the concept of disorder. Among the reasons fueling his position are the facts that in science, concepts are usually not strictly defined; that it does not seem to be a good idea to tie the development of a scientific field to commonsense concepts; and that the defense against antipsychiatrist claims of medicalization of normal life can be accomplished without a definition of what disorders are. Kincaid does not believe that Wakefield's particular analysis of disorders in terms of harmful dysfunction of evolved mechanisms is necessary to evaluate the potential overinclusiveness of diagnostic categories (one of Wakefield's major contributions). Rather, according to Kincaid, psychiatric disciplines need objective and explanatory classifications (which need to delineate real distinctions between people that can be used to successfully explain, predict, and control behavior), and such classifications can be achieved without the analysis of the concept of mental disorder and without references to the evolutionary history of mental mechanisms.

Part II: The Demarcation Problem

In chapter 9, "Psychiatric Disorders and the Imperfect Community," Peter Zachar denounces the inherent essentialism (both causal and psychological) behind Wakefield's definition of mental disorder. Zachar posits that, for Wakefield, attributions of disorder are made on the basis of reasoning rather than empirical evidence: basically, it depends on one's concept of "objective natural function," which, despite what Wakefield claims, is not something that is empirically determined. Rather, it seems that one uses a conception about the responses that are to be expected by someone facing a type of situation (e.g., the death of a love one), and from this conception (which is usually not based on science), one infers whether or not the individual's mental mechanism is disordered. If such is the case, as Zachar points out, HDA cannot do what it set out to accomplish (i.e., factually demarcate valid psychiatric diagnostics from invalid ones). Through a discussion of Paul Meehl's notions of "open concept"

and "construct validation," Zachar explains how one can reject Wakefield's essentialism and ground psychiatric diagnostic in facts. His argument rests on the observation that what we call disorders are the result of a mix of functional disorders (e.g., intrusive thoughts, impulse control difficulties, or decline in functioning), which form an "imperfect community" in that, if they all have been used to identify disorders (they have this in common), they are different in nature (therefore, they are imperfect because they do not necessarily share any other properties). In the literature, it is posited that from a particular mix or pattern of these functional disorders (what you might want to call the "manifest structure" of a disorder), you can infer a latent variable (i.e., a particular disorder, like depression). Latent variables are either thought of in a realist fashion (i.e., the latent variable is understood as being the thing that causes the observable pattern that defines the disorder) or nonrealist fashion (e.g., the latent variable refers to a stable set of functional disorders resulting from mutual interactions between the elements of the pattern). Zachar argues that the latter way of understanding disorders might prove to be much more fruitful for psychiatry than the essentialist way endorsed by Wakefield.

Part III: The Dysfunction Component

In chapter 11, "Is the Dysfunction Component of the Harmful Dysfunction Analysis Stipulative?" Maël Lemoine argues that in his treatment of the notion of "dysfunction," Wakefield is moving away from a conceptual analysis of the commonsense concept of disorder and entering the realm of stipulation. Relying on Hempel's distinction between various types of definition, Lemoine explains that Wakefield's conceptual analysis has elements of *meaning analysis* and elements of *stipulation*. The principal element of stipulation is the notion of dysfunction. According to Lemoine, the correct analysis of the commonsense concept of disorder is probably what he terms the harmful abnormality analysis (or HAA), where someone has a disorder if (1) they have an abnormality and (2) this abnormality is harmful. Wakefield's HDA can be seen as an explicitation of the HAA, as the concept is made less vague and more powerful empirically. Yet interpreting "abnormality" in terms of "dysfunction of an evolved mechanisms" is not an explicitation of the concept (which consists of arranging and stabilizing the sense of the concept) but rather a stipulation. This is made clear by the fact that there are other ways to interpret "dysfunction," for instance, as per Boorse's argument. There would be no conceptual problem to stipulate that abnormality has to be understood in terms of the dysfunction of an evolved mechanism if Wakefield did not view it as his job to provide a conceptual analysis of the commonsense concept. But this is not the case. Moreover, and this is Lemoine's last point, it is not at all sure that Wakefield's notion of dysfunction is the best one available.

In chapter 13, "Function and Dysfunction," Dominic Murphy first reminds us that the requirement of the presence of the dysfunction of a psychological mechanism is motivated in Wakefield's analysis by his rebuttal of purely normative accounts of

mental disorders. Murphy then questions the relevance of Wakefield's evolutionary understanding of function to the field of psychiatry. As per Murphy, a proper account of function in psychiatry should satisfy the following criteria: first, it should be able to ground our intuition that in some cases, something has gone wrong within the mind of an individual. Second, it should be in accordance with standard scientific practice. Third, it should allow us to handle various cases without adopting a "revisionist" attitude, where we exclude some conditions from the list of disorders just to save the theory we favor. Murphy holds that if we adopt these criteria, the systemic view of functions introduced by Cummins, and then revised in a more naturalistic spirit by his followers, fares at least as well as, and in some cases better than, the evolutionary view advocated by Wakefield. The systemic's perspective of Cummins is no less able to justify our intuitions about dysfunctions than the selectionist's view endorsed by Wakefield. Furthermore, it meshes especially well with biological sciences that are more closely related to research in psychiatry, and it does not lead to a drastic revision of our taxonomy of mental disorders because it does not have to identify each dysfunction with a "design failure."

In chapter 16, "The Developmental Plasticity Challenge to Wakefield's View," Justin Garson challenges Wakefield's idea that a mental disorder necessarily involves a dysfunction of a mental mechanism, where dysfunction is understood in terms of the failure to execute the function for which it has been selected. Indeed, Garson claims that some disorders might be the result of mechanisms in perfect working order. This is the case, he proposes, for some developmental mechanisms for which parameters are set by the environment early on in life. In such cases, there is a possibility of "developmental mismatch"—that is, it is possible that the early environment is not at all like the later environment, and what was adaptive in the first environment is maladaptive in the second one. If such is the case, it is possible that what we judge to be a dysfunctional behavior is caused by the working of a perfectly well-ordered mechanism, whose function is to adapt the organism to its environment by sampling earlier environments and taking this environment as a reliable cue of later environment. Garson argues that this is not only a view from the mind, but that there is an actual current of research in psychiatry that takes this possibility very seriously (the "Developmental Origins of Health and Disease" program). Through a careful discussion of Wakefield's view of dysfunction and his answer to the evolutionary mismatch's argument, Garson also shows that Wakefield is committed to the notion that dysfunctional behavior is caused by a functional mechanism. If Garson is right and some mental disorders are the result of developmental mismatches, then Wakefield's claims that dysfunction of a mental mechanism is a necessary condition for mental disorder are invalidated. Moreover, if the possibility of mental disorders caused by intact mechanisms is taken seriously by psychiatrists, then it seems that Wakefield's conceptual analysis does not perfectly capture the intuition of psychiatrists (at least of some of them and for some psychiatric conditions).

In chapter 19, "Harmful Dysfunction and the Science of Salience," Philip Gerrans holds that the most promising research path in terms of the explanation of mental disorders is not to look for psychological mechanisms that could fail to do what they have been designed to do but rather to focus on lower-level, molecular, and neural mechanisms whose integrity is crucial to our ordinary mental activity, even if they are only indirectly related to it. His key example is the salience system, wherein impaired functioning causes aberrant valuation of stimuli, which, in turn, may cause delusional states. If such is the case, delusions are not caused by the failure of an epistemic system whose task it is to produce true beliefs, because the evolutionary history of a salience system (or dopaminergic system) has little to do with the acquisition of true beliefs. Supporting empirical data for this argument are gathered from a body of fast-growing literature that links key symptoms of schizophrenia with abnormalities in dopamine regulation. Gerrans's idea, then, is to look for explanations of disorders in terms of lower-level relevant neural mechanisms, rather than at the level of psychological mechanisms (as Wakefield has typically done), and to suggest that this might lead one to think of psychiatric conditions in a revisionary way.

In chapter 21, "Autistic Spectrum, Normal Variation, and Harmful Dysfunction," Denis Forest focuses on the example of autism to challenge key components of HDA, namely, the evolutionary background of the function/dysfunction distinction and the link between harm and value. Forest observes that it is true that some psychological theories have tried to explain autism through the malfunction of specific, evolved mental mechanisms or modules, and on this basis, it could be argued that HDA and explanatory theories of autism are in complete harmony. However, autism research is more concerned with the ontogeny of mental mechanisms than with their evolutionary origin. In the context of the neurodiversity movement, which claims that we should consider autism as an instance of normal variation in human populations, it is difficult to see how HDA could tell us when behavioral and cognitive differences should be understood in terms of underlying dysfunction and when they are not. Moreover, confronted with the heterogeneity of cases within the autism spectrum, we want to highlight the difference between autism's harmless and harmful features. But what makes a dysfunction harmful is its intrinsic detrimental consequences, not the fact that it would be disvalued in a given social context. As Forest shows, recent shifts in the representation of high-order autism do not change a disorder into a healthy condition; they unmask abilities that had previously remained undetected. In the context of autism and neurodiversity, we need other criteria of dysfunction and harm than those specified by HDA.

In chapter 23, "Naturalism and Dysfunction," Tim Thornton questions the reductionist, objectivist, naturalist account of dysfunction that is central to HDA. Light can be shed on the prospect of reducing the apparently normative notion of dysfunction by comparing it to two distinct reductionist projects in the philosophy of mental

content that stand next to one another as do the contrasting options in the Euthyphro dilemma. A more modest project (Fodor's representational theory of the mind) takes for granted the structure of normative relations between concepts and attempts to solve the engineering problem of how human thought can fit that structure. A more ambitious project (Millikan's teleosemantics) aims to explain that structure itself in naturalistic terms. This ambitious project, however, is undermined by Wittgenstein-Kripke's paradox. Tim Thornton argues that the harmful dysfunction analysis of disorder has to be interpreted as isomorphous with the latter project, as its aim is to explain how disorders are possible within the natural world. It is thus subject to the same objections raised against Millikan's project: if we cannot choose between rival accounts of mental functions, our understanding of mental disorders as natural dysfunctions is also undermined.

Part IV: The Harmful Component

In chapter 25, "Harmless Dysfunction and the Problem of Normal Variation," Andreas De Block and Jonathan Sholl focus on the harm component of HDA, wherein they question the presumed separability of the scientific, objective, factual requirement of a state of dysfunction from the second requirement that concerns harm and value in HDA. First, they point out that clear cases of *harmless* dysfunction are crucial to HDA: it is only because some dysfunctional states are not pathological (because they have no harmful consequences) that we can dissociate a value-free assessment of the loss of functional integrity and a value-laden judgment about the loss of health. Second, they discuss the cases of harmless dysfunction mentioned by Wakefield and offer skeptical counterpoints to his interpretations: one may wonder if fused toes and albinism are genuine cases of harmless dysfunction. Then, in the last and most ambitious part of the chapter, they use the problem of suboptimal variation to directly challenge Wakefield's main thesis. On one hand, if any kind of suboptimal variation is an instance of dysfunction, one departs from the standard use of the term "disorder" and stretches the notion beyond its reasonable limits. On the other hand, if what is suboptimal coincides with what is dysfunctional only when it has detrimental consequences, then we cannot really separate dysfunction from harm and HDA runs into trouble.

In chapter 27, "On Harm," Rachel Cooper expresses both her agreement with Jerome Wakefield (the harm component is crucial to the definition of mental disorders) and her disagreement with him (according to her, harm should not be understood as what is disvalued by a given society). Looking for a different measure of harm, she contrasts what she sees as an overly ambitious goal (grounding a construal of harm on an overall conception of the good life) with a more modest one (defining ways to assess if a given condition is harmful or not). She offers three methods to make progress on this issue: directly assessing the consequences of a given condition, analyzing cases in terms of

cost and benefits, and looking for consistency when we use criteria to judge whether something is harmful. Finally, she holds that we should not think in terms of disorder when people are a cause of harm without being harmed themselves; Cooper stresses that in the context of the *DSM-5*, more than ever, the emphasis on harm is linked to a key concern of a reflection on psychiatry: preventing the unwarranted medicalization of ordinary life.

Acknowledgments

This project grew out of a conference organized by one of us (D. F.) on Jerome Wakefield's analysis of mental disorder held in Paris in 2010 and financially supported by the Agence Nationale de Recherche (ANR). We thank Pierre-Henri Castel for teaming us up for this collection of essays and for his support. We are also grateful to Jeffrey Poland for his advice and encouragements about submitting the manuscript to MIT Press. We would like to extend our gratitude to Philip Laughlin, editor at MIT Press, for his help with the project and his incredible patience and to Alex Hoopes, with whom it has been a pleasure to work in preparing the manuscript. L. F. thanks the Faculté des Sciences Humaines (FSH) de l'Université du Québec à Montréal for its financial support of this edition of the book, as well as Evi Amanda Leigh-Cox for the copy editing of this chapter and Cloé Gratton for her work on the formatting of the manuscript.

References

Bolton, D. 2008. *What Is Mental Disorder?* Oxford University Press.

Boorse, C. 1976. What a theory of mental health should be. *Journal of the Theory of Social Behavior* 6(1): 61–84.

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

Conrad, P. 2007. *The Medicalization of Society*. Johns Hopkins University Press.

Cummins, R. C. 1975. Functional analysis. *Journal of Philosophy* 72(20): 741–764.

Cummins, R. C. 1985. *The Nature of Psychological Explanation*. MIT Press.

Demazeux, S., and P. Singy, eds. 2015. *The DSM in Perspective: Philosophical Reflections on the Psychiatric Babel*. Springer Verlag.

Faucher, L. 2016. Darwinian blues: Evolutionary psychology and depression. In *Sadness or Depression?* S. Demazeux and J. Wakefield (eds.), 69–94. Springer Verlag.

Faucher, L., and S. Goyer. 2015. RDoC: Thinking outside the *DSM* box without falling into a reductionist trap. In *The DSM in Perspective: Philosophical Reflections on the Psychiatric Babel,* S. Demazeux and P. Singy (eds.), 199–224. Springer Verlag.

Forest, D., and M. Le Bidan. 2016. In search of normal functions: BST, Cummins' functions and Hempel's problem. In *Naturalism in Philosophy of Health: Issues, Limits and Implications*, E. Giroux (ed.), 39–51. Springer Verlag.

Godfrey-Smith, P. 1993. Functions: Consensus without unity. *Pacific Philosophical Quarterly* 74: 196–208.

Gold, I., and L. Kirmayer. 2007. Cultural psychiatry on Wakefield procrustean bed. *World Psychiatry* 6(3): 165–166.

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

Horwitz, A. V., and J. C. Wakefield. 2012. *All We Have to Fear: Psychiatry's Transformation of Natural Anxieties into Mental Disorders*. Oxford University Press.

Insel, T., and B. Cuthbert. 2010. Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry* 167(7): 748–750.

Kendler, K. S. 2005. Toward a philosophical structure of psychiatry. *American Journal of Psychiatry* 162(3): 433–440.

Kirmayer, L., and A. Young. 1999. Culture and context in the evolutionary concept of mental disorder. *Journal of Abnormal Psychology* 108(3): 446–452.

McNally, R. J. 2001. On Wakefield's harmful dysfunction analysis of mental disorder. *Behavioral Research and Therapy* 39(3): 309–314.

Murphy, D. 2006. *Psychiatry in the Scientific Image*. MIT Press.

Murphy, D., and R. L. Woolfolk. 2000. The harmful dysfunction analysis of mental disorders. *Philosophy, Psychiatry, and Psychology* 7(4): 241–252.

Neander, K. 1991. The teleological notion of 'function.' *Australasian Journal of Philosophy* 69(4): 454–468.

Nesse, R. 2002. Evolutionary biology: A basic science for psychiatry. *World Psychiatry* 1(1): 7–9.

Nordenfelt, L. 2003. On the evolutionary concept of health: Health as natural function. In *Dimensions of Health and Health Promotion*, L. Nordenfelt and P.-E. Liss (eds.), 37–54. Rodopi Press.

Poland, J. 2003. *Whither Mental Disorder*. Unpublished manuscript.

Richardson, R. C. 2010. *Evolutionary Psychology as Maladapted Psychology*. MIT Press.

Rosenhan, D. L. 1973. Being sane in insane places. *Science* 179: 250–258.

Schwartz, P. 2007. Defining dysfunction: Natural selection, design, and drawing a line. *Philosophy of Science* 74(3): 364–385.

Spitzer, L. 1999. Harmful dysfunction and the *DSM* definition of mental disorder. *Journal of Abnormal Psychology* 108(3): 430–432.

Szasz, T. 1974. *The Myth of Mental Illness*. Harper and Row.

Wakefield, J. C. 1992a. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47(3): 373–388.

Wakefield, J. C. 1992b. Disorder as harmful dysfunction: A conceptual critique of *DSM-III R*'s definition of mental disorder. *Psychological Review* 99(2): 232–247.

Wakefield, J. C. 1999. Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology* 108(3): 374–399.

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.

Wakefield, J. C. 2014. Wittgenstein's nightmare: Why the RDoC grid needs a conceptual dimension. *World Psychiatry* 13(1): 38–40.

Wright, L. 1973. Functions. *Philosophical Review* 82(2): 139–168.

Defining Mental Disorder

Jerome Wakefield and His Critics

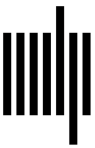
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl

Edited by: Luc Faucher, Denis Forest

Citation:

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

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



The MIT Press

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

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

Library of Congress Cataloging-in-Publication Data

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