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

20 Are Cognitive Neuroscience and the Harmful Dysfunction Analysis Competitors or Allies? Reply to Philip Gerrans

Jerome Wakefield

I thank Philip Gerrans for his lucid and fascinating exploration of the developing interface between cognitive neuroscience and psychiatric nosology as seen through the prism of salience theory and related cognitive neuroscientific theories, and for his assessment of the relationship of these developments to my harmful dysfunction analysis (HDA) of medical, including mental, disorder. The HDA claims that "disorder" refers to "harmful dysfunction," where dysfunction is the failure of some feature to perform a natural function for which it is biologically designed by evolutionary processes and harm is judged in accordance with social values (First and Wakefield 2010, 2013; Spitzer 1997, 1999; Wakefield 1992a, 1992b, 1993, 1995, 1997a, 1997b, 1997c, 1997d, 1998, 1999a, 1999b, 2000a, 2000b, 2001, 2006a, 2007, 2009, 2011, 2014, 2016a, 2016b; Wakefield and First 2003, 2012).

I am excited by Gerrans's attempt in his work on the salience system to bring together an understanding of neurocognitive mechanisms with the phenomenological quality of lived experience, a long overdue but elusive synthesis. Although I sympathize with much of Gerrans's position and aspirations, I focus here on some areas of possible disagreement—"possible" because Gerrans writes with a light rhetorical touch that sometimes leaves his claims ambiguous. I see three primary areas of possible disagreement: the demoted heuristic role rather than conceptually fundamental role that Gerrans suggests for biological design and evolution and thus for the HDA in a neuropsychiatric theory of psychopathology, the sharp dichotomy Gerrans sees between what he calls "evolutionary psychology" or "strong adaptationism" and cognitive neuroscience, and the notion that there is a radical discontinuity between the sorts of insights he pursues and the aspirations of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*). In exploring these claims, I will also make reference to the related views of Dominic Murphy, whom Gerrans cites.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

Mechanical-Causal and Biological-Design Analyses as Complementary in a Theory of Disorder

First, then, is mechanical-causal analysis enough for a theory of psychopathology? Despite agreement that evolutionary theory is relevant to the theory of psychopathology, there seems to be a fundamental divergence between Gerrans and me on the nature of the role of evolutionary theory in an account of psychopathology. Gerrans agrees that evolutionary analysis applies to the neuroscientific salience systems he describes:

> Wakefield is surely right that the explanation of this phenomenon involves the functioning of a salience system "designed" by evolution, which is either malfunctioning (making the wrong information salient) or, if it is functioning correctly … warping the functioning of other systems with which it interacts. The explanation of salience outlined above makes use of evolutionary ideas all the way through.

On first glance, this passage seems to suggest an HDA approach. However, Gerrans also distances himself from the HDA's view that evolutionary theory provides a unique and essential conceptual foundation for disorder judgments. He construes evolutionary insight rather as a useful heuristic for constructing neuroscientific causal-mechanical explanations of symptoms:

> Wakefield is right to say that if we want psychiatric classification to reflect causation, we need to ensure that evolutionary considerations are theoretically incorporated in ways that help reveal underlying neural and cognitive mechanisms. There are, however, different ways to do this. … Once we abandon the strong adaptationist approach in favor of a cognitive neuroscience informed by evolutionary theorizing, psychiatric classification requires substantial revision. … I think that the right approach here is the radical one: a complete reconceptualization of the phenomenon using the vocabulary of cognitive neuroscience, informed by, but not necessarily invoking, evolutionary theory.

I will get to "different ways to do this," "strong adaptationism," and why classification "requires substantial revision" later. For now, observe that evolutionary considerations are important to Gerrans because they can "help reveal underlying neural and cognitive mechanisms." Evolution is thus demoted from the HDA's requirement that biological design must be explicitly or implicitly invoked in identifying a neurocognitive condition as psychopathology (more on this in a moment, too) to a consideration in which cognitive neuroscience is merely "informed by" evolutionary theorizing as a perspective that can yield additional insights in hypothesizing mechanisms, a vague and noncommital relationship with no conceptual bite.

So far as I can tell, the reason that Gerrans—like Dominic Murphy, whom he cites—distances himself from the HDA's evolutionary framework for disorder judgments is that he believes that the HDA's evolutionary perspective is somehow in tension with or an alternative to the neuroscience agenda, rather than a complement to it. However,

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

the idea that the HDA favors evolutionary explanation *over* mechanistic neuroscientific explanation confuses the conceptual-analytic and scientific-theory domains. It makes no sense on its face because an evolutionary explanation is based on and supported by a detailed understanding of the design-like characteristics of mechanisms, neurocognitive or otherwise, and their causal interactions. Even Tinbergen's (1963) classic list of four necessary components of an evolutionary explanation includes a mechanistic-causal understanding of how an evolved system works as essential to any such undertaking. More recently, the HDA was reportedly an inspiration for the National Institute of Mental Health's (NIMH's) Research Domain Criteria (RDoC) project to identify brain circuitry dysfunctions underlying psychopathology (see my reply to Demazeux in this volume), which suggests that NIMH understands the implications of the HDA better than do Gerrans and Murphy. In any event, for practical clinical interventive purposes, a mechanistic understanding of disorders is obviously of overwhelming importance. So, *of course* we should seek a mechanistic-causal understanding, including relevant cognitive neuroscientific mechanisms, that explains disorder symptoms. Such an analysis is totally consistent with, encouraged by, and complementary to the HDA.

It is true that often we can tell from superficial symptoms that something is going wrong with biologically designed processes, thus that there exists some dysfunction, long before we have the slightest idea of the nature of the relevant mechanisms and their functions and dysfunctions. Thus, it may misleadingly appear that evolutionary theorizing can proceed without the need for understanding of underlying neural mechanisms, when in fact an eventual combination of evolutionary and mechanistic-causal explanation is essential to understanding psychopathology.

In discussing delusions and other mental disorders, Gerrans repeatedly uses the language of dysfunction of biologically designed mechanisms ("dopamine dysfunction"; "dopamine dysregulation"; "hypersalience"; "abnormal"; "aberrant activity"), clearly placing his discussion within the domain of psychopathology. However, given that all neurocognitive functioning, whether normal or disordered, can be mechanistically characterized, what makes a neuroscientifically analyzed condition pathological and thus justifies applying the disorder/nondisorder distinction to the described conditions? Gerrans's notion that the discovery of mechanical causal principles is "informed by" evolution misses the fundamental nature of the biological-design analysis, which is to justify disorder attribution. There is no way to draw the disorder/nondisorder distinction among mechanistically described systems without an evolutionary perspective that allows one to judge whether something has gone wrong with their biologically designed functioning.

For example, certain patterns of neuronal firing in the mouse brain cause a male mouse to sexually mount a receptive female, certain patterns of firing of nearby neurons cause the same mouse to attack an encroaching male, and yet a third pattern of firing involving an overlapping region causes the mouse to attack a receptive female

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

(Anderson 2012; Lin et al. 2011). These are all equally good mechanical neuroscientific elaborations of causal relations. However, two of them are plausibly normal behavior and the other is plausibly pathological. To make this distinction, an evolutionary perspective on biological design must be imposed on the causal grid. Gerrans (and Murphy) ignore this additional step and thus cannot turn a mechanical-causal theory into a medical conceptualization of disorder versus nondisorder. (For further comments on this point, see my reply to Murphy in this volume.)

In a recent article (see also his chapter in this volume), Dominic Murphy seems to edge toward recognizing that a sheerly neurobiological causal description does not by itself validly distinguish disorder from nondisorder and that some additional factual criterion is needed:

> There is an important sense in which diagnoses cannot be validated at all, if by "validation" we mean "shown to be a real disorder." All validation can do is show that a pattern of behaviour deemed to be clinically significant depends on a physical process. Whether that pattern of behaviour is really pathological—rather than immoral or harmlessly odd—is another matter. … It requires that judgements of pathology be like findings of positive charge, i.e. scientifically grounded, rather than judgements of ugliness, i.e. human responses. If so, there has to be some natural fact of the matter about whether some physical system is dysfunctional. If this cannot be done, then predictions about physical states can be validated, but disorders cannot be. (Murphy 2017, 4–5)

The solution to the conundrum posed by Murphy in this passage—the need for an extra fact beyond all the neuroscientific facts to warrant a scientifically objective attribution of dysfunction—is simply that a dysfunction is the failure of a naturally selected function. This is a historical fact about the biological design of the described neurobiological system in relation to its current performance, so it is an additional fact beyond all the cross-sectional neurobiological descriptive facts. Indeed, of two identically describable mechanical systems, one can be properly functioning and the other dysfunctioning given divergent evolutionary histories (Wakefield 1999a). The HDA and neuroscientific mechanical-causal elucidation are necessarily complementary elements in the analysis of mental disorder.

Neuroscience versus Evolutionary Psychology in the Quest for Dysfunctions

Although receptive to bringing evolutionary considerations into the theory of mental disorder, Gerrans notes that there are various ways one might do this, and he has several concerns about what he thinks is the HDA's specific approach to evolutionary explanation. Gerrans apparently equates the HDA's evolutionary understanding with two approaches that he calls "evolutionary psychology" and "strong adaptationism" and rejects both of the latter views. Additionally, he is particularly skeptical of evolutionary psychological explanations of pathological belief fixation.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

First, then, regarding evolutionary psychology, both in his paper in this volume and in his broader work (e.g., Gerrans 2002, 2007; Gerrans and Stone 2008; Stone and Gerrans 2006), Gerrans emphasizes the fallibility of commonsense inferences shaped by folk-psychological constructs from the nature of overt symptoms to the postulation of dysfunctions in specific dedicated neurocognitive mechanisms that fit our folk-theoretic schemas. Thus, it is all too easy to postulate "sadness regulating mechanisms," "jealousy regulating mechanisms," "self-esteem regulating mechanisms," and so on. However, the brain works its wonders in mysterious ways that evolved under obscure fitness pressures constrained by unknown earlier adaptations, so, Gerrans argues, surface psychological phenomena are often no royal road to elucidating the structure of the deeper levels of processing that give rise to them.

Gerrans's prototypical example of the evolutionary psychology fallacy is the inference from autistic individuals' difficulty understanding others' intentional states to the postulation of a dedicated "theory of mind" module that is dysfunctional, when in fact the difficulty may be due to dysfunctions in deeper and much more general neurocognitive mechanisms that happen to manifest in theory-of-mind difficulties. It is the hypothesizing of such symptom-close dedicated cognitive modules formulated in terms of folk-psychological experienced variables that Gerrans labels "evolutionary psychology." He contrasts evolutionary psychology with explanation in terms of a combination of both deeper neuroscientific mechanisms (e.g., the salience mechanisms he describes at length that might amplify the salience of certain ideas to the point of delusion) and perceptual-surface neuroprocessing levels (e.g., automatic facial interpretation mechanisms that might bias toward seeing disapproval and thus toward depression) that he considers to be less folk-psychologically inspired and to constitute a more scientific "cognitive neuroscience." (I tend to understand "evolutionary psychology" less narrowly as a general discipline that encompasses all evolutionary understanding of psychological processes, but I will stick to Gerrans's usage here.)

Gerrans rebukes evolutionary psychology, and (I surmise) by implication the HDA, as follows:

> It will, however, rarely be the case that a psychiatric disorder can be explained in terms of the (mal)function of a single well-defined cognitive adaptation. Patterns at the surface level, of behavior, belief, or experience, rarely directly reflect the operations of a single domain-specific cognitive system. More often, such patterns result from a cascade (often a developmental cascade) of events that ramify through the cognitive system.

Any identification by Gerrans of the HDA with his target of evolutionary psychology in the above passage is a mistake. There is nothing in the HDA that says that the cognitive adaptations that constitute psychological functions must be "single, well-defined" or "single, domain-specific" adaptations or mechanisms, nor that the nature of mechanisms and their functions and dysfunctions must be able to be specifiable by directly

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

reading them off from the "surface level, of behavior, belief, or experience." There is nothing in the HDA that precludes that a dysfunction can result from or even be constituted by a "cascade … of events." (That comment by Gerrans seems to leave the door open to something going wrong with salience mechanisms that percolates higher and causes a more specific dysfunction in a particular "evolutionary psychology" middle-level regulating system.) I have argued elsewhere that dysfunctions can even occur in pathological-level interactions between mismatched individual mechanisms that are each individually performing within their normal ranges (Wakefield 1999a, 2006b). The HDA only asserts that to be a disorder, the explanation must attribute the harmful symptoms to *some* dysfunction of biologically designed mechanisms.

My extensive explanations of why dyslexia is considered a disorder despite reading not being selected for make clear that there need be no commonsense relationship between the function of the inferred dysfunctional mechanism and the nature of the consequent symptoms, other than a causal one. An inference from symptoms to the presence of a dysfunction that supports a provisional disorder judgment can be so weak as to be a sheerly existential hypothesis that a dysfunction exists in some mechanism, based on circumstantial evidence and remaining noncommittal on the kinds of mechanisms, functions, and dysfunctions involved. After that, it is scientific open season to theorize about the nature of the mechanisms, their function, and their dysfunctions. This takes time; in physical medicine, it was over 2,000 years from the time Hippocrates inferred a series of diagnostic categories to a scientific understanding of the mechanisms, functions, and dysfunctions underlying many of his categories. A measure of the surprising strength of the circumstantial evidence that supports such dysfunction inferences is that Hippocrates's speculative theories about etiology were wildly wrong, yet virtually every category he baptized has turned out to consist of what we still consider genuine disorders, although of course reorganized and relabeled as etiological knowledge increased.

Gerrans is surely correct that the facile inference from symptoms to the existence of a certain type of underlying biologically designed domain-specific middle-range modular brain mechanism and a certain kind of dysfunction of that mechanism can all too often be an unsupported projection of folk-theoretic notions, giving rise to infamous "just-so stories." For this reason, I tend to exert restraint and entertain skepticism about many evolutionary psychological explanations until adequate potentially falsifying testing occurs, and this is one reason why, throughout my work, I rarely discuss or endorse specific evolutionary psychological or cognitive neuroscientific hypotheses. Yet, it should also be kept in mind that the "just-so story" problem is a general one, and most science-based theorizing initially gets things wrong. Clever and ruthless potentially falsifying testing of theories leading to revisions that correct detected errors, not a priori specification of what kinds of hypotheses are allowable, is what makes science powerful and progressive.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

If there is the sort of rivalry between "evolutionary psychology" and "cognitive neuroscience" that Gerrans portrays, the HDA is neutral on the outcome. The HDA mainly predicts the relationship between background beliefs about causation and consequent disorder attribution (see my reply to De Vreese in this volume for further elaboration of this point). Thus, in arguing for the HDA, I generally attempt to show that beliefs about biological design shape judgments about disorder versus nondisorder. So, my examples often cite current theories as examples of background beliefs. However, the theories' correctness or incorrectness is not relevant, except when I go beyond conceptual analysis and argue for substantive claims about what is and is not a disorder, as Allan Horwitz and I did in arguing for the invalidity of *DSM*'s major depression category in *The Loss of Sadness* (2007). Thus, whether or not the theory-of-mind module theory of autistic pathology turns out to be correct, it supports the HDA because those who believe the theory-of-mind modular account believe that autism is the harmful effect of a dysfunction in the hypothesized biologically designed module and consequently judge it a disorder. Equally supportive of the HDA are "neurodiversity" accounts of autism that are opposed to the "theory-of-mind" account, because they deny that autism is a disorder and thus are forced to argue that it is not due to a dysfunction (see my reply to Forest in this volume for a discussion of neurodiversity and autism from the HDA perspective). The HDA predicts not which theory is correct but that the judgment of disorder is consistently rationalized by a corresponding theory about biological design.

Sometimes, when discussing dysfunction hypotheses in a conceptual-analytic context, I might use abstract descriptions of postulated underlying mechanisms that are meant to be neutral on the kinds of mechanisms involved. For example, I might attribute major depression to a dysfunction of "sadness-generating mechanisms." This could easily be misconstrued as a commitment to middle-level dedicated mechanisms as the locale of the primary dysfunction. True, I do believe that in the case of sadness and other major emotions, there likely are such middle-level mechanisms, and their dysfunctions do play a role in emotion-related mental disorders. However, in such locutions, I mean to refer to *whatever* mechanisms are responsible for generating and regulating sadness responses (given the assumption that such basic emotions do have an evolutionary undergirding), whatever they may turn out to be. Gerrans makes the entirely correct point that, even if such mechanisms exist, it need not be that this is where the deepest or most crucial dysfunction is occurring in depression. Gerrans suggests instead that dysfunctional deeper neurocognitive processes may be feeding problematic information into a middle-range module, causing it to no longer function within the parameters for which it was biologically designed. If so, I agree that the etiology and diagnosis would have to reflect this deeper level of dysfunction that is disrupting downstream functions, perhaps yielding his proposed megacategory of "disorders of the salience system." Yet, Gerrans mentions that delusions tend to form into groups (e.g., grandiosity, jealousy, paranoia), suggesting the need for an explanation

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

of this clumping in terms of middle-level adaptive mechanisms downstream from the salience system.

Neuroscience, Strong Adaptationism, and the HDA

Gerrans continues with a critique of what he calls "strong adaptationism":

> For this reason, strong adaptationist versions of evolutionary psychiatry are unlikely to succeed in directing us to cognitive or neural mechanisms. They are too "top down" in their analysis of the problem. This is obvious in the case, for example, of theories that postulate an adaptive problem for which schizophrenic delusions represent a solution (Stevens and Price 2000; Dubrovsky 2002), but the point generalizes.

However, I am not a "strong adaptationist" as that position is generally understood, and neither is strong adaptationism equivalent to the middle-range modular "evolutionary psychology" approach to which Gerrans seems to equate it. Strong adaptationists think that virtually all human features can be understood as adaptations and tend to explain disorders as adaptations to circumstances in the environment of evolutionary adaptation (EEA). I think, to the contrary, that things really can go wrong with almost any biologically designed system in ways that were never biologically designed to happen, and so there are dysfunctions that are not part of design, and those are what conceptually undergird the category of disorder. Thus, I claim that strong adaptationism about medical disorder is not only false but conceptually incoherent. In various publications, I have explained the fallacies involved in strong adaptationists' attempts to explain disorders as adaptations (Wakefield 2016; also see my reply to Cooper in this volume).

In any event, as Gerrans observes, the interestingly complex explanatory mechanisms identified by cognitive neuroscience are presumably not accidents but quite design-like and thus, barring alternative "spandrel"-type explanations, may be presumed to be biologically designed forms of brain circuitry. Cognitive neuroscience is adaptationist "all the way through" not as an ideology but as an empirical claim. As to Stevens and Price, their attempt to explain the prevalence of psychotic symptoms as the result of natural selection for charismatic leaders initially may appear to be an example of strong adaptationism about disorder, but a careful reading reveals that they in fact explain selection only for risk factors that are themselves adaptive but that in certain unselected combinations yield disorder (for further discussion of Stevens and Price, see my reply to De Block and Sholl in this volume).

Neuroscience, Belief Fixation, and the HDA

It appears that a further reason that Gerrans sees a tension between the HDA and mechanistic explanation is that he understands *DSM* diagnoses as well as the HDA as

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

concerned with contents like beliefs, and he sees belief fixation as not subject to direct neuropsychological explanation in terms of naturally selected mechanisms. Writing of the very inadequate *DSM* criteria for delusion, he says,

> It is especially difficult to preserve a taxonomic role for evolutionary theory in the case of psychiatric disorders like delusion, whose classification involves the concept of belief. … The difficulty with all these proposals is not that they get the phenomena wrong but that they all rely on the language of belief, which is intrinsically agnostic about mechanisms.

However, these proposals are purposely agnostic because, rather than attempting to identify mechanisms, they are elaborating the surface descriptive phenomenology of delusions that tells us that something is going wrong *somewhere*, thus identifying the phenomenon that requires explanation by underlying mechanisms.

Moreover, the language of belief is not intrinsically agnostic about mechanisms describable at that level. Although belief mechanisms must have realizations in brain mechanisms, the belief-system level can possess emergent mechanisms that may have then exerted natural selection force on underlying brain mechanisms. Jerry Fodor observes, "Roughly, if you start out with a true thought, and you proceed to do some thinking, it is very often the case that the thoughts that the thinking leads you to will also be true. This is, in my view, the most important fact we know about minds" (Fodor 1994, 9). He thus refers to a mechanism at the belief level (or an idealization of a mechanism), namely, valid reasoning from premises to conclusion. One does not have to understand deeper processes to formulate a provisional theory of how the reasoning mechanism works at the belief level, as Aristotle already attempted.

Gerrans holds this position on belief fixation because he thinks that, although humans are adapted to engage in belief-related ideational behaviors, the relevant mechanisms belong to lower-level neurocognitive functioning rather than the ideational level:

> The human cognitive phenotype does show entrenched patterns of belief fixation in specific domains, and some psychiatric disorders can be characterized in terms of typical abnormalities in those patterns (e.g., social cognition is a specific domain and autism is characterized by deficits of belief fixation in that domain). However, those patterns are produced by low-level neurocognitive mechanisms that produce an upward cascade of effects ultimately expressed as patterns of belief fixation. It makes no sense to see these neurocognitive mechanisms … as cognitive mechanisms selected for forming particular classes of beliefs. … Rather, they are mechanisms that enable processing of information at specific levels of cognitive complexity, and we can make progress on determining their nature by tracing their evolutionary history at the correct level of cognitive resolution.

I cannot find in Gerrans's paper a cogent defense of this unlikely thesis. He is suggesting that, say, pathological jealousy involves neither dysfunction in any belief-close dedicated jealousy-belief-fixation mechanism or process, nor grandiosity in a

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

belief-close dedicated self-esteem belief-fixation mechanism or process, nor depressive hopelessness in a belief-close loss-response belief-fixation mechanism or process. Of course, as Gerrans explains, belief fixation involves complex contextual background understanding and thus certainly interacts with general cognitive processing of the belief system that is not strictly modularized. But, that generic background processing does not preclude final-pathway canalization by specific mechanisms that account for the recognizable distinctions among primary emotions or major areas of cognition. The situation is no different in physical functioning; liver function involves a complex cascade of general bodily functions such as blood circulation and homeostatic thermoregulation to provide an appropriate context, but that doesn't mean that there are no liver-specific mechanisms or that there is no such thing as a liver disorder.

Neuroscience, the *DSM-5*, and the HDA

Regarding the implications of cognitive neuroscience for psychiatric nosology, Gerrans concludes the following:

> Once we abandon the strong adaptationist approach in favor of a cognitive neuroscience informed by evolutionary theorizing, psychiatric classification requires substantial revision. Wakefield's approach, I suggest, ultimately leads to abandoning the current *DSM* approach in favor of that that recommended by Dominic Murphy (2006)—namely, to allow classification and psychiatric practice to reflect the architecture of the mind disclosed by cognitive neuroscience.

The supposed conflict suggested by this passage between *DSM* and the neuroscientific elucidation of mental disorder etiology is a strawman. Gerrans's suggestion that attention to neuroscientific explanations will correct a deep flaw in *DSM* of addressing only a more superficial level ignores the circumstances in which *DSM*'s modern approach came about and the understanding of its project within which it was born.

The nosological problem facing psychiatry before *DSM-III*, among many other serious problems such as diagnostic unreliability and reliance on psychoanalytic constructs, was not that too little attention was being paid to causal mechanisms by psychiatrists but that too many unestablished causal theories prematurely were being taken seriously, and thus there was a fragmentation of the field along theoretical lines. Every one of those theoretical approaches, ranging from neurobiological, behavioral, and cognitive theories to social stress and family dynamics theories as well as five flavors of psychoanalytic theory, felt they had fresh, illuminating, scientifically valid insights into the sources of mental disorder just as neuroscientists do now. Research was pursued by each school using different diagnostic criteria, so research samples could not be compared and the knowledge base was not cumulative. *DSM-III's* solution was to propose criteria that were agreed on across theoretical perspectives to identify classes

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

of mental disorders and thus create a level playing field for the demonstration of etiological hypotheses. There are many problems with the *DSM* system and how it has evolved, and I have spent a good deal of effort pointing to some of them. However, in no way is the system meant to be some sort of syndromal conceptual account of the nature of mental disorder (contrary to Murphy's [2017] portrayal). The syndromes are markers for likely underlying dysfunction (First and Wakefield 2013), and, as the *DSM*'s definition of mental disorder makes clear, it is causation of harmful symptoms by a dysfunction that makes a condition a disorder. Thus, the goal of the *DSM* is to provide the initial step in bootstrapping to eventually cash out the reference to an inferred dysfunction for actual knowledge of etiological dysfunctions. The *DSM*'s logic does not dictate or prejudge the nature of those dysfunctions.

But you don't need to take my word for it; precisely this point of the provisional nature of *DSM* symptomatic criteria awaiting etiological understanding was explained by Robert Spitzer in his introduction to *DSM-III* at the inauguration of the current descriptive system:

> *Descriptive Approach*. For some of the mental disorders, the etiology or pathophysiological processes are known. . . . For most of the *DSM-III* disorders, however, the etiology is unknown. A variety of theories have been advanced, buttressed by evidence—not always convincing—to explain how these disorders come about. The approach taken in *DSM-III* is atheoretical with regard to etiology or pathophysiological process except for those disorders for which this is well established and therefore included in the definition of the disorder. Undoubtedly, with time, some of the disorders of unknown etiology will be found to have specific biological etiologies, others to have specific psychological causes, and still others to result mainly from a particular interplay of psychological, social and biological factors.
>
> The major justification for the generally atheoretical approach taken in *DSM-III* with regard to etiology is that the inclusion of etiological theories would be an obstacle to use of the manual by clinicians of varying theoretical orientations. . . . For example, Phobic Disorders are believed by many to represent a displacement of anxiety resulting from the breakdown of defensive operations for keeping internal conflict out of consciousness. Other investigators explain phobias on the basis of learned avoidance responses to conditioned anxiety. Still others believe that certain phobias result from a dysregulation of basic biological systems mediating separation anxiety. . . . Clinicians can agree on the identification of mental disorders on the basis of their clinical manifestations without agreeing on how the disturbances come about. (Spitzer 1980, 6–7)

The *DSM* states that it is atheoretical—that is, it does not include a theory of etiology in the diagnostic criteria—because "for most of the *DSM-III* disorders, however, the etiology is unknown." This is a pragmatic, epistemologically based compromise, not a conceptual or ontological statement about the concept of disorder. This atheoretical stance, we are told, applies "except for those disorders for which [etiology] is well established and therefore included in the definition of the disorder." The clear

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

implication of this passage is that symptom syndromes, as in physical medicine, are understood not as ontological foundations but as transient epistemological necessities until etiological knowledge is established, at which time etiological criteria will supplement or replace syndromal criteria. Syndromes are the best that can be done at present to individuate disorders, but they are intended as provisional indicators on the way to more scientifically sophisticated and validated etiological disorder identification. Spitzer's example of multiple theories of phobia illustrates that the problem is not lack of attempts to identify causes but rather lack of established knowledge. The situation has not changed all that much today, to the consternation of many in the field.

Gerrans again cites Murphy in support of his claim that the *DSM* embraces a syndromal conception of disorder that needs to be overthrown for a causal conception. So, I briefly consider Murphy's view of *DSM*. It is useful to compare the above statement from *DSM-III* explaining its descriptive nosology with the following excerpts from Murphy's account of what he calls "the *DSM* Conception of Mental Illness" (references are deleted; consult the original for them):

> The *DSM* treats mental disorders as syndromes. … The previous version of the *DSM* assumed that each diagnosis represented malfunction in some mental, physical or behavioural trait or capacity (*DSM-IV-TR*, xxi). However, the diagnoses were listed without worrying about what that underlying malfunction might be, and in most cases there was (and remains) no agreement about what causes what. *DSM-5* defines mental disorders as syndromes comprising clinically significant disturbances of cognition, emotion or behaviour that reflect underlying dysfunctions. These … cannot be diagnosed if the behaviour is culturally normal or merely socially deviant, unless it reflects a dysfunction. … There are plenty of students of psychopathology who argue that the neglect of causal structure in psychopathology is getting in the way of science. …
>
> The *DSM* approach is often called "neo-Kraepelinian." But Kraepelin … saw classification by clinical description as an interim measure. … Kraepelin's preferred basis for classification and inquiry actually rested on his less well-remembered belief that "pathological anatomy promises to provide the safest foundation" for classification of mental illness in a mature psychiatry. He considered the correct taxonomy would be one in which clinical description, etiology and pathophysiology coincided. …
>
> There is a substantial difference between thinking of clinically-based, syndromic classification in this way and thinking of it as the *DSM* does. The *DSM* classification … is not advertised as the jumping-off point for a mature system of causally organised classification and practice. (Murphy 2017, 5)

Murphy concludes that the *DSM* is committed to syndromal classification and does not see itself as a starting point for development of a more mature etiologically based causal system of diagnoses. As we saw, *DSM-III* clearly indicates precisely the opposite; atheoretical definitions are to be used "except for those disorders for which [etiology] is well established," and it is expected that "with time, some of the disorders of

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

unknown etiology will be found to have specific… etiologies" and then the etiology will be "included in the definition of the disorder." In his next-to-last sentence, Murphy asserts that Kraepelin's view that syndromes are used for initial classification as a stepping stone to etiological discovery is a substantially different way of thinking about syndromal diagnosis than the *DSM*'s, but in fact it is identical to the view enunciated by Spitzer in *DSM-III*. Murphy explains that Kraepelin sees "classification by clinical description as an interim measure," and, contrary to Murphy's portrayal, *DSM-III* sees it in exactly the same way. Kraepelin's model was general paresis, initially defined by syndrome and course, gradually distinguished phenomenologically from symptomatically similar conditions through syndromal analysis, and then, when eventually it was discovered to be caused by syphilitic brain infection, the etiology replaced the syndromal diagnostic approach. The hope was that *DSM*-based diagnosis would give rise to similar progress. The *DSM*, like Kraepelin, uses syndromes to try to pick out unknown dysfunctions, and the different possible symptom presentations of a given disorder, sometimes nonoverlapping, are united in being thought to pick out the same or a similar dysfunction. The quest to identify dysfunctions also explains many other features of the *DSM*'s diagnostic criteria sets, such as number of symptoms and durational thresholds (First and Wakefield 2013) and contextual exclusions (Wakefield and First 2012). All these features are aimed at trying to align the criteria with the target dysfunction(s).

Murphy's account confuses the epistemology of diagnosis with the ontology of mental disorder. The problem seems to be in part a matter of the limitations of Murphy's theory of concepts. As the above passage indicates, Murphy tends to see either syndromes or explicit, known causal etiologies as exhausting conceptual logic. Thus, he can ignore the *DSM* "dysfunction" clause because "the diagnoses were listed without worrying about what that underlying malfunction might be, and in most cases there was (and remains) no agreement about what causes what." This misses the basic point that a syndrome is being used to pick out an unknown but inferred dysfunction, the presence of which is conceptually essential, and this is precisely the conceptual feature that looks beyond the syndrome and awaits elucidation as the etiology of the symptoms. This structure is analogous to what I call "black-box essentialist" concepts, as elaborated in psychology by Medin and Ortony (1989) and in philosophy by Putnam (1975) and Kripke (1980), in which an essential factor unifying a category is unknown but is picked out by way of an observable "base set" of instances. The necessity of the reference to a known or inferred dysfunction in an analysis of "medical disorder" is a crucial insight that took Spitzer half a decade to come to, and without it, one gets the watered-down implausible syndromal view of disorder that Murphy mistakenly attributes to the *DSM*.

Murphy reports that many writers criticize the *DSM* for a "neglect of causal structure in psychopathology," but neglect is different from lack of scientific success despite great effort. *DSM-III* does not neglect causal structure; it is a compromise with the fact

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

that we don't yet know causal structures. This is illustrated by Spitzer's phobia example; those who criticize the *DSM* for neglect of etiology are not in agreement about what the appropriate causal structures should be. If Gerrans (or anyone else) were actually to present psychiatry with serious *consensually persuasive* scientific evidence for mental disorder etiology—mere excitement at novel proposals by a subdiscipline's enthusiasts after some initial testing is *not* the same as established, persuasive scientific fact—the *DSM-5.1* committee would likely jump at it.

Murphy's skewed reading of the *DSM* leads to the sort of misunderstanding we saw in Gerrans. Rather than opposing or radically altering the *DSM*'s program, neuroscientific theories of disorders would advance and even vindicate the *DSM* program of starting from syndromes and bootstrapping to underlying dysfunctions and reorganized etiologically coherent syndrome categories. In fact, the *DSM-5* Task Force completely *agreed* with Gerrans and Murphy that we should move beyond the descriptive system, and early on, they infamously declared that *DSM-5* would be a "paradigm shift" (cf. Gerrans's "radical" change) in favor of brain circuitry and biomarkers, with eventual rectification with RDoC sure to follow. So, why didn't it happen? The goal was abandoned late in the revision process simply because there were no adequately confirmed brain-mechanism theories of disorder to put into the manual as diagnostic criteria. (Yes, there are lots of findings of group-level differences, but not ones that adequately distinguish disorder from normality to be used diagnostically.) This humiliating and disruptive late admission of failure could have been avoided by a good literature search at the outset of the *DSM-5* process, but the salience of the vision of what would be desirable blinded the task force to seeing what is.

Rather than disagreeing with *DSM*, Gerrans and Murphy are agreeing with *DSM* aspirations but making the same mistake of confusing wishes and aspirations with current reality. The problem is with reality, not with *DSM* doctrine. Despite Gerrans and Murphy being gripped by the enormous salience of advances in neurocognitive research, sobering pitfalls likely lie ahead for cognitive neuroscientific explanations of psychopathology as they did for every earlier salient research approach (Paulus and Thompson 2019). Gerrans's own theory of delusion as "salience overshoot" suggests that scientific ardor may be a constructive quasi-delusion and that caution is warranted. The difficult truth is that we are just not there yet.

References

First, M. B., and J. C. Wakefield. 2010. Defining 'mental disorder' in *DSM-V*. *Psychological Medicine* 40(11): 1779–1782.

First, M. B., and J. C. Wakefield. 2013. Diagnostic criteria as dysfunction indicators: Bridging the chasm between the definition of mental disorder and diagnostic criteria for specific disorders. *Canadian Journal of Psychiatry* 58(12): 663–669.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

Fodor, J. A. 1994. *The Elm and the Expert: Mentalese and Its Semantics*. MIT Press.

Gerrans, P. 2002. The theory of mind module in evolutionary psychology. *Biology and Philosophy* 17(3): 305–321.

Gerrans, P. 2007. Mechanisms of madness: Evolutionary psychiatry without evolutionary psychology. *Biology and Philosophy* 22(1): 35–56.

Gerrans, P., and K. Scherer. 2013. Wired for despair: The neurochemistry of emotion and the phenomenology of depression. *Journal of Consciousness Studies* 20(7–8): 254–268.

Gerrans, P., and V. E. Stone. 2008. Generous or parsimonious cognitive architecture? Cognitive neuroscience and theory of mind. *British Journal for the Philosophy of Science* 59(2): 121–141.

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

Kripke, S. A. 1980. *Naming and Necessity*. Harvard University Press.

Levy, N. 2017. Hijacking addiction. *Philosophy, Psychiatry, and Psychology* 24(1): 97–99.

Medin, D. L., and A. Ortony. 1989. Psychological essentialism. In *Similarity and Analogical Reasoning,* S. Vosniadou and A. Ortony (eds.), 179–196. Cambridge University Press.

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

Murphy, D. 2017. Philosophy of psychiatry. In *The Stanford Encyclopedia of Philosophy,* 1–31, E. N. Zalta (ed.). https://plato.stanford.edu/archives/spr2017/entries/psychiatry/. April 7, 2019.

Paulus, M. P., and W. K. Thompson. 2019. The challenges and opportunities of small effects: The new normal in academic psychiatry. *JAMA Psychiatry* 76(4): 353–354.

Putnam, H. 1975. The meaning of meaning. In *Mind, Language, and Reality: Philosophical Papers,* H. Putnam (ed.), 215–271, vol. 2. Cambridge University Press.

Spitzer, R. L. 1980. Introduction. In *Diagnostic and Statistical Manual of Mental Disorders,* 1–12. 3rd ed. American Psychiatric Association.

Spitzer, R. L. 1997. Brief comments from a psychiatric nosologist weary from his own attempts to define mental disorder: Why Ossorio's definition muddles and Wakefield's "harmful dysfunction" illuminates the issues. *Clinical Psychology: Science and Practice* 4(3): 259–261.

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

Stone, V. E., and P. Gerrans. 2006. What's domain specific about theory of mind. *Social Neuroscience* 1(3–4): 309–319.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20: 410–433.

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

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

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

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

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

Wakefield, J. C. 1997a. Diagnosing *DSM-IV*, part 1: *DSM-IV* and the concept of mental disorder. *Behaviour Research and Therapy* 35: 633–650.

Wakefield, J. C. 1997b. Diagnosing *DSM-IV*, part 2: Eysenck (1986) and the essentialist fallacy. *Behaviour Research and Therapy*: 35: 651–666.

Wakefield, J. C. 1997c. Normal inability versus pathological disability: Why Ossorio's (1985) definition of mental disorder is not sufficient. *Clinical Psychology: Science and Practice* 4: 249–258.

Wakefield, J. C. 1997d. When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder. *Development and Psychopathology* 9: 269–290.

Wakefield, J. C. 1998. The *DSM*'s theory-neutral nosology is scientifically progressive: Response to Follette and Houts. *Journal of Consulting and Clinical Psychology* 66: 846–852.

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

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

Wakefield, J. C. 2000a. Aristotle as sociobiologist: The "function of a human being" argument, black box essentialism, and the concept of mental disorder. *Philosophy, Psychiatry, and Psychology* 7: 17–44.

Wakefield, J. C. 2000b. Spandrels, vestigial organs, and such: Reply to Murphy and Woolfolk's "The harmful dysfunction analysis of mental disorder." *Philosophy, Psychiatry, and Psychology* 7: 253–269.

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

Wakefield, J. C. 2006. What makes a mental disorder mental? *Philosophy, Psychiatry, and Psychology* 13: 123–131.

Wakefield, J. C. 2007. The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry* 6: 149–156.

Wakefield, J. C. 2009. Mental disorder and moral responsibility: Disorders of personhood as harmful dysfunctions, with special reference to alcoholism. *Philosophy, Psychiatry, and Psychology* 16: 91–99.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

Wakefield, J. C. 2011. Darwin, functional explanation, and the philosophy of psychiatry. In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory,* P. R. Andriaens and A. De Block (eds.), 143–172. Oxford University Press.

Wakefield, J. C. 2014. The biostatistical theory versus the harmful dysfunction analysis, part 1: Is part-dysfunction a sufficient condition for medical disorder? *Journal of Medicine and Philosophy* 39: 648–682.

Wakefield, J. C. 2016a. The concepts of biological function and dysfunction: Toward a conceptual foundation for evolutionary psychopathology. In *Handbook of Evolutionary Psychology,* D. Buss (ed.), 2nd ed., vol. 2, 988–1006. Oxford University Press.

Wakefield, J. C. 2016b. Diagnostic issues and controversies in *DSM-5*: Return of the false positives problem. *Annual Review of Clinical Psychology* 12: 105–132.

Wakefield, J. C., and M. B. First. 2003. Clarifying the distinction between disorder and nondisorder: Confronting the overdiagnosis ("false positives") problem in *DSM-V*. In *Advancing DSM: Dilemmas in Psychiatric Diagnosis*, K. A. Phillips, M. B. First, and H. A. Pincus (eds.), 23–56. American Psychiatric Press.

Wakefield, J. C., and M. B. First. 2012. Placing symptoms in context: The role of contextual criteria in reducing false positives in *DSM* diagnosis. *Comprehensive Psychiatry* 53: 130–139.

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

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

Defining Mental Disorder

Jerome Wakefield and His Critics

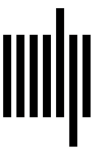
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl

Edited by: Luc Faucher, Denis Forest

Citation:

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

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



The MIT Press

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024

MIT Press Direct

© 2021 Massachusetts Institute of Technology

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

Subject to such license, all rights are reserved.

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

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

Library of Congress Cataloging-in-Publication Data

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

Downloaded from http://direct.mit.edu/books/oa-edited-volume/chapter-pdf/2268329/9780262362931_c001900.pdf by guest on 13 October 2024