22 Do the Challenges of Autism and Neurodiversity Pose an Objection to the Harmful Dysfunction Analysis? Reply to Denis Forest

Jerome Wakefield

In his chapter in this volume, my friend Denis Forest presents a provocative overview of the increasing complexities and controversies surrounding the diagnosis of autism, including the neurodiversity movement's arguments against psychiatric labeling of autism as a disorder. He argues that in considering these developments, difficulties lie in wait for my harmful dysfunction analysis (HDA) of medical, including mental, disorder (First and Wakefield 2010, 2013; Spitzer 1997, 1999; Wakefield 1992a, 1992b, 1993, 1995, 1997a, 1997b, 1997c, 1997d, 1998, 1999a, 1999b, 2000a, 2000b, 2001, 2006, 2007, 2009, 2011, 2014, 2016a, 2016b; Wakefield and First 2003, 2012), especially regarding the key question of whether autism should be understood as normal variation or disorder. He argues that the "data" of the autism controversy, although not fundamentally disconfirming the HDA, require modification of both its dysfunction and harm components: "In a nutshell, I shall advocate a more mechanistic view of dysfunctions and an understanding of harm in terms of diminished ability." I will consider Forest's critique of the dysfunction criterion and most of his concerns about the harm requirement. I'll address other harm-related concerns in my reply to Cooper.

The diagnosis of autism emerged from observations of a triad of syndromally associated severe symptoms, including impairment of social development, impairment of communication, and display of rigid and repetitive behavior. The *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5)* officially expanded classic autism into an autistic spectrum of conditions varying in severity, engulfing the former milder diagnosis of Asperger's disorder. *DSM-5* also reduced the triad to a dyad of dimensions by combining impairment of social communication and impairment of social development into one overarching dimension of deficits in social communication and social interaction.

Forest observes that autism was recognized as a developmental disorder from the earliest days of its identification, but he does not try to explain why it seemed, and still seems, so obvious to almost everyone (other than those arguing for the most extreme neurodiversity position) that the initially identified severe condition—which, like Forest, I will refer to as "classic autism"—is a disorder rather than an unusual variant

of normality. (I will return to this question below.) Instead, Forest focuses on recent developments that, he says, have clouded the initial picture of clear disorder. Newly recognized phenomena such as the autistic spectrum (Wing 1997), high-functioning autism with preserved general intellectual abilities, the broader autistic phenotype (BAP) consisting of personality traits that are mild versions of autistic-like symptoms, subthreshold cases satisfying one rather than both autistic dimensional criteria, and isolated but sometimes quite dramatic special talents in otherwise seriously impaired autistic individuals complicate the classic picture and make the category of autism increasingly problematic as to disorder status, he suggests.

It is this complex of autistic conditions that, Forest thinks, poses a challenge to the HDA. Based on his account of autism and the neurodiversity movement, Forest lodges three objections to the HDA. First, its reliance on the evolutionary model of dysfunction fails to reflect how dysfunction is actually used by researchers. Second, its analysis in terms of dysfunction is unhelpful in guiding disorder attributions in difficult cases within the autism category. Along with these first two objections, Forest presents several subsidiary concerns about the dysfunction requirement, such as that it is subject to "just-so" stories and that it fails to help us appreciate the many varieties of normality. Third, Forest argues that the harm criterion is too narrow in virtue of its being mistakenly linked to social values—a point that, as noted, I will partly address here and return to elsewhere in this volume. I focus here mainly on Forest's two objections to the "dysfunction" requirement as opposed to the "causal role" approach preferred by Forest, as well as his subsidiary criticisms of the dysfunction requirement.

First, then, Forest challenges whether HDA's evolutionary approach actually guides research on autism: "Is research about autism and its explanation concerned with the discovery of dysfunctional mechanisms, where 'dysfunctional' has the precise meaning that is attached to it within the framework of HD analysis?" Forest surveys three of the main theories of autism—mindblindness, executive dysfunction, and weak coherence—and concludes that only mindblindness theorists explicitly refer to evolutionary considerations. He thus concludes, "The answer … is negative: in the literature, with few exceptions, 'dysfunction' of psychological or neural mechanisms is not understood as their failure to perform what they have been 'designed' to do."

If not a failure of biologically designed function, what, then, is a dysfunction? Forest suggests that what is at work instead of the HDA is a concept of dysfunction that combines Robert Cummins's (1975) causal-role model with a normative component that determines which causes are dysfunctions: "Cummins's view of functions (Cummins 1975) according to which the function *F* of a component *C* is its contribution to the explanation of a capacity of the system in which *C* is embedded, seems appropriate for psychological as well as physiological mechanisms"; "Moreover, factual judgements about impaired performance and dysfunction are usually inseparable from implicit evaluative claims." That is, for Forest, a dysfunction in an internal component of the

organism occurs when the component causes a species-atypical negative condition in the organism.

Forest's argument—that because some theorists who study autism do not explicitly couch their theories in terms of evolution, and therefore "dysfunction" must have no essential connection to failure of biological design—is invalid. True, some theories of autism are not explicitly evolutionary, but, even if the HDA is correct, why should they be? No reference to evolution is necessary for doing studies of the proximal causes and potential treatments of autism, which, as Forest observes, is all that matters from a practical perspective in most autism research. Even Tinbergen's (1963) list of four basic features of evolutionary explanation includes proximal causal explanation as a component. When researchers pursuing mechanistic causal understanding assume that autism is a disorder, something beyond a mechanistic understanding must be involved, for a mechanistic causal explanation can equally be given for disorders and nondisorders. That implicit additional assumption, the HDA claims, is that something has gone wrong with the organism in the sense that there is a failure of biological design. This implicit assumption need not be made explicit in most causal research.

To take an analogy, the science of water, hydrology, was around long before anyone understood let alone stated that water is H_2O. The Nile was dammed about 4000 BC for agricultural irrigation reasons, and other ancient civilizations, including the Greeks, Romans, and Chinese, manipulated water with irrigation canals, aqueducts, and flood-control structures. They did not have any trouble identifying, studying, and manipulating the liquid they were aiming to control despite not knowing it was H_2O. Theories of the water precipitation cycle existed in ancient times and began to be quantified in the seventeenth century, whereas the chemical constitution of water as two parts hydrogen and one part oxygen was identified by Henry Cavendish in 1781. Surely neuroscientists have similarly ample grounds for recognizing certain homologous structures as causative of autism and devising ways to intervene without knowing or explicitly stating the evolutionary history that explains the existence and natural function of that kind of mechanism, although to some degree, assumptions about natural functions are presupposed in theorizing about what is going wrong.

Against the evolutionary view, Forest raises the usual complaint that evolutionary explanations can be "just-so" stories. This is true of all theorizing; there are endless "just-so" stories in every domain of human thought, and sifting through the theories and establishing which is correct is precisely what science is about. In any event, this is an objection to being overly gullible in accepting superficially appealing evolutionary explanations, not an objection to the conceptual claim that biological design is integral to the concept of disorder. One must not mistake the evolutionary framework for any particular theory of biological design. Incorrect theories of human nature—often promoted socially to provide an objective veneer to social values—yield incorrect theories of normality and disorder. When an unsatisfactory theory of human nature is

used to support an oppressive diagnostic regime, the way to attack it is not, absurdly, to deny that human nature has anything to do with what is normal and disordered, but rather to provide evidence that the specific theory is a flawed account of human nature. Finally, this concern violates parity of reasoning because Forest fails to raise the same objection to the often implausible "just-so" stories of the neurodiversity movement itself that suggest, for example, that because there is some occasional highly specific feature of autism that is potentially useful under some specific modern circumstances, autism was adaptive and naturally selected as a normal variation earlier in our species history.

Returning to an earlier question, why was classic autism perceived as a disorder? It cannot simply be, as Forest's causal-role view would have it, that classic autism has negative statistically uncommon symptoms caused by an internal state. That criterion is hopelessly invalid and would mistakenly imply that illiteracy, criminal behavior, marital infidelity, and many other clearly nondisordered normal-range problematic conditions are disorders as well. The HDA explains the judgment that classic autism is judged a disorder by a combination of two judgments about autism: that the condition is harmful (to this extent, it is like illiteracy, criminal behavior, etc.) and that it is likely caused by a dysfunction, that is, by a failure of some internal mechanism to perform its biologically designed function (in this respect, it is unlike illiteracy, etc.). Though we cannot directly observe dysfunctions of internal mechanisms and have virtually no valid biomarkers for mental disorders and so the judgment that there is a dysfunction remains inferential and fallible, in the case of classic autism, this inference appears justified in virtue of the gross failure of presumptively biologically designed human capacities to socially interact, detect others' mental states, flexibly regulate one's actions, and communicate effectively. It is the strong circumstantial evidence of a dysfunction in classic autism that caused it to be classified as a disorder.

Forest's second objection is that the HDA does not help to guide us in resolving the many nosological questions that arise about autism given the series of expansions of the autism category noted above: "My second question is as follows: is the definition of disorders as harmful dysfunctions helping us to settle the debate about deficiency versus diversity, by providing us a standard for the application of the notion of disorder? To this second question, my answer will also be negative."

I believe that the HDA account is essential to achieving the explanatory discriminations that Forest requests. Of course, no analysis of the concept of medical disorder alone will tell us how to explain the phenomenon of autism or whether autism is a normal variant or a disorder, for these are factual matters that must be empirically investigated. Given the complexities of research on the brain and mind, no quick answers are to be expected. However, in terms of guidance with regard to the schematic form that such an explanation should take, the HDA implies that Forest's two questions are intimately related; whether autism is a disorder or a normal variant will depend in

part on how it is explained. Part of the explanation, as Forest emphasizes, must be a mechanical causal explanation of how various autistic conditions come about. However, as noted, all normal and disordered conditions have mechanical explanations, so the causal-role analysis is at best necessary but not sufficient. The HDA implies that whether a given form of autism is a disorder versus a normal variation will depend on how the causal explanation relates to our species' history of biological design and specifically whether the explanation involves a dysfunction. The HDA further implies that if the explanations of various conditions comprising the autistic spectrum differ with regard to the involvement of dysfunction, then the judgments of whether the conditions are disorders or normal variants will also differ.

Thus, the HDA suggests that the answers to Forest's questions may differ depending upon where along the spectrum one focuses. This sort of differentiated view may eventually undo the confusion that has resulted from the premature expansion of the classic notion of autism into the autistic spectrum and beyond. This expansion has followed legitimate scientific pathways of generalization, locating classic autism and other prima facie clearly disordered autistic conditions within a broader context that can yield fresh scientific insight. However, these expansions, while scientifically and clinically useful, have been accomplished without adequate attention to the conceptual underpinnings of the concept of medical disorder, leading to an inevitable confusion of broader autism-related normal variation and autistic disorder. The HDA can serve as a corrective to the uncritical expansion of the autism category that ignores the requirements for disorder. This is a service that the causal-role model cannot provide because both normal variants and disorders can be mechanistically explained, statistically infrequent, and problematic. Moreover, in offering this guidance, the HDA provides a provisional explanation of how it is that people on various sides of the neurodiversity dispute can hold opposed views about the diagnostic status of high-functioning autism while knowing the same basic facts about the condition: the dispute is over whether the described conditions allow one to plausibly infer that they are caused by dysfunctions and over whether they are harmful or not.

In a comment on one potentially problematic addition to the autism category, Forest says, "Instead of a sharp contrast between autistic and nonautistic people, research has pointed out the presence of autistic features within some parts of the nonautistic population—which has led to the introduction of the notion of a broader autistic phenotype (Piven et al. 1997)." So, there is a puzzle and some disagreement in the literature about how to think about BAP. It is difficult to see how Forest can even begin to address the BAP puzzle with his account of dysfunction. The causal-role model of dysfunction can be applied equally to classic autism, high-functioning autism, and BAP. In contrast, this extension of autism-related conditions poses no real problem for HDA classification. Prima facie, the BAP concerns normal variation in personality. Even Piven and colleagues, who were pioneers in this area and developed the scale

most frequently used to measure BAP, protest that BAP is not a disorder and that the scale they devised to measure it should not be used to identify pathology of any kind, let alone autism (Piven and Sasson 2014). Moreover, based on emerging evidence, BAP is most plausibly considered at this time not to be based on dysfunctions. There is evidence that single milder autistic traits are advantageous and have been positively selected, and it is only certain confluences of them that for unknown reasons become deleterious (Polimanti and Gelernter 2017). On this theory, BAP, although statistically infrequent and perhaps sometimes causing mild harm in our social environment—and thus potentially misclassified as a disorder according to the causal-role account—is not in fact a disorder but a normal variation.

More generally, the fact that features of a disorder show up in more moderate levels in population personality dimensions without constituting disorders is absolutely routine for virtually all symptoms, even psychotic ideation. Even if such extensions introduce a degree of fuzziness to the category, this need have no implication for the disorder status of many clear cases. The existence of some boundary fuzziness does not imply the illegitimacy of an overall distinction that has clear cases on both sides of the boundary. Orange versus red, night versus day, and adult versus child are legitimate distinctions with clear cases on both sides, even though there are not the "sharp contrasts" at the boundaries between these categories that Forest might desire. The developments cited by Forest do raise important conceptual questions about boundary setting, but prima facie, they do not cloud the picture regarding many clear cases, including classic autism.

According to the HDA, the question of whether various forms of autism are disorders remains an empirical question, and the question can be raised and answered differently about various subgroups currently engulfed by the autism label. One obvious reason why the diagnostic status of various forms of autism is problematic is that we just don't know much about the causes of each of the varieties of autism. With further research and theory, and with interpretation of the results guided by the HDA, many problems of diagnosis of the sort raised by Forest can be put to rest. That is, according to the HDA, there are possible empirical routes to resolving issues of dysfunction versus function and thus of disorder versus nondisorder. (Harm poses different challenges.) For example, recently, evidence has emerged in animal models of autism for a specific kind of dysfunction in which genetic functioning in certain areas of the genome becomes constricted due to the chromatin being packed too tightly and the genes prevented from expression by closing them off from the cell's transcriptional machinery (Qin et al. 2018). In the animal model, when those constricting structures were loosened and the genes allowed to function, social capacities of the sort impaired in autism were restored (Qin et al. 2018). Now, suppose that it should turn out that this form of harmful dysfunction is at the heart of classic autism (or some forms of classic autism) and that we then discovered that high-functioning autism is due to milder levels of

this same dysfunction. This would then support the conclusion that high-functioning autism, if judged harmful, is a mild form of autistic disorder. Alternatively, we could discover that high-functioning autism is not caused by this same process at all and is in fact a separate naturally selected variant due to advantages it confers and thus not a disorder. Or, we could find that high-functioning autism is due to a different type of dysfunction and is some disorder other than autism. Guided by the discriminating ability of the HDA's evolutionary dysfunction component, these issues can be addressed through evidence and theory.

Essentialist Confusions in the Extension of Diagnostic Categories

Aside from sheer ignorance of the empirical facts, there is a deeper source of confusion about the diagnostic status of autism that also applies to many other disorders and generally afflicts nosology these days. The puzzlement to which Forest refers is due to a subtle conflation of concepts that often occurs in dimensionalizing what start out as presumed disorder categories. Although for convenience we might choose syndromal language at times that refers to any conditions with certain symptoms, in the long run, we gravitate toward essentialist meanings and concepts in which categories of disorder are understood as determined by etiologies that amount to a specific type of dysfunction. Most disorder categories start out as essentialist concepts defined using some (presumptively, but defeasibly) clearly disordered base set that is syndromally manifested in harmful symptoms that are provisionally assumed to be due to the same etiology based on the same dysfunction (or possibly multiple dysfunctions that eventually can be distinguished and separated into multiple distinct disorders). The idea is, roughly, that the disorder is the etiologically homogeneous category consisting of the base set *and any other harmful condition that has the same dysfunction as its etiology*. Thus, as science advances, disorders tend to be extended in accordance with shared dysfunction etiology, not shared symptoms. This is why symptomatically, quite diverse conditions presumed to indicate the same dysfunction can fall under the same disorder and symptomatically similar conditions presumed to involve different dysfunctions can fall under different disorders.

However, the essence of the syndromal base set can be theorized in many different ways. How one extends essentialist concepts given the many properties possessed by a base set depends on two things. First, there are empirical or theoretical discoveries about the underlying nature of the base set. Second, there is the choice of a semantic or ontological marker to indicate the kind of category one intends to formulate and thus which kinds of features of the base set are relevant to guiding the extension of the category to new instances. Each base set can be generalized in many different ways based on different properties, so the ontological marker that one is defining as a presumptive disorder is crucial to the process of defining a nosological category.

For example, for water, the base set is the clear liquid in the familiar lakes and rivers, so "water" means "whatever is essentially like the clear liquid in the familiar lakes and rivers." However, without an ontological marker, this formula remains ambiguous in the extreme. That base set has many properties that might serve as an essence for a larger category depending on the ontological marker. Water is H2O if the ontological marker is "same substance," but water is H2O-that-also-has-the-essence-of-liquidity if the marker is "same liquid," and if the ontological marker is "matter (versus energy)," then the essence is different from either of these and is something like "composed of elementary particles." Which of these senses one is using matters pragmatically. A glass of ice is not a glass of water in the intended sense when asking for a glass of water in a restaurant, but it may be a glass of water in the relevant sense when asked for in a chemistry lab.

This point about specifying the ontological marker is often taken for granted, but it is where confusion about disorder categories can occur. In the case of autism, the initial base set consists of severe classic autism, and in principle, the category is extended from there in accordance with a postulated common etiology (or etiologies) of the base set. However, if the category is being essentialistically defined as a category of disorder, then, given that "disorder" means "harmful dysfunction," the essence must consist of a dysfunction, and the included conditions must be harmful. The ontological marker of "same disorder as" determines these constraints on how the category is extended. Anything falling outside of these regulative principles is not autism in the intended sense of a pathology in the same category as the identified base set of presumed pathological conditions. If this is the intended meaning of "autism," and if BAP and certain forms of high-functioning individuals with some autistic traits are in fact not suffering from disorders, then, with apologies to those whose identities may be tied up with this term, they are not autistic (in this historically anchored semantic sense). It was just a mistake by overreaching nosologists to ever apply this term to them based on the mistaken view that they are mild cases of the harmful dysfunction underlying classic autism. Such mistakes are not uncommon; for example, whales were thought of as fish for millennia until a deeper understanding of biology revealed that the kind of thing we refer to as fish do not share a deep biological nature with whales, which thus turn out not to be classifiable as fish after all. Some similar analysis may well apply to what we call "high-functioning autism"; it may not be autism at all.

However, very often, the greatest insight into a category of disorder can come from casting a broader research web and seeing the disorders as part of a larger category not constrained by harm or even by dysfunction and looking at all those conditions falling within the broader category, whether they are disorders or not. To take a very simple example, it is scientifically extraordinarily illuminating and explanatorily potent, and yields multiple fruitful lines of research, to consider what gives rise to the sickle cell gene and how the gene functions, despite the fact that, at least in malaria-endemic

locations, having a single sickle gene is not necessarily a medical disorder. In effect, there is a three-point dimensional variable ranging from zero to two sickle cell genes that defines a scientifically important target of study, and it is only through studying that broader dimensional domain—and thus identifying the adaptive malaria-resistance properties of having a single sickle cell gene (the precise causal-role workings of which are still being explored and yielding surprising and potentially useful scientific insights) that we have come to a fuller understanding that sickle cell disease, like many genetic disorders, is the result of the genetic lottery leading to an individual having too many genes of a kind that are adaptive and selected for in more moderate amounts but that together yield a nonselected and harmful dysfunction. In malaria-endemic areas, a single sickle cell gene need yield neither dysfunction nor harm, yet it lies within a crucial domain of study that includes the disorder of sickle cell anemia.

Disorder is often, as in this case, the accidental confluence of selected features that, when occurring together, yield dysfunction (meaning a zone of outcome never selected for). Indeed, there is recent evidence that autism is precisely this sort of disorder in which individually advantageous genes that were naturally selected for cognitive and social advantages occur in specific combinations that for as yet unknown reasons shift to being jointly deleterious and causing dysfunction (Polimanti and Gelernter 2017). If this is correct, then the argument for studying a broader domain that includes, for example, the BAP and high-functioning autism should have some weight as a potentially fruitful pathway to new insights. However, this research strategy and the use of the term "autism" for the dimensional expression of features associated with the disorder must not be confused with the expansion of the domain of disorder. If "autism" is extended from classic autism with the ontological marker not being "the same disorder" but rather something like "some of the same cognitive or personality traits as in the triadic syndrome at varying levels of severity" (which encompasses both high-functioning autism and the broader autism phenotype), there is then no implication that something that falls under "autism" in this sense must be a disorder, any more than something that falls under "water" with the ontological marker of "same substance" must be a liquid. Scientists may want to study dimensions defined the former way to gain general understanding about variables linked to autism-the-disorder, but that does not mean they are studying autism-the-disorder.

The issue with autism, as with many categories of pathology, is that scientists like dimensions. They like the dimensions to be as encompassing as possible so that they can formulate the most perspicuous theories and do the most decisive statistical tests of data. Consequently, our initial clearly pathological categories are regularly generalized into dimensions. This has now confusedly gotten inflated into a supposed dimensional approach to pathology. Such dimensional generalizations often ignore the fact that if the marker for "same disorder" is cast aside and other features without that constraint are seen as essences and the category extended in accordance with those features, then

generalization may occur in ways that are unrelated to pathological status. The larger more all-encompassing categories that result may socially still cling to their "disorder" status as a result of being tied to the original base set by which the relevant disorder category term was defined but may no longer represent the intended ontological constraint of pathology. It would be like extending "water" based on the H2O theory to all oxides and calling that much larger set of chemical substances "water."

Varieties of Health

With all the varied conditions now collected under "autism," Forest expresses the quite legitimate concern that "we have to be especially careful not to define these standards in an excessively narrow way and to remain sensitive to what we could call the *varieties* of mental health." Contrary to Forest's implication, this goal is entirely consistent with and best served by the HDA's evolutionary approach to human normality. Indeed, the misidentification of varieties of normal-range mental health as purported disorders is one of the primary ways I have deployed the HDA in my extensive work on false positives in psychiatric diagnosis. The HDA gives one a place to stand in formulating such critiques. Variation within normality is a routine and essential feature of evolutionary thinking across almost all human features. A focus on how human beings are biologically designed—as opposed to a focus on what is locally useful or culturally valued—leads to a critical examination of proposed expansions of diagnostic categories, and such skepticism can liberate us from diagnostic oppression. Without the in-principle objective touchstone of how human beings are in fact biologically shaped to be, diagnosis can and often does run amok in the direction of greatly overdiagnosing disorder as a tool of socialization and social control (Wakefield, Lorenzo-Luaces, and Lee 2017).

It is difficult to see how, without the HDA's biological design constraint, Forest proposes to recognize unusual varieties of health, including statistically deviant and socially disvalued varieties. Forest's causal-role approach to function and dysfunction offers no coherent account of function and dysfunction that might help here. Every condition, from "sluggish schizophrenia" (applied to political dissidents in the Soviet Union) to female clitoral orgasm in Victorian England (seen as a disorder by many physicians), could be analyzed as potential dysfunctions on a statistical-infrequency account, and they certainly have mechanical explanations that in the causal-role account can be translated into "dysfunctions." (Whether there is truly harm in these conditions despite the social disapproval is of course another matter; see the discussion of harm below and in my reply to Cooper in this volume.) So, where is the standard that tells us that we are dealing here not with dysfunctions but with "varieties of mental health"? For that, you need the HDA's evolutionary conceptualization of the enormous variation within normal-range human biological design.

Is Autism a Trade-off for Savant Talents and Thus a Normal Variation?

A further argument Forest puts forward against the HDA's explanatory powers concerns both autistic special talents and the "weak coherence" theory of autism. Here, he considers that the fact that a significant percentage of autistic individuals have been found to have unusual specific abilities within such spheres as memory, calculation, drawing, or music, or in attention to detail, has been used as an argument that autism is not a disorder. However, sometimes, features that are desirable come into existence as side effects of disorders that are in themselves harmful. As Forest ultimately concedes, "Autistic talent has been 'unmasked' by scientific research and by the exposure of exceptional cases. But this does not prove wrong the view that in many cases, autism may be harmful." The question is whether the HDA can illuminate how to think about this issue.

Regarding the isolated special abilities sometimes found in autistic individuals, Forest says, "These results also mean that instead of thinking in terms of cognitive deficits in autism, we may have to think in terms of cognitive style (Happé 1999) and a different trade-off between abilities that are impaired and other abilities that may be enhanced, at least in high-functioning autism." However, many disorders have some positive effects, but mostly these are either accidental side effects (e.g., cowpox inoculates against smallpox) or compensatory changes in response to the disorder (e.g., enhanced echolocation in blindness). The fact that a negative condition causes this sort of positive side effect or compensatory adjustment does not undo the disorder attribution to the negative condition. Forest's argument that autism may not be a disorder because its negative features are a trade-off for special talents depends on a stronger sense of "trade-off." His use of that term and his provisional limitation of his point to high-functioning autism indicate that he is considering the following hypothesis: perhaps the autistic individual's distinctive cognitive functioning is a normal variant that occurs as a result either of the random normal distribution of cognitive strengths and weaknesses in the population (as, for example, Einstein's extraordinary development of spatial ability may have been a normal-variational brain-developmental trade-off for early language learning) or as a biologically designed trade-off in which the autistic individual's weaknesses are the inevitable side effect of naturally selected strengths (as, for example, the negative features of pregnancy such as diminished physical agility in late stages of pregnancy and pain during childbirth are trade-offs for the naturally selected process of pregnancy). Such a trade-off account appears prima facie implausible for classic autism given that the severe global challenges would not seem to be even remotely offset by or required by the potential for isolated talents. The trade-off hypothesis makes most sense when applied to high-functioning autism or BAP.

However, there is simply no evidence that the savant skills of autistic individuals are naturally selected with social and emotional impairments as necessary trade-offs.

Several considerations weigh against such a hypothesis. First, savant skills do not appear regularly and they vary enormously. Estimates of the percentage of autistic individuals having any such skill range from classic studies suggesting 10% to around 25% up to as high as 63% to 88% (Happé 2018; Meilleur, Jelenic, and Mottron 2015) depending on the range of skills and perceptual acuities measured and the methodology used. However, specific skills do not exist in a substantial number of cases and in any event are highly varied, not always functionally meaningful, and sometimes transient. Moreover, savant skills are often extremely narrow (e.g., calendar calculations, jigsaw puzzle solving), placing in doubt their adaptive significance as the basis for a naturally selected trade-off. In no study does autism emerge as regularly accompanied by some uniform set of talents for which it might be hypothesized to be a side effect or trade-off, making it problematic as to how the negative autistic features could be explained as trade-offs for these skills. Finally, we know that a trade-off theory is not necessary to explain the occurrence of autism's advantageous isolated talents because certain brain disorders, including brain trauma and frontotemporal dementia, can occasionally yield the same sorts of talents, and in these cases, there is no question that the skills are the side effect of pathology, perhaps where some brain areas become disinhibited due to damage elsewhere.

Forest also suggests a second form of trade-off argument in which it is not savant talents in general but the characteristically decontextualized, detail-oriented cognition characteristic of many autistic individuals, sometimes called "weak coherence" (Frith 1989; Frith and Happé 1994), that may be a benefit of such magnitude that impairments in social skills may be a biologically designed or biologically normal-range trade-off for it. The claim is that, under some circumstances, a decontextualized focus on detail can yield divergent insights or perceptions that elude those with greater context sensitivity.

However, sensitivity to context seems to be a sophisticated normal-range biologically designed developmental achievement. Along with autistic individuals, children in general are not yet fully sensitive to context. For example, with their immature perceptual systems, children, like autistic individuals, are more resistant than neurotypical adults to optical illusions that result from context sensitivity (Doherty, Campbell, Tsuji, and Phillips 2010). Forest emphasizes that some companies are hiring autistic individuals with potential strengths in mind, such as being outspoken or attending to details that others might miss. This may be a step forward in terms of social justice, but lack of contextual sensitivity could render potential strengths ineffective because one requires a contextual understanding to know when speaking out or bringing disparate details to the attention of others is useful rather than distracting. Further serious dangers of lack of contextual sensitivity are revealed in numerous reports of autistic people experiencing such tragedies as drowning (e.g., McLaughlin and Sutton 2018; Sanchez 2018) or needless violent interactions with the police leading to imprisonment (Furfaro

2018) because of lack of sensitivity to relevant contextual cues. Overall, diminution in context sensitivity, whatever its marginal advantages, must be considered a serious intrinsic harm and not a plausible basis for a trade-off against other negative features of autism.

Neurodiversity versus Neurotypicality

Forest quotes a much-cited neurodiversity-inspired definition of the neurotypical syndrome (NT) that characterizes nonautistic individuals in a way that mimics the supposedly invalid description of autistic individuals as pathological. For example, it defines "neurotypical syndrome [as] a neurobiological disorder characterized by preoccupation with social concerns, delusions of superiority, and obsession with conformity. … Tragically, as many as 9,625 out of every 10,000 individuals may be neurotypical." Forest implies that perhaps the HDA is stymied by how to interpret such descriptions.

Humor aside, what is actually wrong with the neurotypicality-mocking passage quoted by Forest? First, it suggests that neurotypicality is privileged for the vacuous reason that it just happens to be statistically typical. However, neurotypicality is privileged not because it is statistically normal but because it is inferred to be functionally normal. Many statistically infrequent problematic conditions, ranging from illiteracy to criminality, are not considered disorders, and many statistically normal conditions, such as dental caries and periodontal disease (which characterize roughly 80% of people's gums worldwide), are considered disorders. Statistical normality is at best a fallible indicator of functional normality when it comes to disorder judgments. Second, the passage suggests that not all aspects of neurotypicality are beneficial or desirable. However, health is not simply a matter of whether a condition is positive or negative. Most people would prefer not to be anxious at an upcoming test and not to experience the pain of grief, yet these negatively valued features are not considered disorders because they are functionally normal in terms of human biological design. All of his makes sense within the HDA framework. (For further discussion of autism and the HDA, see Wakefield et al. 2020.)

Does the Harm Criterion Need to Be More Factual?

Finally, Forest accepts the need for a harm criterion for disorder but, like many critics (see Cooper, this volume, for a similar objection and see my reply to Cooper for a fuller response), objects that the HDA's reliance on social values to determine harm is problematic and would prefer that values be more factual: "what counts as 'harm' is less dependent on 'social values' than what is required by HD analysis. … What is needed is a view of harm that is more factual and not dependent on 'social values.'" Forest thinks that socially defining harm unduly restricts disorder judgments that could be based on

harm "in itself": "The deprivation of something that is both widespread and useful is in itself harmful … because it is a source of disadvantage. For a given subject, dysfunctions result in harm when they reduce significantly and repeatedly his autonomy, the range of his opportunities, or the probability of success of his actions."

Forest is basically right in suggesting that the "social values" addendum to the "harm" requirement cannot be a general and absolute requirement. If harm can be understood independently of social values, it still satisfies the HDA's harm requirement. The social values addendum cannot be general and absolute because the HDA applies to all organisms that can become disordered, including to creatures that are nonsocial and, even if social, do not possess social values as a filter through which harm is understood. Thus, the social-values codicil to the harm requirement of the HDA ought to be read analogously to "dysfunction," where evolutionary theory offers the best empirical framework for understanding the concept.

Though desirable, and in principle consistent with the HDA, a theory of harm that would apply universally to the human case is elusive and philosophically highly controversial. Moreover, in the human case, the appeal to a culture-transcendent standard for harm that resembles current Western philosophical views but is to be applied universally in medical diagnosis raises worrisome issues of implicit Western triumphalism and of turning medicine into another battlefield in culture wars in which some people's needs are ignored because their condition is not deemed to be truly harmful (for an example of this danger, see Powell and Scarffe 2019; Wakefield and Conrad 2019). I tend to think that the social values addendum suitably broadly interpreted remains relatively benign and useful, that what are claimed by critics to be culture-transcendent "factual" human values are implicit in every human cultural value system, and that many of the harms human beings suffer are *pro tanto* harms related to their social roles and expectations. (Further reasons why I added the social values codicil in the human case on which my analysis has focused are detailed in my reply to Cooper in this volume, and I will not repeat them here.) So, I will briefly try to explain or defend here the social-values addendum from Forest's arguments for a more factual criterion.

It is true that some values are deeper, more presupposed by other values, and more widely embraced across cultures, and in this sense, one might say they are more "factual" than others. However, their factuality is not something independent of what actual human beings in actual cultures value but simply an expression of a more general valuing of them across human cultures. However, it is almost always true that the realization of such values will vary across cultures based on local more specific values and practices that will influence the evaluation of harm in medical diagnosis. A simple extrapolation of what we specifically value to the criterion of harm for all human cultures smacks of Western triumphalism.

For example, in terms of the above passage, surely usefulness is valued universally, but whether a specific feature is useful is culturally relative. Similarly, values such as

advantage, *autonomy*, *opportunity*, and *success* are virtually universal cultural values, but the nature of their fulfillment is heavily culturally loaded. There is also some variation in social attitudes toward even such seemingly "factual" values. At moderate levels, autonomy may be a common value, but some societies, it is well known, disvalue high autonomy and instead value group devotion and cohesion, whereas those in the West tend to place autonomy and development of the unique aspects of the self at the highest level of desirability. While success in some sense is a universal human goal valued in all cultures, what actually constitutes success varies enormously among cultures and times. Similarly, opportunity in the form of access to social roles as a form of justice and self-realization exists in advanced liberal industrialized states with highly differentiated work roles that include some scarce and coveted positions but is not readily applicable to the way most human nomadic hunter-gatherer groups lived throughout history, so its deprivation would not constitute a disorder in that context. In my view, Forest's substantive value considerations fall comfortably within the HDA's social values–based harm component.

Forest further asserts, "Impairing language acquisition in a neurodevelopmental disorder is causing significant harm because in this case, the range of opportunities is severely reduced and results in a disadvantage for the child. And to know this, we don't need an evolutionary scenario about the benefits of language mechanisms. And we don't need to think of social values, because we cannot figure out a society where failing to learn how to speak would not be intrinsically detrimental to a human child."

According to the HDA, of course Forest is correct that we don't need evolution to know whether impaired language is harmful, because only dysfunction, and not harm, is evaluated relative to a baseline of biological design. More important, the claim that we don't need to refer to social values in judging the cited harms is immediately contradicted by the evidence Forest offers on its behalf, namely, that the reason it is clearly harmful is that it is harmful according to the values of every human culture because in every such culture, lack of linguistic ability is manifestly detrimental.

In a final argument, Forest offers a thought experiment in which a culture reveres classical autism as a sign of being chosen by the gods for a higher spiritual purpose, with the symptoms seen as trade-offs for supernatural powers. (This hypothetical is reminiscent of reports of schizophrenic individuals with their delusional symptoms being deified in certain cultures.) Forest argues that, according to the culture's social values, there is no harm to these children, but yet objectively, "we still have the underlying cognitive dysfunctions, and the reduced abilities. … Classical autism is still a disorder, even if the child is placed in a most favorable environment." Forest concludes that "it does not seem possible, then, to claim that, underlying dysfunctions being kept constant, the presence or the absence of a mental disorder depends on social values."

This example does not seem to me to support Forest's point. Forest agrees with the HDA that there is a dysfunction, whatever the community thinks: "we still have the

underlying cognitive dysfunctions. … A reappraisal has not in itself the power to transform a disorder into a nondisorder, as if functions and dysfunctions were dependent on our values. In this I would side with Wakefield." (I would caution that in this last passage, Forest appears to be running together dysfunction and disorder; a reappraisal cannot change a dysfunction into a nondysfunction, but if the altered appraisal implies a lack of harm, then it can change a disorder into a nondisorder.)

Is there harm to these children according to the community's social values? Forest says that the community theorizes that the children's deficiencies are in fact divinely caused trade-offs for their unusual powers. That very theory implies that the children's deficiencies are generally seen by the community as *pro tanto* harms and are not seen as harmful in this instance only because of the theory that there is a trade-off. However, that theory is incorrect. Forest observes, "We still have the underlying cognitive dysfunctions, and the reduced abilities. … Whatever values we adopt or reject, the harmful consequences, in the sense above defined, are there to stay." True, and the community would presumably agree, if it did not believe its false theory. The community positively values cognitive and other abilities, and their loss is a real harm by the community's own lights. However, due to an incorrect theory, they mistakenly think that in this particular case, the limitations result from a purposeful trade-off for a greater good, and so there is no harm of the sort relevant to disorder attribution. Consequently, in reality, given that the trade-off theory is false, the classic autism symptoms are harmful to these children *as judged by the community's own value system* even if that is not the belief of the community members. Once the situation is made explicit, I do not see any divergence here between harm and social values of the sort Forest suggests. (For the reader interested in further discussion of the sorts of objections raised by Forest to the harm criterion, see my reply to Cooper in this volume.)

References

Doherty, M. J., N. M. Campbell, H. Tsuji, and W. A. Phillips. 2010. The Ebbinghaus illusion deceives adults but not young children. *Developmental Science* 13(5): 714–721.

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.

Frith, U., and F. Happé. 1994. Autism: Beyond 'theory of mind'. *Cognition* 50: 115–132.

Happé, F. 1999. Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Science* 3(6): 216–222.

Happé, F. 2018. Why are savant skills and special talents associated with autism? *World Psychiatry* 17(3): 280–281.

McLaughlin, E. C., and J. Sutton. 2018. Autistic man who went overboard on Carnival cruise was traveling with special needs group. *CNN,* December 20. https://www.cnn.com/2018/12/20/us/autistic-man-overboard-carnival-cruise/index.html.

Meilleur, A.-A. S., P. Jelenic, and L. Mottron. 2015. Prevalence of clinically and empirically defined talents and strengths in autism. *Journal of Autism and Developmental Disorders* 45: 1354–1367.

Piven, J., P. Palmer, D. Jacobi, D. Childress, and S. Arndt. 1997. Broader autistic phenotype: Evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry* 154(2): 185–190.

Piven, J., and N. J. Sasson. 2014. On the misapplication of the broad autism phenotype questionnaire in a study of autism. *Journal of Autism and Developmental Disorders* 44: 2077–2078.

Polimanti, R., and J. Gelernter. 2017. Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genetics* 13(2): e1006618.

Powell, R., and E. Scarffe. 2019. Rethinking "disease": A fresh diagnosis and a new philosophical treatment. *Journal of Medical Ethics* 45(9): 579–588.

Qin, L., K. Ma, Z.-J. Wang, Z. Hu, E. Matas, J. Wei, and Z. Yan. 2018. Social deficits in *Shank3*-deficient mouse models of autism are rescued by histone deacetylase (HDAC) inhibition. *Nature Neuroscience* 21: 564–575.

Sanchez, R. 2018. An autopsy report says a missing North Carolina boy with autism likely drowned. *CNN,* November 29. https://www.cnn.com/2018/11/29/us/maddox-ritch-autistic-boy-death/index.html.

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.

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.

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: 233–246.

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.

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 J. A. Conrad. 2019. Does the harm component of the harmful dysfunction analysis need rethinking? Reply to Powell and Scarffe. *Journal of Medical Ethics* 45(9): 594–596.

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.

Wakefield, J. C., L. Lorenzo-Luaces, and J. J. Lee. 2017. Taking people as they are: Evolutionary psychopathology, uncomplicated depression, and the distinction between normal and disordered sadness. In *The Evolution of Psychopathology,* T. K. Shackleford and V. Zeigler-Hill (eds.), 37–72. Springer.

Wakefield, J. C., D. T. Wasserman, and J. A. Conrad. 2020. Neurodiversity, autism, and psychiatric disability: The harmful dysfunction perspective. In *Oxford Handbook of Philosophy and Disability,* D. T. Wasserman and A. Cureton (eds.), 501–521. Oxford University Press.

Wing, L. 1997. The autistic spectrum. *Lancet* 350: 1761–1766.

Defining Mental Disorder

Jerome Wakefield and His Critics

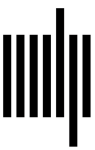
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl

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