Defining Mental Disorder

Jerome Wakefield and His Critics

The title-level DOI for this work is:

doi:10.7551/mitpress/9949.001.0001

4 Do the Empirical Facts Support the Harmful Dysfunction Analysis? Reply to Luc Faucher

Jerome Wakefield

I have long appreciated Luc Faucher's provocative interweaving of conceptual and empirical arguments (e.g., Faucher and Blanchette 2011), and I am delighted that he brings the discussion in this volume around to the implications of empirical studies for the conceptual analysis of "disorder." In particular, his contribution offers an opportunity to consider the empirical testing of 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, 2006, 2007, 2009, 2011, 2014, 2016a, 2016b; Wakefield and First 2003, 2012). I am also extremely grateful to him for his role as an editor of this volume.

Faucher urges philosophers of psychiatry to "go factual" with empirical studies. Not only do I agree, but as Faucher notices, I early did so myself in a series of studies of clinical judgment about adolescent conduct disorder. In those days, I was listed as doing X-Phi on the empirical philosophy website. However, "the facts" always need interpretation, and any hope that going factual will in some simple way resolve conceptual disputes is overly optimistic. As a case in point, despite my basic agreement with Faucher as to the value of empirical work on the concept of mental disorder, I disagree with the anti-HDA interpretations Faucher offers of some extant studies of models of mental disorder, and I also disagree with his speculations about what empirical studies with varying outcomes are likely to mean. I thus respond to Faucher's points as a cautionary tale about the potential pitfalls that face such empirical work. I will also revisit my own HDA-related empirical studies of clinical judgments about conduct disorder (CD), considering not only Faucher's critique of them in his paper but also previous critiques by Dominic Murphy and Robert Woolfolk as well as by Arthur Houts. Finally, I comment briefly at the end on Faucher's additional comments regarding supposed epistemological challenges facing the HDA's application.

Do the Colombo et al. and Harland et al. Studies Challenge the HDA?

I start by considering Faucher's claims that two clinical-judgment type empirical studies pose significant problems for the HDA. Specifically, Faucher claims that because different groups have different judgments, that must mean that there are multiple concepts of disorder. I am going to closely examine the studies that Faucher claims are problematic for the HDA and explain how he misconstrues their results and how the different judgments by different groups of subjects have no implications at all for the concept of disorder.

Faucher describes the first study and how it supposedly diverges from the HDA as follows:

> Colombo and colleagues (2003a)…give vignettes to subjects from distinct groups (psychiatrists, psychiatric nurses, patients, informal carers, social workers, etc.) [that]…describe someone who shows symptoms of schizophrenia, with their background (life situation, childhood, etc.).…Researchers then code responses according to six models they constructed and that are supposed to reflect current conceptualizations of mental disorders: for instance, the biomedical model, the social model (the disorder is within society), the family model (the whole family is sick, not just the patient), and so on.…Their results show a quite different picture from Wakefield's; for instance, while 91.3% of psychiatrists agreed with the medical model, only 8.8% of social worker agreed with it, instead showing preference for the social model (47.5%).

Faucher concludes, "What this set of studies reveals is that different people (and different professional groups) seem to use different models of mental illness." He presents this as a challenge to the HDA, suggesting that it shows that there are variant concepts of mental disorder.

In fact, the cited statistics are irrelevant to the HDA and to the concept of disorder. In the Colombo et al. (2003) study, the respondents started by reading a vignette about a 30-year-old male whose behavior, including social withdrawal and strange ideas such as that a religious group is putting thoughts into his mind, suggests the onset of schizophrenia, but the vignette also mentions various social stressors both early (a death in the family) and recent (business failure) in the man's life and notes that there was no previous history of psychiatric problems. After reading the case vignette, participants, twenty each of the psychiatrists and social workers on whom I will focus, responded to twelve open-ended interview questions about the nature of the individual's problem and its treatment. The questions fell into categories including, for example, diagnosis/definition, interpretation of behavior, labels, etiology, treatment, and prognosis. For example, the question concerning etiology was "what do you think caused Tom to behave like this?" (2003, 1558). The transcribed open-ended answers to the questions were then qualitatively scored by raters for agreement with one or more of six models of disorder: medical-organic, social-stresses, cognitive-behavioral, psycho-therapeutic, family interactions, and conspiratorial-myth. Guidelines for when answers fit each

model were provided by the researchers (see below). Although generally a response fit one model, raters could score agreement with from zero to three models.

Now, regarding the Colombo et al. report's summary differences in psychiatry's and social work's endorsed models noted by Faucher, those percentages lump together many different indicators, most of which are irrelevant to judgments of disorder. Fortunately, detailed data are presented on the psychiatrists' and social workers' responses to specific study questions, allowing a look past the summary statistics. (Colombo et al. appear to use endorsement of a model by at least half of respondents as the threshold for attributing that model to the group, and I will follow this metric in my discussion.)

Surveying the responses to some specific questions most relevant to disorder attribution, we find that, despite their general sympathy with the social model, none of the social workers (0%; n = 0) agreed with the social model's definition of diagnosis that implied health versus illness (described in the guidelines as "Health/low stress—illness/high stress continuum"), and none endorsed the medical model ("physical health—illness continuum"), which 100% of the psychiatrists endorsed. Half of the social workers (50%; n = 10) agreed with the "labels" social model item that asserts explicitly that the person's condition is not a disorder ("person is seen as a victim of social forces and not as ill"), whereas none (0%) of the psychiatrists did so. Most social workers (75%; n = 15) agreed with the social model of etiology ("social and economic stress, cultural conflict, marginal status, etc."), while few psychiatrists did so (15%; n = 3), and correspondingly, the social workers agreed that, other than long-term individual psychotherapy, treatment should be "social change to reduce stress" (50%; n = 10), which does not directly address internal states at all, whereas 0% of psychiatrists endorsed that answer. All this is consistent with the sharp divergence in prognosis, for which by far the most frequent social worker's answer (80%; n = 16) was the social model's, "Good if changes made at the social level"; 0% of psychiatrist's agreed with this outlook.

It thus seems that a number of social workers rejected the label of mental disorder for the described individual in this study because they judged that the individual was having a problematic reaction to a stressful environment and that there was no internal dysfunction sustaining the symptoms independent of the social stressors. The study is aimed at exploring potential issues on professional teams due to different ways of modeling disorder, and one issue pointed to by the study is the potential for confusing caregivers by their getting caught in the crossfire between professionals who believe there is a mental disorder and those who do not, as in the following report by a caregiver: "my daughter had some problems and got schizophrenia and that's an illness so I was told … I started to do things for her because she was sick … she got annoyed and said she wasn't sick. The social worker told me to give her space and said she was just depressed because of her problems … I mean what is going on?" (1567).

However, none of this suggests that psychiatrists and social workers have different concepts of mental disorder. Rather, the pattern of answers reveals that the psychiatrists'

and social workers' different views of whether the described individual has a disorder track their respective beliefs about whether the individual has a dysfunction, as the HDA predicts.

The basic problem here is that Faucher fails to distinguish *theories* of mental disorder from the *concept* of mental disorder. To construct a credible scientific theory of mental disorder, one must understand the target phenomenon, so one must already possess the concept of mental disorder. Colombo et al. do at one point refer to what they are examining as "conceptions of mental disorder" (1565), but this is misleading; their results have nothing to do with the *concept* of mental disorder and rather address different *theories* or *models* of the nature of mental disorder. The mental health professions entertain many models or theories of mental disorder ranging from brain disease, repressed conflict, and cognitive distortion to family dynamics, behavioral reinforcement, and social stress. To some degree, members of different professions—particularly psychiatrists versus social workers versus psychologists—tend to be trained in and to embrace different models that are distinctively relevant to their profession's expertise and focus. The *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) recognizes that it is essential that its diagnostic criteria be designed in a theory-neutral way precisely because a single disorder, say depression, may be explained by rival theories in different ways. The HDA, like the *DSM*, provides a theory-neutral formulation of what constitutes a mental disorder, independent of theoretical orientations. Thus, differences over theories of mental disorder, and even differences over whether specific conditions are mental disorders, do not imply differences over the concept of mental disorder.

Turning to the second study cited by Faucher as challenging the HDA, Faucher says the following about Harland et al.'s (2009) "A Study of Psychiatrists' Concepts of Mental Illness": "Even worse (for Wakefield), according to some other studies (Harland et al. 2009), different disorders seem to activate different models. If these studies are on the right track, then there is a more complex picture (different groups have different concepts, at different times, for different types of patient, etc.) than the unified and universal picture that Wakefield proposed."

In fact, despite its promising title, the Harland et al. study is irrelevant to testing the HDA and to illuminating the concept of disorder. The title is an instance of how researchers often use "concept" not for concepts in the philosopher's sense but for models and theories of what falls under the concept.

Harland et al. (2009) is a study of "how a group of trainee psychiatrists understand familiar mental illnesses in terms of propositions drawn from different models" (967). Despite the study's promising title and the fact that Harland et al. call the models "conceptual paradigms" (968), the "models" are *not* competing analyses of the concept of "mental disorder" but in fact mostly competing theories of the etiology of mental disorder (biological, cognitive, behavioral, psychodynamic, social realist, and spiritualist) along with a couple of views that question whether standard categories really are

mental disorders (social constructivist, nihilist). The study's questionnaire asked four Likert-scale "agree-disagree" questions about each of the eight models for each of four standard psychiatric disorders (schizophrenia, major depressive disorder, generalized anxiety disorder [GAD], and antisocial personality disorder).

In regard to the six etiological theories, the question most pertinent to the concept of mental disorder is the etiology question, and all of these questions presuppose that the condition is a disorder and ask about what causes it ("The disorder results from brain dysfunction"; "Maladaptive thoughts and beliefs are normally distributed in the population and it is the extreme ends of this distribution that account for the disorder"; "The disorder results from maladapted associative learning"; "The disorder results from the failure to successfully complete developmental psychic stages"; "Social factors such as prejudice, poor housing, and unemployment are the main causes of the disorder"; "Neglecting the spiritual or moral dimension of life leads to the disorder"). The fact that that every question uses the term "disorder" to describe the condition is a fatal problem for drawing any conclusions regarding the distinction between disorder and nondisorder because there is no "nondisorder" option. The etiological questions for the skeptical social constructivist and nihilist models do seem to imply nondisorder ("The disorder is a culturally determined construction that reflects the interests and ideology of socially dominant groups"; "All classifications and 'treatments' of the disorder are myths," respectively), but there is nothing in the study that shows what other beliefs are correlated with them, so there is no information about what the distinction between disorder and nondisorder means to the participants. Moreover, few endorsed those categories, and there are no questions that explore associated beliefs about dysfunction or whether something has gone wrong with psychological functioning. The point is not whether the respondent has a certain view of a condition as being a disorder versus nondisorder but why the respondent has that view of the condition, and for that you need to test for correlated variations in other potentially related beliefs.

Note that the judgment that a condition is or is not a disorder says nothing in itself about the nature of the concept of mental disorder, which is supposed to offer an explanation for why people judge disorder versus nondisorder. In the mental health field, some think that virtually no standard disorder categories are really disorders, and others think that virtually all are, and many think some are and some aren't, and the HDA predicts the basis for such judgments but says nothing about which judgments people will actually make.

So, to return to Faucher's claim, these studies' results are decidedly *not* "even worse (for Wakefield)." Faucher claims there is a problem for the HDA because, he says, Harland et al. show that "different disorders seem to activate different models." But this has nothing at all to do with the *concept* of mental disorder; it has to do with different theoretical views of the etiology of specific mental disorders. It is entirely consistent with the HDA to hold various different theoretical positions about the nature and causes of

dysfunction. For example, it is consistent with the HDA to believe that schizophrenia and GAD are both harmful dysfunctions in which something has gone wrong with biologically designed functioning but believe that schizophrenia is caused primarily by a genetic or brain dysfunction, whereas GAD is primarily the result of cognitive issues. Faucher says, "If these studies are on the right track, then there is a more complex picture (different groups have different concepts, at different times, for different types of patient, etc.) than the unified and universal picture that Wakefield proposed." That is not what the study shows. The study shows not that the concept of mental disorder varies with group, time, or condition but, at most, that different groups may have different theories of the etiology of mental disorders. To address the HDA, at a minimum you would need to allow respondents to judge whether a described condition is or is not a mental disorder and then correlate that judgment with indicators of harm and dysfunction. The Harland et al. study lacks these minimal requirements because it was never meant to address the conceptual issue, in the philosophers' sense of the term.

In sum, in citing the Colombo et al. and Harland et al. studies as addressing the concept of disorder and having implications for the HDA, Faucher confuses theories of the nature and etiology of mental disorder (which in the literature are often referred to as "models" of mental disorder) with disputes about the concept of mental disorder itself. Obviously, there are many competing theories of mental disorder, some of which tend to break down along professional disciplinary lines but many of which vary from individual clinician to clinician. One can disagree about how the dysfunctions in various mental disorders are caused and even disagree over whether a certain condition is caused by a dysfunction, while having exactly the same concept of mental disorder.

Conceptual analysis is in my view a form of psychological theorizing about shared cognitive structures underlying shared classificatory judgments. Conceptual analysis generates hypotheses and identifies evidential support but should be continuous with empirical work to evaluate claims about both the existence and the nature of the hypothesized shared representational structure in a target linguistic community of interest. Empirical studies thus have an important role as an adjunct to conceptual analysis if designed, executed, and interpreted with care. However, isolated studies almost never prove anything taken individually; they must be part of a research program in which alternative hypotheses about the meaning of the results of a study are generated and progressively tested, as occurs in any science. Moreover, empirical study of conceptual issues is challenging, especially when it comes to designing the experimental manipulation for testing rival hypotheses to yield relatively unambiguous outcomes. This is because concepts interact in a variety of ways with the background web of beliefs to yield classificatory judgments, so judgments in response to a target vignette can represent many different things. For example, one time when I was testing vignettes on graduate students in clinical social work, I noticed that if I specified that the individual described in a vignette had been sexually abused as a child, then

that enormously increased the percentage of responses agreeing that the individual has a mental disorder, irrespective of the environmental-context versus internal-causation manipulation. The reason was not divergent concepts of disorder but rather that at that time, it was an article of deeply held theory among mental health professionals that anyone who is sexually abused almost certainly will develop a mental disorder, a belief that has been called into question to much controversy (Rind et al. 1998). There are many such alternative hypotheses available to explain most judgments, and the art of vignette and questionnaire construction is largely the art of narrowing the range of plausible interpretations.

Faucher, in being concerned about variations in individual uses of "mental disorder," seems to me to have lost track of the primary motivation for the conceptual analysis of "mental disorder." It is not a linguistic exercise or conceptual fishing expedition to discover all the various ways people use the term "disorder" and argue for one. Surely, like almost all interesting abstract and theoretical terms, "disorder" has a large range of subtle variations in usage. It is rather an attempt to address specific foundational challenges to psychiatry as a science raised by the antipsychiatry movement and some Foucaultian and postmodernist theorists. The question is whether there is a widely shared meaning of "disorder" in our professional/lay linguistic community that simultaneously (a) is consistent with psychiatry being a medical specialty, in the sense that "disorder" in "mental disorder" is used in the same sense as "disorder" in "physical disorder" and "medical disorder" and thus locates mental disorders within the "medical model" of conditions that are medical disorders, and (b) offers at least an in principle distinction between genuine mental disorders and various normal-range problems in living, emotional distress, social deviance, disapproved and undesirable behavior, and other acknowledged misapplications of "disorder" that are not true medical conditions. Many researchers, theoreticians, and clinicians clearly believe that "mental disorder" has such a meaning despite the many antipsychiatric challenges. The analysis of "mental disorder" in my view is first and foremost an attempt to resolve this issue and show that "mental disorder" does have a widely understood meaning with both of the aforementioned properties, thus securing the conceptual and scientific foundations of psychiatry as a medical discipline.

Conduct Disorder Studies

I now turn to my series of conduct disorder studies (Kirk et al. 1999; Pottick et al. 2003; Wakefield et al. 1999; Wakefield et al. 2002; Wakefield et al. 2006) that Faucher discusses, in which subjects judged mental disorder in a described youth engaging in antisocial behavior that satisfied *DSM-IV*'s criteria for conduct disorder. I will focus on the data, reported and unreported, from the 2006 study in which we compared four samples: three lay samples (nonclinical social work graduate students who had not

taken a *DSM* course and generally had no mental health experience; nonpsychiatric general or pediatric nursing graduate students, with most having nursing experience but no mental health experience; and undergraduates in sociology courses reporting no mental health experience [for details, see Wakefield et al. 2006]), and the clinician sample reported earlier in the 2002 study consisting of 117 graduate students in clinical psychology and clinical social work with an average of four years of clinical experience (the two clinical groups were pooled when initial analyses showed that their responses were extremely similar). (As an aside, these studies yielded results of interest regarding the concept of disorder that are not here pursued. Notably, for example, contrary to what some philosophers have suggested, neither clinicians nor laypeople equate treatment with disorder status [see Wakefield et al. 2002].)

The *DSM-IV* diagnostic criteria for CD are stated purely in terms of symptomatic behaviors, for example: "often bullies, threatens, or intimidates others"; "often initiates physical fights"; "has used a weapon that can cause serious physical harm to others (e.g., a bat, brick, broken bottle, knife, gun)"; "often stays out at night despite parental prohibitions, beginning before age 13 years"; "is often truant from school, beginning before age 13 years"; "has broken into someone else's house, building, or car"; "often lies to obtain goods or favors or to avoid obligations"; "has stolen items of nontrivial value without confronting a victim (e.g., shoplifting…)," and so on. It is also required that "the disturbance in behavior causes clinically significant impairment in social, academic, or occupational functioning."

In many of its diagnostic categories, the *DSM* uses contextual exclusions to discriminate genuine disorders from normal-range responses to problematic environments (Wakefield and First 2012). However, it does not do so in the CD category. Contrary to the *DSM*, the HDA predicts that whether CD symptoms are interpreted as indicators of mental disorder will depend on what the diagnostician infers about the explanation of the symptoms. Specifically, if the symptoms are seen as likely due to an internal dysfunction, they would tend to be seen as psychopathology, but if they are seen as a reasonable response to environmental circumstances, the same symptoms would tend to be understood as not pathological. Thus, to test whether the HDA or *DSM* more accurately reflects intuitions about disorder, we constructed vignettes with the same symptoms satisfying *DSM* criteria for CD but added additional contextual information designed to trigger causal attributions either to the environmental context or to an internal dysfunction (without using those or related labels so as not to bias the subjects' reactions) and evaluated whether subjects judged the symptoms to indicate a disorder.

In these studies, for each of two described youths (Carlos, Judy), there were three vignettes: "symptom only," "environmental context," and "internal dysfunction." Each subject responded to one version for each youth. The symptom-only vignettes included some demographic and history information of a kind common in case descriptions and described the youth's symptoms, which were formulated to satisfy

DSM-IV CD diagnostic criteria (i.e., three or more symptoms from the *DSM*'s list plus resulting role impairment), as follows:

> Judy is a 13 year old white junior high school student who has been in trouble with school authorities for over a year for frequent truancy, which has markedly impaired her academic performance. She has also been caught shoplifting. Recently, an incident in which she was arrested for breaking into a car brought her into court. … She often lied to escape from her responsibilities around the house, she often stayed out until late at night despite their prohibitions.
>
> Because of many disciplinary actions initiated by his teachers, [Carlos] was referred to the school social worker for an evaluation. In addition to his often being truant, teachers have reported that Carlos often bullies or threatens his classmates and often initiates physical fights, which has seriously limited his social relationships. He was recently caught using a baseball bat as a weapon in a schoolyard fight.

These symptom-only vignettes formed the first paragraph of the other two vignette conditions, which each added a paragraph of contextual information to the symptom description. The environmental-context vignette offered background information that tended to explain the symptoms as understandable reactions to environmental circumstances, specifically as Judy's reaction to attempted sexual abuse by her stepfather and Carlos's self-protective reaction to a gang violence–infested school environment, as in these excerpts:

> When Carlos first arrived at the school, he was terrified by the violence. Eventually, to avoid being preyed on, he and many of his classmates joined one of the rival gangs. Gang fights at the school often involve weapons like bats and bricks on both sides … Carlos learned over time that the most effective defense … was to be highly aggressive and intimidating to others. However, within his gang and outside in the community, he has close relationships and a keen sense of loyalty. Last summer, when Carlos returned to Mexico for his first extended visit with his grandparents, … he got into no trouble. … But, once he returned to Los Angeles, his problematic behavior began again.
>
> [Judy's] troubles began shortly after her stepfather started attempting to sexually abuse her when she was 12. After that, Judy often made up excuses to get out of the house, stayed out late, and even ran away overnight to avoid him. … She became truant partly to avoid her stepfather, who often waited to pick her up at the end of the school day. … Because she had no money for food when out of the house, Judy began shoplifting and once she broke into a car to get some change she saw on the dashboard, even though she felt bad about doing so.

The "internal-dysfunction" vignettes provided information that suggested an internal source of the symptoms that seemed beyond normal range, as these excerpts illustrate:

> Judy's schoolmates … reported that she was often unreliable and dishonest with them. … Her problematic behavior was not confined to home and school; in the residential facility, Judy did not obey the house rules and lied in order to get out of her assigned chores, and she tried to run away on her third night there.

> Carlos reacts to the slightest perception of provocation with severe anger. … He often escalates fights from fists to weapons. … Discipline seems to only exacerbate his problematic behaviour. Even with those he hangs out with, Carlos is easily irritated and frequently initiates fights. Last summer, when Carlos returned to Mexico for his first extended visit with his grandparents, he got into trouble.

As Faucher explains, each vignette was followed by a set of questions asking the subject to rate their agreement or disagreement (recorded on a Likert scale ranging from 1 = "I strongly agree" to 6 = "I strongly disagree") with several items, among which was the following: "*According to my own view, this youth has a mental/psychiatric disorder*." As Faucher also explains, "the results were quite univocal," with the symptoms judged as less indicative of the presence of mental or psychiatric disorder in the environmental-context condition than in the internal-dysfunction condition, thus strongly supporting the HDA over the *DSM* as a better predictor of judgments of mental disorder.

Having explained the nature of the CD studies at some length, I now turn to Faucher's critical comments on the CD studies. First, Faucher cautions, "If these results are suggestive that Wakefield's analysis is valid, I want to argue that they might not be decisive." Of course the CD studies are not *decisive* proof of the HDA, because decisive support is not what they—or almost any other empirical studies in psychology—are designed to provide. The studies do provide solid evidence that the HDA is superior to the *DSM* diagnostic criteria as an account of disorder intuitions. However, as for any successful study, there are likely multiple alternative hypotheses that suggest that further studies would be useful.

Faucher offers one general and three specific reasons for concern about the CD studies. The general reason is a standard concern about "armchair" conceptual analysis. Faucher quotes my (Wakefield 1999c) comment explaining why, despite appearances to the contrary, conceptual analysis is a quasi-empirical process that relies on a confidence that one shares certain judgments with others in the linguistic community and objects, "But, one might wonder what are the grounds for such a presupposition? … After the recent wave of X-Philosophy investigations, one becomes suspicious of what is presented as an armchair analysis of a concept." Perhaps, but the "armchair" concern can be overdone in the case of mental disorder. There is a world of difference between the situation in this regard in philosophy of psychiatry versus most other areas of philosophy where "armchair analyses" often involve bewildering counterfactuals that no one has ever considered outside of the philosophical context. Unlike, say, the nuances of the "justified true belief" account of knowledge or the variations in judgments of moral responsibility under various scenarios of determinism (these are examples of X-Phi that Faucher presents), the area of psychiatric and medical diagnosis has been subject to exhaustive scholarly and public analysis and debate over the distinction between disorders and nondisordered problems in living. This has been not only in response to the antipsychiatric challenge but also as part of the vigorous airing of disputes during

the revision process leading to each new edition of the *DSM*. Thus, the concept of mental disorder has a plentiful and well-elaborated professional and lay public history of discussion. Consequently, in this area, "armchair" methodology need not be a matter of idiosyncratic and untethered intuitions of the individual philosopher about esoteric counterfactual cases but rather a matter of evidential judgment informed and constrained by a broad base of public data about usage and the reasons that emerge in disagreements. This is not "armchair" philosophy in the usual sense. In some areas of philosophy, the armchair—unless one goes with X-Phi—is all there is, but in philosophy of psychiatry, one can simultaneously be a philosopher and a member of the mental health community who can inform on community linguistic practices.

Faucher asks what the basis is for the presupposition that one's judgments are not idiosyncratic and that one shares judgments with a community. Here is Robert Spitzer, the leading psychiatric nosologist of the past century, answering that question: "What is remarkable—and is in keeping with Wakefield's analysis of the problem—is the great degree of consensus that exists about whether particular psychological or physical conditions are or are not disordered in the absence of a definition of disorder in general. Neither physician, psychologist, nor the public have any problem in agreeing that childbirth (painful), being in love (overevaluation of the loved object), and normal grief (marked distress) are not disorders and that unprovoked panic attacks (dysfunction of the anxiety system), severe depression (dysfunction of mood regulation), and schizophrenia (dysfunction of reality testing and motivation) are disorders" (Spitzer 1999, 430).

Turning to Faucher's concerns specifically about the CD studies, he writes that the sample, which is deliberately heterogeneous, still might be biased in some unknown way. This is a very speculative concern that appears to be based on no immediate weakness identified in the research. It is of course legitimate, but it could be applied to all such studies, even those that, like mine, purposely used heterogenous samples.

Faucher's second concern is that the CD studies did not specifically test the evolutionary component of the HDA and that they "tested only HD [i.e., harmful internal dysfunction but without the evolutionary interpretation]. To conclude that it tested [the evolutionary component as well], one would need to assume that the only interpretation of "function" possible is some kind of designed function. Yet there are many different interpretations of function, for instance, in terms of system-function (Cummins-Function), propensity function, and so on."

These studies did not attempt to test the evolutionary component because, as I have repeatedly explained (see, for example, my response to Lemoine in this volume), the evolutionary component is a theoretical scientific discovery about what constitutes biological design and not part of the conceptual-analytic understanding of "disorder," and thus the HDA makes no simple prediction about general lay judgments on this technical scientific matter. People can understand the concept of "disorder"

as "harmful dysfunction" without understanding or agreeing with or even knowing about the evolutionary account of function and dysfunction (e.g., Hippocrates did not know about it and Christian fundamentalists reject it, yet both make mostly the same judgments of disorder as contemporary medical professionals do). So, it makes no sense to test evolutionary beliefs as part of a study of "disorder" unless it is a specially selected sample with that specific issue in mind.

Moreover, it is questionable whether there is really a need for empirical studies comparing the sorts of philosophical views of function that Faucher mentions. It is a fallacious argument to reason, as Faucher does, from the correct premise that within philosophy of biology, "there are many different interpretations of function" (e.g., propensity, systems, biological design) to the implicit conclusion that there are many different interpretations of function potentially relevant to understanding the concept of medical or psychiatric disorder. This is where philosophy has a role in helping to identify what are prima facie plausible hypotheses worthy of empirical effort. If one accepts that there are several possible meanings of "function" in biology, the question for philosophy of medicine is which of those accounts of "function" provide a corresponding account of "dysfunction" as failure of function that is prima facie plausible as the specific sense of "dysfunction" that explains medical judgments of disorder. This is where many critics of the HDA go awry. They note that evolutionary function is just one of several competing philosophical analyses of the use of "biological function" and leap from there to the objection that my choice of evolutionary function is arbitrary (e.g., see Murphy's chapter in this volume). However, my use of evolutionary function is based on an analysis of which biological meaning or meanings of "function" are prima facie plausible candidates to undergird medical notions of function and dysfunction. The step missing from such objections is the systematic testing of the proposed alternative accounts of "function" against common medical judgments to establish whether they are plausible accounts of function and dysfunction in the medical sense.

In fact, the alternative analyses of "function" mentioned by Faucher, even if they explain some judgments about "biological function," do not work as accounts of function and dysfunction in the senses relevant to judgments of medical disorder and thus can be safely ignored in the context of philosophy of medicine. For example, the "propensity theory" (which holds that the effect of a condition on current reproductive fitness is the criterion for function and dysfunction) cannot even explain why, say, dyslexia is considered a disorder but illiteracy due to lack of education is not, supposing that these conditions with similar effects have similar negative impacts on reproductive fitness in our modern social environment. Similarly, the endless dispute about attention-deficit/hyperactivity disorder's (ADHD's) diagnostic status is over whether normal-range rambunctious children are being misdiagnosed, yet even if such rambunctious children suffer a fitness-propensity insult the same size as those with true ADHD (due, perhaps, to their unfortunate interaction with our constrictive school

environments), they are still not considered disordered. Do physicians or patients really think that if broken legs or blindness or chronic pain are shown not to influence reproductive fitness in our modern environment, they are no longer disorders? Certainly, according to the propensity account, deciding not to have children would be a dysfunction. Moreover, the propensity theory opens psychiatry up to uses for social control, because social rules can be designed to influence fitness. The propensity theory thus fails the test of answering the antipsychiatric challenge.

The systems view (which holds that the effect of any part on the capacities or properties of a larger containing system is a function of the part) is even less applicable to diagnosis. Everyone, including Cummins (Cummins and Roth 2009), agrees that the systems view makes no distinction between health versus medical disorder as "functions" of whatever internal structures bring them about in the organism. If you are interested in understanding the etiology of CD, then CD is a function of its causes, and an internal state that prevents the development of conduct disorder is a "dysfunction" in that context, according to this view. In biological research, where one is trying to understand how things work, this usage of "function" does often occur. However, the systems account is so inclusive of functions that it is not clear how one would actually formulate an empirical test within the medical context. (For an exhaustive analysis of why the systems view does not work in the medical context, see my reply to Murphy in this volume.) Granted that the understanding of "internal dysfunction" in the sense relevant to diagnosis could use additional empirical exploration (indeed, see below for some unpublished results on this from my CD studies), the types of alternatives mentioned by Faucher can be dismissed out of hand.

Third, Faucher objects that the experimental situation may not represent what happens in actual clinical practice: "subjects have to react to an abstract situation described by vignettes. … One might wonder if this kind of reaction … captures concepts that are used or that are more operative in practice." It is possible that the experimental results don't represent what people would do in some clinical situations, but this is just a speculation, and contrary to Faucher's assumption, even if it were so, this does not imply that what people do in practice involves a different concept. Practice involves many compromises with considerations other than the concept of mental disorder. What we want here is not just abstract generic worries but alternative hypotheses with some basis in theory or empiricism. (See my reply to De Vreese in this volume for further discussion of the way that excessive focus on what is done in practice can confuse the conceptual investigation.)

Faucher also notes that some philosophers have argued that self-report and forced-choice instruments of the kind I used in the CD studies "are open to numerous distortions due to wording, framing of questions, question order, and so on." These are routine concerns in experimental work, and addressing them is the art and skill of vignette and questionnaire construction and experimental methodology. The point is

to test predictions that distinguish rival hypotheses, and methodology aims to eliminate as many extraneous explanations for the outcome as possible. The precise nature of wording, the balanced ordering of stimuli across subjects, the nonbiasing framing of questions, and so on take months of focused attention to get as close to right as possible, and Faucher offers no reason to think the CD studies suffered from this sort of defect in a way that could jeopardize the meaning of the results. And, of course, it is good to use multiple methodologies, as I have in pursuing epidemiological analyses of HDA-related hypotheses (e.g., Wakefield 2013; Wakefield et al. 2017; Wakefield and Schmitz 2013; 2014; Wakefield et al. 2007). As objections to the CD studies, these concerns are too abstract and generic to be credible.

Replies to Murphy and Woolfolk's and Houts's Critiques of the Conduct Disorder Studies

Faucher's concerns stay at too abstract a level to cast any serious doubt on my CD studies as support for the HDA. However, others have been more targeted and empirically grounded in their criticisms, and I take this opportunity to answer the two most salient objections.

First, Arthur Houts (2001), defending a symptom-based behaviorist account of disorder that denies any inference to internal dysfunction, argued that the CD studies' results do not support the HDA because there is no evidence that subjects in the "internal-dysfunction" vignette condition actually inferred a dysfunction or anything else about the described individuals' minds. Houts observed that, whereas environmental-context vignettes described environmental circumstances that triggered antisocial behavior, the "internal-dysfunction" vignettes did not specify or mention internal dysfunctions but only described behavior without any environmental explanation. Houts thus in effect challenged the validity of the study's fundamental experimental manipulation: how do we know that the internal-dysfunction vignettes actually triggered a dysfunction inference about the described youth and the environmental-reaction vignettes did not? Houts argues that the subjects may have instead attributed disorder simply based on lack of any information about environmental contingencies:

> These outcomes were interpreted as supporting Wakefield's claim that people infer there is a mental disorder when they infer a dysfunction, but in fact, the investigators did not report what inferences led to the differential frequency of seeing a mental disorder when antisocial behaviors were presented under different collateral information conditions. Based on the information provided in this study, a more consistent conclusion is that the social work students attributed antisocial behavior to a mental disorder when they could not otherwise explain it based on current environmental conditions. In other words, the inference to mental disorder is an inference based not on knowledge of function or dysfunction, but an inference based on ignorance. (Houts 2001, 1122–1123)

It always strengthens a study to test for the success of the experimental manipulation, and fortunately, we anticipated Houts's type of concern and added some additional questions to address what if anything the subjects inferred about the described youth. So, I can report analyses that directly address Houts's objection. Due to word limitations, we did not include these analyses in the published versions, so keep in mind that these are un-peer-reviewed results. However, the relevant analyses were done at the time of the publication of the studies using the same methodology and analytic techniques that were peer-reviewed in other analyses.

The most direct way to test Houts's hypothesis is to establish whether subjects did in fact infer internal dysfunction versus no internal dysfunction or related properties from the internal-dysfunction versus environmental-context experimental manipulations, respectively, contrary to his claim. We collected two kinds of data aimed precisely at this point. First, we presented the following item to the psychologist and nurse samples: "*This youth's problematic behaviors likely result from a dysfunction of some cognitive, affective, or other mental mechanism in the youth*." This item explicitly identifies the cause of the problem specifically as a dysfunction of a mental mechanism inside the individual.

The results decisively falsified Houts's hypothesis. For both psychology and nurse samples, in both youth conditions (i.e., Judy and Carlos vignette sets), as the HDA predicted, internal-dysfunction vignettes generally caused subjects to infer dysfunction, and environmental-reaction vignettes did not. Averaging across the four cells (two samples, two youths), the average percentages agreeing with "dysfunction" in the internal-dysfunction versus environmental-context conditions were 81.6% versus 24.0%, respectively (one-tailed Fisher's exact test, $p < .01$ in each of the four cases). These large differences disconfirm Houts's hypothesis that subjects did not infer internal dysfunction and confirm the validity of the study's context manipulation.

However, Houts might argue that the "dysfunction" item in isolation remains potentially ambiguous. Fortunately, we went further. The "dysfunction" item seemed to us a bit technical for nonclinical samples and so was presented only to the two professional samples mentioned earlier. To test for inferences to dysfunction in all of our lay and professional samples, we used a less technical item: "*It seems likely that something is wrong with this youth's mind*." Many theorists (e.g., Klein 1978; Spitzer and Endicott 1978) state that a person has a mental dysfunction when "something has gone wrong with" the person's mind, and this language seemed the closest we could get to colloquial, nontechnical usage for dysfunction in the HDA sense.

All samples answered the "something wrong with the mind" item, and professional and lay responses were similar. In all groups, the percentage agreeing with "something wrong" was substantially and significantly greater in the internal-dysfunction than environmental-reaction context (average percentage agreeing to "something wrong" across all groups and both youths was 76.2% versus 16.9%, respectively; one-tailed

Fisher's exact test, $p < .001$ in all cases). Thus, the results support the initial results for the "dysfunction" item and again strongly disconfirm Houts's hypothesis that subjects did not infer anything wrong internally with the described youths in the internal dysfunction experimental condition.

Second, in a similar vein, Murphy and Woolfolk (2000) lodged the following objection to taking the CD studies as evidence for the HDA: "A careful analysis of the experimental materials employed suggests that the study confounded severity of the anti-social behavior presented in vignettes with the ostensible and intended independent variable manipulation: whether or not the anti-social behavior was readily attributable to external circumstances" (289). That is, perhaps subjects judged disorder in the internal-dysfunction context and nondisorder in the environmental-reaction context based not on inferences to dysfunction versus nondysfunction but based on the severity versus nonseverity of the symptoms portrayed in the respective vignettes.

The suggestion that disorder was distinguished from nondisorder on the basis of symptom severity immediately runs into some problems. This hypothesis implies that the severity level of the symptoms in the environmental-context vignettes was so low that, despite fully satisfying *DSM-IV* diagnostic criteria, they nonetheless fell below the minimal threshold of severity for intuitive disorder attribution. Yet, there is nothing incoherent about a mild disorder, and people judge that they have mild disorders (e.g., colds, rashes) all the time. Moreover, as noted, all of the study's CD vignettes were designed to fully satisfy *DSM* standards for diagnosis, including the *DSM*'s clinical significance criterion requiring that the symptoms cause impairment in some area of role functioning (e.g., school, family, or job problems), which would seem to place them over any reasonable minimal severity threshold for disorder.

Nonetheless, the objection that the case vignettes confound severity with dysfunction has prima facie merit and cannot be dismissed out of hand. As noted, all case vignettes for a given youth started with the same symptom description, which by itself constituted the symptoms-only vignette, and the internal-dysfunction and environmental-context versions were obtained by adding a paragraph of contextual information to the symptom description. This procedure that separated symptom description from contextual information was designed to control for symptom severity. However, it was flawed, and in retrospect, the additional contextual information aimed at indicating internal dysfunction versus reaction to the environmental context did create potential confounds with symptom severity because the contextual information indicated differential intensity, duration, and generality of symptoms. For example, in the vignette excerpts reproduced above, it is clear that the internal-dysfunction vignettes implied in multiple ways that symptoms occurred across a broader domain of circumstances than in the environment-context vignettes (e.g., with close associates and strangers as well as rival gang members; in the facility versus only at home; in Mexico visiting the grandparents as well as at home), and this could be interpreted

as greater severity. This confound is exceedingly difficult to avoid because indicators of dysfunction, such as situational nonspecificity, irrationality, and disproportionality, are interpretable as indicators of severity. Of course, severity itself might be interpreted merely as an indicator of dysfunction, and if so, the severity hypothesis is not inconsistent with the HDA. However, Murphy and Woolfolk's hypothesis is that it is severity itself, unmediated by internal dysfunction, that could be the basis for disorder attributions, and the published analyses do not address this possibility. So, the "severity" objection needs to be addressed.

The most compelling test of the Murphy and Woolfolk hypothesis would be to hold severity constant and evaluate whether judgments of disorder and dysfunction still go together within the constant severity condition. In theory, one could look within any one vignette condition to hold severity constant, but in fact, the robust responses to the internal-dysfunction and environmental-context vignettes created rather homogeneous response sets (but see below). Consequently, the best way to test the severity hypothesis is to examine the results within the symptom-only vignette condition, which due to its minimal information triggered diverse "disorder" responses and also has the simplest symptom descriptions.

The severity account predicts no particular relationship between disorder and dysfunction when severity is constant, whereas the HDA predicts that disorder and dysfunction judgments should tend to go together even when severity is constant. The point here is not whether subjects judged disorder versus nondisorder but whether their disorder versus nondisorder judgments, whatever they were, went together with dysfunction versus nondysfunction judgments, respectively. To do this test, we defined "congruent" responses to disorder and dysfunction items as follows: a congruent combined judgment was defined as either (1) agree that there is a disorder *and* agree that there is a dysfunction (or something wrong) or (2) disagree that there is a disorder *and* disagree that there is a dysfunction (or something wrong).

The results are that within the symptom-only context, psychologists and nurses responded congruently to "disorder" and "dysfunction" items 80.5% and 74.4% of the time, respectively, both significantly greater than incongruent rates (one-tailed large sample test, $z = 3.90$, $p < .001$; $z = 3.20$, $p < .001$). Clinical and lay "disorder" and "something wrong" responses were congruent 71.4% and 73.2% of the time, respectively, both significant (one-tailed large sample test, $z = 3.93$, $p < .001$; $z = 4.20$, $p < .001$). These results are unexplained by the severity account and make it unlikely that symptom severity independent of dysfunction inference played the sole or dominant role in disorder response.

One might still ask: not just in the symptom-only condition, but in the other vignette conditions as well, did "dysfunction" or "something wrong" judgments tend to go along with disorder judgments, as predicted by the HDA? To answer this question, we reanalyzed the data for the internal-dysfunction and environmental-context vignettes

using as our "items" congruent versus incongruent combined disorder-dysfunction judgments. The results confirm the HDA predictions. Combining internal-dysfunction and environmental-context replies, there was high overall congruence between "disorder" and "dysfunction" items, with nurses' responses congruent 79% of the time, the mental health (clinical social workers and clinical psychology graduate students) sample's responses to "disorder" and "something wrong with the mind" congruent 86% of the time, and the lay sample's responses congruent 84% of the time.

I conclude that these attempts to cast doubt on the CD studies' supportiveness of the HDA fail. The alternative hypotheses proposed by Houts and by Murphy and Wolfolk are inconsistent with the relationships found among both professional and lay subjects' judgments of "disorder," "dysfunction," and "something wrong with the mind."

Epistemology

I now turn briefly to Faucher's discussion of epistemological obstacles to the application of the HDA to psychiatry. Faucher acknowledges at the outset that epistemological concerns have no bearing on whether the HDA is a correct analysis of what we mean by mental disorder and thus are not a critique of the HDA itself. Nonetheless, Faucher points out that to the degree that epistemological obstacles block our ability to distinguish disorders from nondisorders, that will limit how effectively the HDA can accomplish its "normative" role of providing an intellectual justification for reining in diagnostic abuse. The HDA might then meet the antipsychiatric challenge in principle but not in practice.

Faucher focuses on epistemological challenges in establishing evolutionary hypotheses and the fallibility of our judgments about biological design and function. There are indeed limits to our current ability to distinguish disorder from nondisorder due to limitations in our understanding of what is evolved psychological functioning. Like every empirical domain, judgments of biological design and function are fallible and must be subjected to continued scrutiny and testing against alternative theories.

However, the extreme skepticism expressed by Faucher is unwarranted. Faucher objects to the claimed obviousness of many instances of design and function: "the fact that the function of a mechanism seems obvious to us is not a good guide to the evolutionary function of the mechanism or even to its designed nature." This blanket claim is bewildering and itself begs for epistemic support. What percentage of hypotheses that a system is biologically designed have been proven wrong in the history of biology and medicine? Was Aristotle wrong to take acorns growing into oak trees as an obvious example of biological design requiring a special "final cause" explanation? Despite their ignorance of evolution, Hippocrates and Galen surely had adequate grounds for hypothesizing that, say, psychotic delusions, profound melancholia, mania, social phobia, and hysterical paroxysms—all recognized disorders in their time and still so

recognized over two millennia later in ours—were failures of human biological design, whereas erotic love and grief—highly disruptive and often distressing mental states—were not in themselves disorders (although they might trigger disorders). Most historical hypotheses identifying conditions as disordered, of the kind that fill the pages of Hippocrates and Galen, have not been overturned even as our scientific understanding has grown. How is that possible if judgments of design and failure of design are not at least a reasonably good prima facie guide to design and failure of design? Is the *DSM* generally wrong to assume that human thought and perception; human emotional systems such as fear and sadness; human biopsychological systems such as sleep, sex, and eating; and human developmental mechanisms are part of our biological design and thus their failure warrants disorder status? The categories of disorder in the *DSM* are by and large not subtle categories; does Faucher dispute that it is prima facie plausible that such conditions as autism, schizophrenia, bipolar disorder, panic disorder, reactive attachment disorder, conversion disorder, anorexia nervosa, erectile disorder (primary impotence), and many others are likely failures of human biological design? Yes, these judgments can be challenged, as in the case of the "neurodiversity" movement that argues that autism (or at least certain forms higher up on the "autistic spectrum") is a normal variation of cognitive functioning (Wakefield et al. 2020) or in R. D. Laing's (1967) arguments that schizophrenia is a normal response to an abnormal family situation. However, like all scientific disputes, these disputes about whether there is in fact a dysfunction are subject to argumentation on evidential grounds and not epistemically untouchable, despite the limits of our knowledge.

Faucher's claims about great epistemic difficulties afflicting the HDA are based on overly pessimistic assumptions about the intractability of our ignorance. For example, how can you possibly empirically address the debate over whether some children labeled ADHD due to their behavior in school are in fact normally rambunctious or disordered? Here are some indirect ways. First, it has been found that within a class, the youngest students consistently get diagnosed with ADHD more frequently, and the only explanation that makes sense is that relatively younger developmental age is being mistaken for disorder. Also, one can study brain development and see whether there are any abnormalities in those labeled ADHD, and it turns out there are not—but there is slower maturation, consistent with the school-age finding. Also, you can see whether ADHD is associated with an enduring impairment, as is generally assumed, or is transient and would then be better explained as normal variation in developmental rates—as we see in every area of physical and mental development—that is interacting badly with our age-specified educational system. Then, if you are really ambitious, you can examine whether children with a specific genetic variation known to be associated with ADHD are indeed impaired in more nomadic cultures that do not have the lockstep educational system we do (because the main alternative hypothesis to failure of designed attentional systems is that normal children are being overly constrained by

our school environments), and perhaps you find that the very same genetic variation appears adaptive in nomadic groups but maladaptive in settled groups with constricting educational environments—even where the groups are from the same genetic and cultural population otherwise. Such work has a real impact on whether such children are considered disordered or nondisordered (see my reply to De Vreese in this volume for references and details). Scientists are far from helpless; formulating tests that distinguish hypotheses that seem difficult to distinguish is what scientists do for a living.

Moreover, what is epistemically challenging can change radically and relatively quickly (think of longtime impossibilities like being able to see the back of the moon or tell the sex of a child before birth), making it additionally important to keep epistemic and conceptual/ontological issues separate. Actually, this is already happening in the area of biological design. Enormous and dramatic progress has recently been made in evaluating hypotheses about what has been naturally selected in our species. This is because of the remarkable fact that our evolutionary history is largely preserved in our genetic heritage, which only recently has become accessible to detailed analysis. There are now scientific methods that did not exist a few years ago for examining the patterns of genetic loci near a target locus and determining whether it is likely that the target locus was the result of positive selection pressures or not. These methods have been applied to variety of loci, some potentially related to disorder pathogenesis (e.g., Ding et al. 2002; Lind et al. 2019; Polimanti and Gelernter 2017). Such methods of genetic analysis of selective pressures offer a degree of ability to enter into evolutionary inquiry without need for a time machine, in a way unimaginable a few years ago.

Faucher says, "We should beware that we have a natural inclination to make essentialist inferences based on the fact that certain behaviors are deviant to the presence of dysfunction, yet *that does not guarantee* that we are in presence of a dysfunction" (emphasis added). True, we are essentializing fallibly causal-theorizing creatures, and that is one of our major strengths as a species. Faucher is of course correct that such a hypothesis "does not guarantee" that there is a dysfunction, and the correction of such incorrect hypotheses about nondisordered deviance is enhanced by the HDA's analysis of the concepts being deployed. However, one cannot leap from the modest "does not guarantee" point to the claim: "Our intuitions have no special evidentiary status, quite the contrary. In the past, our inclinations or intuitions have shown to be insufficient guides to dysfunction, as cases of masturbation, female orgasm, drapetomania, and so on have proven." If some past mistakes and abuses were sufficient to show "no special evidentiary status," then a parallel argument would wipe away medicine in general and science altogether, for there are similar occasional errors and abuses there. Fortunately, the fact that Victorians allowed their morals to be confused with medical judgments does not reduce the prima facie likelihood that eyes that can't see, hands that can't grasp, thoughts that lack coherence, anxiety unrelated to any threat, and so on are dysfunctions in systems that are biologically designed. Faucher cites my statement that

"one can go wrong in such explanatory attempts," but the fact that one can go wrong using a method does not imply that a method is generally unreliable or that it cannot be used as a beginning for successful bootstrapping to the truth: the fact that there are hallucinations and dreams does not mean that we cannot generally rely on perception for knowledge of the world, and the fact that there are diagnostic errors and abuses does not imply that we are generally wrong about design and function.

Faucher offers two examples of how functions and dysfunctions are not obvious, but his examples also show that they are not beyond our ability to study scientifically. The first is Darwin's own wonderful example of how one can be misled by the unfused skull sutures in the infant, which "have been advanced as a beautiful adaptation for aiding parturition" because they allow the bones of the skull to move and adjust during the birth process and thus allow movement of the baby's large skull through the constrictive birth canal and fuse solidly only at around the age of two years. Darwin observed, however, that, even though the sutures may be "indispensible" to human birth, "as sutures occur in the skulls of young birds and reptiles, which have only to escape from a broken egg, we may infer that this structure has arisen from the laws of growth, and has been taken advantage of in the parturition of the higher animals" (Darwin, as quoted by Gould and Vrba, as quoted by Faucher, this volume). Gould and Vrba use this example to support their distinction between adaptation, which they limit to the original selective pressure that brought about a feature, and what they label "exaptation," which includes subsequent selective pressures that co-opt an already existing feature for new functions. However, structures are routinely exploited for new purposes in evolution, and later selective pressures that maintain a feature are still "adaptations" and instances of "design" or at least "biological functions" in the modern evolutionary sense.

This example does not support Faucher's case. First, the intuition that the sutures are biologically designed is correct. Second, the "obvious" function of the unfused sutures of allowing noninjurious birth is indeed a function, and almost certainly the suture-skull-deformation/birth-canal interaction has been shaped by natural selection and so is biologically designed. Darwin's point is that there is another earlier evolved function of the sutures of allowing unencumbered spherical brain growth. Given these functions, early fusing is a genuine dysfunction and disorder (craniosynostosis). If anything, this example shows that it is possible to identify scientific evidence, including comparative evidence, allowing one to correct false intuitive hypotheses about the evolutionary history of a clearly designed feature.

Faucher's second example concerns the debate over whether *Archaeopteryx* could fly. *Archaeopteryx* was a winged creature that lived about 150 million years ago and had features of both dinosaurs and birds, including feathered wings. It initially seemed obvious by a simple analogy to modern birds that *Archaeopteryx*'s feathered wings as well as its claw and wing structure showed that it was well designed for flight and that it

could fly—indeed, its German name was *urvogel* ("first bird"). Faucher points to the fact that such intuitions are fallible and are open to debate. Contrary structural arguments concerning claw structure, wing structure, and feather structure have been put forward suggesting *Archaeopteryx* was not designed to fly but was more likely a ground-dweller that may sometimes have climbed on trees to evade predators, perhaps assisted by the wings for short-distance soaring. For example, *Archaeopteryx* has a long, heavy unfeathered tail and a flat sternum unlike living birds, which, except for flightless birds, have a keeled sternum to which their large, powerful flight muscles attach—although it has been argued that the *Archaeopteryx*'s large collar bone might have served as such an anchor for flight muscles. It is also generally accepted that feathers initially evolved for thermoregulation, not flying, and so it is arguable whether one or both functions were served by *Archaeopteryx*'s transitional feathered wings.

Based on this dispute, Faucher concludes, "So obviousness of design does not fare better than intuition: it is not a reliable guide to adaptation." This does not follow. Both sides in the dispute over *Archaeopteryx* agree that its feathers are the result of design and have some function. The disagreement is about the precise function. Faucher emphasizes that the function is not necessarily the initially obvious one, and that is often true. Indeed, the functions of even familiar clearly designed processes, such as sleep or grief, may be unclear. In the case of *Archaeopteryx*, it is an extinct creature and so of necessity, the evidence about how its wings functioned is a matter of inference open to dispute. Presumably, Faucher would not have similar doubts that a robin's wings have the biological function of enabling flight. The details of the *Archaeopteryx* debate illustrate the amazing power of scientific method to tackle even remarkably difficult questions of function and design.

In no way do I intend to minimize the very substantial obstacles that exist in many cases to establishing biologically designed human psychological nature, and Faucher points to some of these obstacles. Yet, I see no reason to think that the situation is more dire than that occurring in the initial stages of many other areas of science, including the beginnings of physical medicine 2,500 years ago. The need to take the distinction between biological design and dysfunction seriously is underscored by the fact that it is regularly through bogus attributions of this distinction that socially oppressive uses of medical power are justified. Serious attention to this difficult distinction is thus in the long run a corrective to such exploitation. In sum, Faucher's concerns are real, but they are precisely the sorts of concerns that science is designed to confront and overcome in the long run and is in the process of doing so.

References

Colombo, A., G. Bendelow, B. Fulford, and S. Williams. 2003. Evaluating the influence of implicit models of mental disorder on process of shared decision making within community-based multidisciplinary teams. *Social Science and Medicine* 56: 1557–1570.

Cummins, R., and M. Roth. 2009. Traits have not evolved to function the way they do because of a past advantage. In *Contemporary Debates in Philosophy of Biology,* F. Ayala and R. Arp (eds.), 72–86. Blackwell.

Ding, Y. C., H. C. Chi, D. L. Grady, A. Morishima, J. R. Kidd, K. K. Kidd, et al. 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proceedings of the National Academy of Sciences of the United States of America* 99(1): 309–314.

Faucher, L., and I. Blanchette. 2011. Fearing new dangers: Phobias and the cognitive complexity of human emotions. In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory,* P. R. Adriaens and A. De Block (eds.), 33–64. Oxford University Press.

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.

Harland, R., E. Antonova, G. S. Owen, M. Broome, S. Landau, Q. Deeley, and R. Murray. 2009. A study of psychiatrists' concepts of mental illness. *Psychological Medicine* 39: 967–976.

Houts, A. C. 2001. Harmful dysfunction and the search for value neutrality in the definition of mental disorder: Response to Wakefield (part 2). *Behaviour Research and Therapy* 39: 1099–1132.

Kirk, S. A., J. C. Wakefield, D. Hsieh, and K. Pottick. 1999. Social context and social workers' judgment of mental disorder. *Social Service Review* 73: 82–104.

Klein, D. F. 1978. A proposed definition of mental illness. In *Critical Issues in Psychiatric Diagnosis,* R. L. Spitzer and D. F. Klein (eds.), 41–71. Raven Press.

Laing, R. D. 1967. *The Politics of Experience*. Penguin Books.

Lind, A. L., Y. Y. Y. Lai, Y. Mostovoy, A. K. Holloway, A. Iannucci, A. C. Y. Mak, et al. 2019. Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nature Ecology & Evolution* 3(8): 1241–1252.

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

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.

Pottick, K. J., J. C. Wakefield, S. A. Kirk, and X. Tian. 2003. Influence of social workers' characteristics on the perception of mental disorder in youths. *Social Service Review* 77: 431–454.

Rind, B., P. Tromovitch, and R. Bauserman. 1998. A meta-analytic examination of assumed properties of child sexual abuse using college samples. *Psychological Bulletin* 124(1): 22–53.

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.

Spitzer, R. L., and J. Endicott. 1978. Medical and mental disorder: Proposed definition and criteria. In *Critical Issues in Psychiatric Diagnosis,* R. L. Spitzer and D. F. Klein (eds.), 15–39. Raven Press.

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: 233–46.

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.

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. 2013. The *DSM-5* debate over the bereavement exclusion: Psychiatric diagnosis and the future of empirically supported practice. *Clinical Psychology Review* 33: 825–845.

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.

Wakefield, J. C., A. V. Horwitz, and L. Lorenzo-Luaces. 2017. Uncomplicated depression as normal sadness: Rethinking the boundary between normal and disordered depression. In *Oxford Handbook of Mood Disorders,* R. J. DeRubeis and D. R. Strunk (eds.), 83–94. Oxford University Press.

Wakefield, J. C., S. A. Kirk, K. Pottick, and D. Hsieh. 1999. Disorder attribution and clinical judgment in the assessment of adolescent antisocial behavior. *Social Work Research* 23: 227–241.

Wakefield, J. C., S. A. Kirk, K. J. Pottick, X. Tian, and D. K. Hsieh. 2006. The lay concept of conduct disorder: Do non-professionals use syndromal symptoms or internal dysfunction to distinguish disorder from delinquency? *Canadian Journal of Psychiatry* 51: 210–217.

Wakefield, J. C., K. J. Pottick, and S. A. Kirk. 2002. Should the *DSM-IV* diagnostic criteria for conduct disorder consider social context? *American Journal of Psychiatry* 159: 380–386.

Wakefield, J. C., and M. F. Schmitz. 2013. When does depression become a disorder? Using recurrence rates to evaluate the validity of proposed changes in major depression diagnostic thresholds. *World Psychiatry* 12: 44–52.

Wakefield, J. C., and M. F. Schmitz. 2014. Predictive validation of single-episode uncomplicated depression as a benign subtype of unipolar major depression. *Acta Psychiatrica Scandinavica* 129: 445–457.

Wakefield, J. C., M. F. Schmitz, M. B. First, and A. V. Horwitz. 2007. Extending the bereavement exclusion for major depression to other losses: Evidence from the National Comorbidity Survey. *Archives of General Psychiatry* 64: 433–440.

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