

This PDF includes a chapter from the following book:

Defining Mental Disorder

Jerome Wakefield and His Critics

© 2021 Massachusetts Institute of Technology

License Terms:

Made available under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International Public License
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

OA Funding Provided By:

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

The title-level DOI for this work is:

[doi:10.7551/mitpress/9949.001.0001](https://doi.org/10.7551/mitpress/9949.001.0001)

3 Facts, Facts, Facts: HD Analysis Goes Factual

Luc Faucher

Wakefield's harmful dysfunction (henceforth HD in what follows) analysis has received well-deserved attention in psychiatry and philosophy. Although many commentators have considered Wakefield's analysis to be correct (or at least on the right track), there is also no shortage of criticism as to his position. I want my contribution to be of a different "flavor" than what is usually found in literature on Wakefield (with the exception of De Block 2008). In fact, I want to indicate a few places where his analysis could make fruitful contact with empirical research. To be more specific, I want to accomplish two things in this chapter. First, I hope to show that "going factual" might reveal why disputes about which analysis of the concept of mental disorder is the right analysis are so endemic in both psychiatry and philosophy. Indeed, Wakefield claims that his analysis is the result of a conceptual analysis. After the recent wave of X-Philosophy investigations, one becomes suspicious of what is presented as an armchair analysis of a concept. I want to suggest that we might want to "go X" in the case of the concept(s) used in psychiatry too and conduct empirical investigations of the concept of mental disorder that ordinary individuals and specialists have (as I will demonstrate, Wakefield is aware of the limits of conceptual analysis and has "gone X"). I want to propose a few experiments that should be performed in order to probe our concepts about mental disorders, but I also want to present different techniques available to the X-Philosopher and suggest that some techniques might be more adequate to investigate our concept of mental disorder. These techniques might reveal a diversity of ways to conceptualize mental disorders that explain why discussions on this topic have endured. Second, I want to argue that an empirical attitude might make the dysfunction component of the analysis more speculative (but not entirely so) than Wakefield seems to think. Indeed, Wakefield has suggested that hope for a more "factual" psychiatry lies in the development of evolutionary psychiatry. I want to show that, at least for some disorders, prospects are not promising in terms of arriving at something other than speculation about mechanisms that are malfunctioning. This does not put into question Wakefield's conceptual analysis, but shows that it might be of limited use in certain cases.

Before discussing these points of contact, let me briefly summarize Wakefield's analysis.

I. HD Analysis

In a series of papers (1992a, 1992b, 1993, 1999, 2000, 2007a), Jerome Wakefield proposes a definition of the concept of "mental disorder" that he hopes will provide psychiatry with an objective criterion for declaring a mental condition a disorder. His definition is presented as an "explicitation" of the intuitive concept of disorder used not only by health professionals (in medicine in general and in psychiatry) but also by the general public (e.g., Wakefield et al. 2006, 212; Wakefield 2007a, 150). Wakefield has also argued that his HD analysis "has proven useful for thinking about the validity of diagnostic criteria; in particular, explicitly formulating issues in terms of function and dysfunction seems to help identify false positives and limit undue diagnostic expansiveness" (2006, 159).¹ As Jeffrey Poland (2003) puts it, Wakefield's HD analysis has two distinct aspects: a descriptive one (which consists of adequately capturing and reconstructing our shared concept of mental disorders) and a normative one (which consists of using the shared concept to evaluate current diagnostic criteria).

Wakefield believes that we can analyze the intuitive concept of "mental disorder" underlying the field of psychiatry (and our ordinary judgments about who's disordered and who's not) by saying that it is the result of a "dysfunction" of a psychological or mental mechanism² that is judged "harmful" (e.g., 1993, 163). This definition is a hybrid account of disorder for it has both a purely scientific and factual component (the notion of dysfunction) and a value component (the notion of harm). According to Wakefield, both of these components are jointly necessary to capture our intuitive concept of mental illness (1992a, 374). Wakefield has little to say about the "value" component of his definition³; he is far more interested in the notion of dysfunction that, he expects, will provide psychiatry the objective foundations it needs.

Although the notions of "function" and "dysfunction" or "malfunction" have been used in medicine and psychiatry for a long time, according to Wakefield, only evolutionary theory can analyze these in causal and scientific terms. Thus, he proposes understanding previous uses of function as cases of what he calls "black-box essentialism." This theory is an extension of Putnam's theory of reference that asserts that we use concepts on the basis of defeasible stereotypical properties (e.g., using "water," the fact of being transparent, liquid, drinkable, etc.) before the underlying essence of what we refer to is scientifically discovered (e.g., the concept of "water" existed long before we finally discovered its underlying essence, in this case, its molecular structure). Wakefield's idea is that the notion of function (and malfunction) used by Aristotle, Harvey, and others has been based on prototypical instances of "non-accidentally beneficial effects like sight [in the case of the eyes] and on the idea that some common underlying process must be

responsible for such remarkable phenomena” (2000, 39). However, the process responsible for such phenomena was unknown until the advent of Darwinian theory. Since then, though, one must make reference to evolutionary biology in order to determine the “real” function of mental mechanisms that are supposed to be dysfunctional. As Wakefield puts it, “[If] the HD analysis is at least roughly correct, however, then validly distinguishing disorder from non-disorder depends on an evolutionary-functional analysis” (2006, 158; see also 1993, 170). One can thus understand that Wakefield is making two separate claims: (1) the correct analysis of our concept of mental disorder has to be made in terms of harm and dysfunction of a mental mechanism, and (2) the correct understanding of the concept of dysfunction is in terms of evolutionary function. For this reason, some (e.g., Poland 2003) distinguish HD analysis from HDW analysis because one could agree with (1) and not with (2) (see, e.g., Woolfolk 1999; Roe and Murphy 2011).

I will say a few words about the evolutionary analysis of function. According to evolutionary theory, the presence of certain traits (including psychological mechanisms responsible for behaviors) is explained by the fact that these traits (or mechanisms) performed certain functions in the organisms’ ancestors, the effects of which had been beneficial enough for the organisms’ ancestors to preserve them in their species through natural selection. The function for which a trait (or a mechanism) had been selected is what has been called in philosophical literature the “normal function” or “proper function” of that trait (or mechanism) (Millikan 2002; Neander 1991). In other words, the normal or proper function of mechanism X is to do what it has been designed to do by natural selection. It follows that there is a dysfunction or a malfunction when a trait (or a mechanism) is not able to properly accomplish its normal function. It should be noted that the notion of “normal function” is independent of the current adaptivity of the trait (or of the mechanism). Thus, the fact that a trait (or mechanism) is maladaptive in a current environment is not necessarily a sign of a dysfunction. To use one of Wakefield’s examples, the fact that we are not capable of breathing under water is not an indication of a malfunction of the lungs, but rather of the fact that lungs can’t perform their function in environments for which they have not been designed.⁴ It should also be noted that the notion of function is presumed to be independent of our values.⁵ For instance, imagine that killing and rape have been found to be produced by mechanisms that have been selected for (as some would argue; for references to killing, see, e.g., Buss 2006; for rape, see Thornhill and Palmer 2000). If such were the case, we would have to judge that, when the mechanisms responsible for these behaviors are in good working order, these behaviors are adaptive, even though we abhor them.

II. Conceptual Analysis as a Form of Empirical Psychology

As mentioned earlier, Wakefield sees his analysis as the result of a form of conceptual analysis of concepts used not only by those working in the psychiatric field but also by

ordinary people when they are judging a condition to be disordered or not. Conceptual analysis is a philosopher's tool and seeks to provide the set of necessary and sufficient conditions for the application of a concept. Proposed analyses are usually tested against the philosopher's intuitions or against what philosophers think individuals' intuitions are (see, e.g., Horwitz and Wakefield 2012, 79). Indeed, generally, philosophers use imagined cases where one condition deemed to be necessary for the application of a concept would be missing, or where all conditions would be present, and see if the concept is judged to apply in those circumstances or if it generates counterintuitive consequences and is judged not to apply. For instance, Aristotle famously proposed an analysis of the concept of responsibility in which someone is said to be responsible if and only if they meet two conditions: the knowledge condition (someone has to know what they are doing) and the control condition (someone has to be in control of what they are doing). A philosopher could test this analysis by imagining cases where someone does not know what they are doing (e.g., someone becomes entangled in an argument and forgets that they left their dog in the car) or where that person knows what they are doing but has no control over it. An analysis is said to be valid if it agrees with one's intuition and invalid if it disagrees with intuition (at least when applied to clear cases). Thus, intuitions—more specifically, the philosopher's intuitions—are considered “evidence” in determining the adequacy of a conceptual analysis.

Wakefield rightly sees this kind of investigation as a form of empirical psychology. As he puts it, “Our conceptual enterprise is also an empirical enterprise aimed at discovering a certain fact about the world, namely what conceptual criterion or definition in the heads of people in our linguistic community ultimately determines and explains their judgments about whatever conditions are mental disorders” (1997, 257).⁶ He also identifies quite well one of the premises on which conceptual analysis relies when he writes that “the process of conceptual analysis does not look empirical because one generally uses one's own intuitions about the clear cases rather than going out and collecting data. However, *this oddity results from the presupposition that one is dealing with a culturally shared concept, and the confidence that one's clear intuitions about the application of the term are likely to be shared*” (1997, 257, my emphasis).

But, one might wonder, what are the grounds for such a presupposition? After all, the individual performing the analysis (in our case, Wakefield or another philosopher of psychiatry) might not be representative of others' intuitions about a concept. His intuition might be idiosyncratic or group or culture relative. Moreover, the method of conceptual analysis (which usually describes rather abstract cases and seeks to describe explicit concepts) might not be able to unveil certain aspects of our conceptual knowledge (e.g., implicit aspects of our knowledge) or the variability of our concepts in different contexts.⁷

These kinds of shortcomings in conceptual analysis inspired a new movement in philosophy: experimental philosophy (or X-Phi in what follows). As one of the early

advocates of this form of philosophy claims, “Experimental philosophy focuses on many of the same types of intuitions that have long been at the center of philosophical study, but it examines those intuitions using the methods associated with contemporary cognitive science—systematic experimentation and statistical analysis” (Knobe 2007, 81). Such methods have been used in philosophy of mind, epistemology, ethics, aesthetics, philosophy of race, philosophy of science, philosophy of language, and many more areas of philosophy, leading to many unexpected results. Allow me to briefly describe to you what Knobe has in mind when he talks about X-Phi using two examples.

The first example comes from the work of Nichols and colleagues (2003) on epistemic intuitions. They wanted to test the classical conceptual analysis of the concept of knowledge as justified true beliefs (so in order to have knowledge, it is not sufficient to have a true belief, you must have a *justified* true belief). In order to test this analysis, Nichols et al. submitted the following vignettes to members of different ethnic communities: “Bob is a friend to Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?” (234). They then asked subjects the following question: “Does Bob really know or does he only believe that Jill drives an American car?” The results were quite surprising as 74% Western-heritage subjects said that Bob *only believes* that Jill drives an American car while 57% of subjects from East Asia and 61% of subjects from India said that he *really knows* that Jill drives an American car. If the results of this study were to generalize, it would prove that the concept of knowledge is understood differently in different cultures and that one cannot extend their conceptual analysis to other cultures.⁸

My second example involves the concept of responsibility. Intrigued by results from Nahmias (2006; Nahmias et al. 2005) that demonstrate that, contrary to philosophers’ expectations, people judged an agent as morally blameworthy even if they have performed their immoral action in a deterministic universe, Nichols and Knobe (2007) designed a series of experiments to test people’s intuitions about moral responsibility. They probed the intuitions of subjects by presenting them randomly one of two vignettes concerning a universe (Universe A) in which every event unfolds according to deterministic laws. Subjects in the “abstract condition” were given the following question:

In Universe A, is it possible for a person to be fully morally responsible for their action?

Subjects in the “concrete condition” were given the following question:

In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible

to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family. Is Bill fully morally responsible for killing his wife and children?

The results were that only 5% of subjects in the abstract condition said that the agent (Bill) was fully morally responsible, while 72% of subjects in the concrete condition said he was fully morally responsible.

Experiments such as the following have led Knobe and Doris (2010) to argue that recent work in experimental philosophy about moral responsibility shows that our attribution of moral responsibility depends on and changes as a function of certain contextual variables. They call this idea “variantism,” as opposed to “invariantism.” They argue that invariantism has dominated philosophical discussions up to now and consists of the idea “that people should apply the *same* criteria in *all* of their moral responsibility judgments. In other words, it is supposed to be possible to come up with a single basic set of criteria that can account for all moral responsibility judgments in all cases” (322). However, they observe that “it seems that people do not make moral responsibility judgments by applying invariant principles. Instead, it appears that people tend to apply quite different criteria in different kinds of cases. Thus if one wants to understand why people make the judgments they do, it is no use looking for a single basic set of criteria that fits all people’s ordinary judgments. A more promising approach would be to look at how and why people may adopt different criteria in different cases, depending on the way an issue is framed, whether an agent is a friend or a stranger, and so on” (322).

Thus, as these two examples show, X-Phi takes seriously the idea of empirically probing the intuition of ordinary people (but also of scientific communities; see, e.g., Griffiths and Stotz 2008) and does so by applying methods inspired by psychology and the social sciences to do it. By so doing, it has revealed surprising and puzzling facts about people’s intuitions. So great is the success of X-Phi that one could even argue that it’s now the only way to probe people’s intuitions. Given this, it is natural to propose that Wakefield should go the way of X-Phi. Given the limits of traditional conceptual analysis, Wakefield should apply X-Phi methods to his topic of interest. Interestingly, this is precisely what the *Research Agenda* (Kupfer et al. 2002) suggested back at the beginning of the process that was to lead to the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. As the authors of the agenda stated, the question of what is the concept of mental illness used by clinicians could be addressed by

conduct[ing] surveys... to elucidate the concepts of disease or of mental illness or disorder used, explicitly or implicitly, by psychiatrists, other physicians, clinical psychologists, research workers, patients, health care providers, and members of different social and ethnic groups. This could be done either by exploring the meaning they attribute to such terms or by asking them to decide which of a list of contentious conditions they themselves regarded as disease or mental disorders. (7)

One will probably be surprised to learn (as I was initially) that Wakefield actually went X! In a pilot study (Kirk et al. 1999) and another study (2002), Wakefield and his colleagues tested social workers' judgments of mental disorders. More precisely, they tried to determine "to what extent do social workers use contextual information to distinguish between disordered and non-disordered adolescent antisocial behaviour?" (Kirk et al. 1999, 84). The question that Wakefield was interested in was the following: are social workers making the distinction between someone whose symptoms are normal reactions to their social environment (and therefore is not disordered) and someone whose symptoms are not explained by their social environment but by internal dysfunction (and who therefore are suffering from a disorder), or are they "reluctant to pathologize deviant and nonconforming behaviours and would instead have a tendency to see them as normal responses by people in stressful or oppressive social environments" (Kirk et al. 1999, 85)?

In order to answer this question, Wakefield and his colleagues asked subjects to read three kinds of vignettes that they identify as "neutral," "environmental reaction," and "internal disorder." "Neutral" just describes a case where one youth, "Carlos," has three key symptoms of antisocial personality disorder. "Environmental reaction" asserts that environmental factors might explain why it was rational and adaptive to behave that way. Finally, "internal disorder" suggests that his reactions are the result of neither a rational nor an adaptive strategy. Here's an example of a vignette (for the sake of space, I am presenting an abridged version; for a complete description, see Kirk et al. 1999, 91):

Carlos' family appears to be stable and caring. Carlos attends a respected public junior high school that has very little violence and provides a secure learning environment. However, Carlos reacts to the slightest perception of provocation with severe anger. Once he gets angry he often escalates fights from fists to weapons like bats and bricks even when the other boy wants to stop. Carlos consistently ignores his teachers' requests and discipline seems to only exacerbate his problematic behaviour. Even with those he hangs out with, Carlos is easily irritated and frequently initiates fights.

They then asked subjects (master's students in psychology or social work programs) if they agree or disagree with the following item⁹ (they had to reply using a Likert scale ranging from 1 = "I strongly agree" to 6 = "I strongly disagree"): "According to my own view, this youth has a mental has a mental/psychiatric disorder."

The results were quite univocal: when the symptoms could be explained by the social context, only 2.9% of the subjects thought that the youth had a mental disorder; when the symptoms were not explained by the social context and could indicate an internal disorder (condition "internal disorder"), 95.2% agreed that he was suffering from a mental disorder. In a similar experiment, performed on nonprofessionals (non-psychiatric nurses, nonclinical social workers, undergraduates), they received somewhat similar results: with 7.9% of undergraduates saying that Carlos had a disorder in the environmental reaction vignette and 73.9% stating he had a disorder in the

internal disorder vignette. Wakefield and colleagues conclude from these studies that “as predicted, identical behavioural symptoms meeting the *DSM-IV* criteria for conduct disorder are judged as less indicative of the presence of mental or psychiatric disorder in the environmental reaction context than they are in the internal dysfunction context” (Kirk et al. 1999, 92).

If these results are suggestive that Wakefield’s analysis is valid, I want to argue that they might not be decisive for three different reasons (some of them recognized by Wakefield and his colleagues). First, the sample being tested is composed of students in psychology and social work, nonpsychiatric nurses, nonpsychiatric social workers, and undergraduates. This sample might be representative of the professional community and share its concepts of disorders, but it is not at all obvious that it allows the probing of the layperson’s concept (after all, students in psychiatry or social work are being trained to think about disorders in a certain way, while nurses and social workers have also been taught, even if their subject matter concerns physical disorders). Studying American undergraduates, even if they never did psychiatry classes, is not necessarily a better way to probe the layperson’s mind. Indeed, the idea that studying students’ intuitions or cognition is a way to study the layperson’s intuitions or cognition has been contested by many. In a now famous paper, Heinrich and his colleagues (2010) reviewed reasons why researchers should “worry about the representativeness of prevalent undergraduate samples in behavioural sciences... [because, as studies show] the sample of contemporary Western undergraduates... is frequently a distinct outlier vis-à-vis other global samples” (22). According to Heinrich et al., undergraduates are not only different from people from other cultures (e.g., Westerners are more likely to explain behavior in decontextualized terms and allude to internal causes, more so than East Asians, who explain behavior in more holistic terms¹⁰) but also from members of their own culture: for instance, from people belonging to different social classes or generations. Therefore, Wakefield’s studies might suffer from a sampling bias.¹¹ Second, Wakefield proposed a version of HD (HDW) according to which disorders are identified by a default of their designed or proper function (whoever or whatever gave them that function). The previous experiment tested only HD. To conclude that it tested HDW, 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. Finally, in the experiment, subjects have to react to an abstract situation described by vignettes and are forced to choose an answer (there is no option to abstain). One might wonder if this kind of reaction is representative of all aspects of their conceptual representations or even if it captures concepts that are used or that are more operative in practice. Many philosophers (Cullen 2010; Kauppinen 2007; Woolfolk 2013) have criticized X-Phi for its exclusive reliance on self-report methods based on questionnaires of the sort used by Wakefield. According to these researchers,

questionnaires are open to numerous distortions due to wording, framing of questions, question order, and so on. For instance, Cullen (2010) provided subjects vignettes that Nichols and colleagues (2003, 288) used but framed the questions differently. Instead of asking “Does X really knows” or “Does X only believes,” he asked “Does X know” or “Does X not know.” Instead of having a majority of individuals saying that X “only believes,” as was the case with the Nichols et al. experiment, Cullen’s results were that a majority of people stated that X “knows.” This is not to say that vignettes and self-reporting are not good tools for probing concepts, but rather that a more comprehensive view that would include naturalistic clinic data, surveys, and structured interviews may be more desirable.

In order to overcome these shortcomings, I propose two things.

First, I would test a representative sample of the general American population but also people belonging to different cultures, using vignettes seeking to verify how the two components of HDW fare with them. I would therefore use a series of vignettes where a trait has a designed function (for simplicity’s sake, below I am using natural selection as the “designer” of the function, but it can be replaced by “God” or by “Nature,” if needed), but where the function is now maladaptive (so I would try to see if something can be judged functional but considered disordered). I would also create vignettes where someone would have a disorder with a designed function but that causes no harm to the subject (I would suppose that if subjects say that someone is disordered even if there is no harm done to them, that goes against HD analysis). I would then test the intuitions of my sample.

Here are a few examples of the kind of questions I would ask. In the group of designed function now considered maladaptive, I would for instance offer subjects an “internal disorder” vignette like Wakefield used to describe Carlos, to which I would add the following:

Psychologists have found that it has been adaptive in the environment of our ancestors, for a small fraction of the population, to show a high level of aggressiveness and to have a very short fuse. Therefore, natural selection has conserved this trait in the population by selecting the genes responsible for it. Carlos is known to have such genes.

Then I would ask, “Is Carlos suffering from a mental/psychiatric disorder?”

I would also try the same sort of vignette with a disorder such as attention-deficit/hyperactivity disorder (ADHD). For instance (following a suggestion made by Panksepp 2007), I would propose vignettes describing children with typical symptoms of ADHD to which I would add the following:

In the environments of our ancestors, it was adaptive for some children to move a lot and to have a very short attention span in order to be able to respond to ever-changing environmental conditions. These children were, biologically speaking, identical to children suffering from ADHD.

Then I would ask, “Are these children (who are similar to our ancestors) suffering from a mental/psychiatric disorder?”

I would also offer subjects more intricate vignettes, like the following inspired from De Block (2008):

Seasonal affective disorder (SAD) is an adaptive pattern of responses that contributes to the individual's reproductive success in higher latitude regions.

Heidi is born in Sweden. Mild SAD was part of her ancestors' phenotype. At one point, Heidi moved to Africa. Like every winter in Sweden, she experienced mild symptoms of SAD, even if it was not adaptive. These symptoms caused her great pain.

Then I would ask, “When in Africa, is Heidi suffering from a mental/psychiatric disorder?”

For the second set of vignettes, those testing the harmful component, I would propose vignettes such as what follows where someone has a disorder, but it is not seen as harmful (inspired by Sacks 1997):

There are some people living on an isolated Micronesian island who are incapable of identifying or recognizing human faces and emotional responses from faces, and for that reason, they cannot form friendships or interact with people. On the other hand, they are better than anyone else at identifying comestible varieties of plants and mushrooms. Until now, this state has been beneficial to them in allowing them to occupy a niche (they are exceptional gatherers) in which they are now thriving. For instance, because of their capacity to identify plants and mushrooms, they turn out to be the most prosperous inhabitants of the island. It appears that, despite their difficulty in forming friendships, they haven't suffered from loneliness or isolation because they are so absorbed by their work.

Then I would ask, “Are people on this island suffering from a mental disorder/psychiatric disorder?”

I'll stop here with my suggestion of vignettes, but it seems plain to me that not only Wakefield but also philosophers, psychiatrists, cultural anthropologists, and sociologists should engage in that kind of enterprise. I'll explain why after my second suggestion.

My second suggestion is inspired by Colombo and colleagues' (2003; Fulford and Colombo 2004) series of studies. In these studies, they used a different technique than Wakefield and his colleagues. As Wakefield did, they gave vignettes to subjects from distinct groups (psychiatrists, psychiatric nurses, patients, informal caregivers, social workers, etc.) to read. These vignettes are longer and describe someone who shows symptoms of schizophrenia, with their background (life situation, childhood, etc.). Then, instead of asking subjects to make a choice between stated options as in Wakefield's experiment, researchers conduct interviews in which they ask this group of people about possible etiology, the individual's level of responsibility, how the individual should be treated, and so on. 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. Colombo et al. used this method because they adopted a “linguistic-analysis model” of concept (inspired by Wittgenstein) where the content of a concept is not exhausted by what one says about it but rather by what does with it, how it is used. According to them,

Asking people, whether professionals or users, directly about mental disorders will elicit, mainly, their explicit views. The most familiar explicit model, nowadays, is perhaps the so-called “biopsychosocial” [medical] model. ...If the linguistic-analytic insight is right, on the other hand, if such concepts use is a surer guide to meaning than explicit definition, then... how they actually respond to... mental disorders, will be driven by their implicit models of disorder. (Fulford and Colombo 2004, 136)

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 workers agreed with it, instead showing preference for the social model (47.5%).

What this set of studies reveals is that different people (and different professional groups) seem to use different models 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. This makes it even more important that we also use and offer subjects the kinds of vignettes I have proposed (varying different aspects of them—like wording, framing, part of the context, or even mood) and apply different kinds of methods to different kinds of people (professional groups/culture/social classes, etc.). It is possible that, after performing such research, Wakefield would find that laypeople’s intuitions and/or some professional people’s intuitions are incompatible with his particular brand of harmful dysfunction analysis (HDW). He could discover that laypeople’s intuitions are not compatible with HDW (or not always compatible with it), while professionals’ intuitions are. If such were the case (I am not claiming that it is, but it is a possibility and until we have run more experiments and probed more deeply the minds of different kinds of people, it is hazardous to claim that there is only one concept shared by everyone), which concept should be preferred? Should one be preferred? On which basis should we decide on one or another? In replying to these questions, one would not be able to avoid normative considerations. I think these that these considerations would have to be invoked even if at the end of the day Wakefield was right and that there was only one concept of mental disorder shared by everyone. If such were the case, one would need to explain the origin of the consensus somehow. For instance, one might invoke psychological dispositions that inclined us to think about mental illness

the same way. But why should we assume that our natural way to think about mental disorder is the right one? Our natural inclinations to think about physical phenomena in a way that goes against our best science (McCloskey 1983) and our natural essentialism about species are an obstacle to a correct understanding of biology (Gelman 2010). What this shows is that even if Wakefield's description of our concept was correct, he would still have to invoke normative considerations to explain why we should adopt this concept in psychiatry.

III. Evolutionary Psychology to the Rescue

In this section, I want to return to an objection that has been previously made to Wakefield's analysis: the so-called epistemic objection. According to this objection, "because the harmful dysfunction analysis holds that whether one has a disorder depends on facts about internal mechanisms and their evolutionary history, and we are largely ignorant of these facts, therefore the analysis implies that it is impossible to know at this time whether conditions are disorders or non-disorders" (Bergner 1997, 255). Wakefield rightly points out that this objection does not target his conceptual analysis (see also his reply to McNally 2001, 349). Indeed, those who endorsed the epistemic objection could grant that Wakefield's HDW analysis is correct, but they would point out that the adoption of an evolutionary framework would be impracticable for psychiatry. So the epistemological objection is important because it is hard to see how Wakefield's analysis could play its normative role without a picture of the normal mind.¹² Indeed, if one is to criticize the *DSM's* criteria of depression or anxiety because it includes cases of normal sadness or normal fear (i.e., of nondysfunctional reactions to some stimuli), one should have a means to tell what is normal from what is disordered. But where would that knowledge come from? It is thus important that Wakefield provides an answer to the epistemic objection, even if it does not touch upon the core of his conceptual analysis.

One can find three different replies to the epistemic objection through Wakefield's writing: (1) we have intuitions about what is and what is not functioning correctly, (2) sometimes the function of a mechanism is obvious, and finally (3) sometimes we need help from evolutionary psychology to identify the function of known components of our mind or to identify these very components.¹³ In this section, I want to argue that there are problems with (1) and (2) and that we should thus rely on (3). However, I will also show that for a class of mechanisms, reply (3) won't be able to deliver the expected results to us. If such is the case, there will be a hole in the middle of our nosology concerning mechanisms that might be crucial to understanding some disorders.

Let start with answer (1). Wakefield writes that the fact that we do not know about the precise design of an artifact does impinge our judgment that we are facing a dysfunction. For instance, he writes that "I do know almost nothing about the design of automobiles, but I am perfectly capable of recognizing many cases of automobile

malfunction and regularly discriminate such cases from proper automotive functioning" (1997, 256; see also 2001, 349). That is, I can discriminate cases where the car doesn't start because the tank is empty (in which case, the car is not broken) from cases where it does not start because something is broken. Basically, I can recognize that something is disordered or malfunctioning even if I know next to nothing about its exact function. But I see two problems with this answer. First, while I can sometimes distinguish cases of malfunction from proper functioning, my understanding of what went wrong is still very rudimentary—I still need to go to the garage to know what is wrong with my car. The fact that my car sometimes still has problems even when exiting the garage proves that it is not necessarily easy to identify these problems. Likely, I might have intuition that this behavior is not normal, that it must be produced by a malfunctioning mechanism, but without a precise knowledge of which mechanism, it is hard to develop or apply adequate treatment. What I mean to say (and I am sure Wakefield will agree) is that for psychiatry to fulfill its role(s) (identifying disorders for diagnosis and prognosis, guiding research, providing indications as to which treatments will work better, etc.), the discipline will have to move away from intuitions to a more scientific basis. At best, intuitions can be a starting point for scientific inquiries into the mechanisms responsible for disorders. Second, relying on intuitions to distinguish malfunctioning from proper functioning is a dangerous game. This is why cars are often equipped with an indicator on the dashboard to signal problems with the engine. Often times, we have no clue that something is going wrong with the engine. Now take an artifact that you do not know very well. For instance, when I got my first iPod years ago, I tried to use it while running. Yet as soon as I was running, the iPod would skip from song to song. I thought there was a problem with the iPod and was ready to throw the thing away (or to go back to the store) until I finally discovered that I had to shut the screen off before starting off on my run. In that case, I was thinking that there was something wrong with my iPod because I did not know its proper functioning. The human mind might be more like the new gadget for which you haven't taken the time to read the instruction book than to a car, as we don't know much about how it is supposed to be functioning (this is a claim that evolutionary psychologists frequently make; see, e.g., Cosmides and Tooby [1994], who discuss the power of evolutionary psychology to go beyond intuition and instinct blindness). Therefore, 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. 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 the cases of masturbation, female orgasm, drapetomania, and so on have proven. As Wakefield himself claimed, ignorance of the facts about the functions of mental mechanisms has left the door open to the use of social norms or values and leads to classifying behaviors that were normal as pathological.¹⁴

So let's turn to reply (2). Wakefield writes that sometimes, the function of a mechanism is obvious. For instance, in his first book with Horwitz (Horwitz and Wakefield 2007), he writes that "in some cases, a mechanism's biological function is immediately obvious: for example, it cannot be accidental that the eyes see, the hands grasps, the feet walk, or the teeth chew, and it is clear that these beneficial effects explain the existence via natural selection of the respective mechanisms," and he continues, "Sadness is somewhat like sleep; the function is not obvious, yet the designed nature is" (47). This claim deserves a few remarks. First, 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. Take the case of sutures in the skulls of mammals. As Darwin observed long ago, "The sutures of the skulls of young mammals have been advanced as a beautiful adaptation for aiding parturition and not doubt the facilitate, or may be indispensable for this act; but 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" (quoted by Gould and Vrba 1982, 5). To use another example, take Dennett's claim that it is obvious that the *Archaeopteryx* was designed for flight: "Did Archaeopteryx, the extinct birdlike creature that some have called a winged dinosaur, ever really get off the ground? ... *An analysis of the claw curvature*, supplemented by aerodynamic analysis of Archaeopteryx wing structure, *makes it quite plain that the creature was well designed for flight*" (1995, 233, my emphasis). Unfortunately, things are not that simple. Some (Nudds and Dyke 2010) have argued that *Archaeopteryx* was not designed to fly. For instance, they suggest that the central shaft of the feathers is thinner and weaker than required by modern birds to fly.¹⁵ Moreover, as Naish observed, "Claims that Archaeopteryx possesses a claw geometry indicating an arboreal lifestyle (Feduccia 1993) are contradicted by newer analyses (Hopson 2001; Glen and Bennett 2007), and virtually all non-avian maniraptorans lack features indicative of a climbing lifestyle" (Naish 2011, 435¹⁶). If such is the case, *Archaeopteryx* would more resemble birds that stay on the ground (ground dwellers) and sometimes climb on trees (maybe assisted by their wings) to evade predators (like chickens) than flying birds (Richardson 2007, 49).

So obviousness of design does not fare better than intuition: it is not a reliable guide to adaptation. Wakefield agrees with this as he writes, "Obviously, one can go wrong in such explanatory attempts; what seems non-accidental may turn out to be accidental," but he adds that "often one is right" (1992a, 383). But one might ask, where is the evidence that we are often right? How do we know we are right? If there is no such evidence, we would again be relying on our intuition, which can be misleading.

So it seems that the only way to answer the epistemological objection is to turn to a discipline able to identify the mental mechanisms that our mind comprises and describe their functions, as well as their normal environment of functioning: evolutionary

psychology (henceforth EP; this is reply (3)). According to Wakefield, only EP is able to reveal our “human species-typical biological design” (Horwitz and Wakefield 2007, 38). That’s why Wakefield thinks that “the destiny of the professions of mental health in regard to theoretical and scientific process in the comprehension of the etiology, the diagnostic and the treatment of mental disorder might depend in a large part from the progress in evolutionary psychology” (2005, 900). As I will show, reply (3) is not without problems of its own.

Evolutionary psychology’s central commitment, which allows the use of both of its paradigmatic methods (i.e., adaptive thinking and reverse engineering¹⁷), is the existence of a strong relationship between biological form and adaptive forces. Without such a relationship, there would be no reason to expect that isolating adaptive problems would be of any help in discovering the architecture of the mind or that starting with known mechanisms will lead to the reliable discovery of adaptive pressures that have acted on them in the past. But, as Griffiths (1996) observed (this is what he called the “historical turn in the study of adaptation”), “adaptive generalizations... cannot explain form except in conjunction with a rich set of historical initial conditions” (515). According to Griffiths, the reason why historical initial conditions are important is that they act as constraints on the two aforementioned modes of reasoning. For instance, adaptive problems depend on the biological features of the organism (e.g., in what kind of ecological niche ancestors of that creature were living), features of the environment of evolution adaptation, and the variations available for natural selection. To put it differently, knowledge of initial historical conditions is crucial to the identification of adaptation, given that an adaptation is relative to

- (1) traits or mechanisms that were present at the moment of selection,
- (2) a particular selective regime at the time (the selective pressures that existed at the time).

So in order to establish that something is an adaptation, one needs information about at least three things: (1) the traits that were present at the same time as the moment of selection, (2) the traits possessed by immediate ancestors of the bearer of the studied trait, and (3) the particular selective regime under which selection has taken place. Because of this, Griffiths suggests (1996; also Richardson 2007) that adaptations are best identified using the comparative method, which consists of comparing a trait to those of phylogenetic ancestors and to prevalent environmental conditions. To give just one example of the kind of surprise one can get from using this method, take the case of the descended larynx that has been hypothesized (using the reverse engineering method; Lieberman 1998) to be a uniquely human adaptation to speech. Fitch and Reby (2001) have shown that many other species have a descended larynx, including some deer species (*Cervus elaphus* and *Dama dama*), but also roaring cats and some bird species. According to them, “these comparative data suggest that vocal-tract elongation

is a relatively widespread response to ubiquitous constraints imposed by basic physics intersecting with the physiology of vertebrate vocal production,” and they go on to conclude that “it suggests that a descent of the larynx serving simply to exaggerate size could have pre-dated, and perhaps served as a preadaptation for, speech- or language-specific functions of the descended larynx” (173–174). Another surprising example concerns human hands. It has long been thought that the particular shape of human hands (the fact that we have a relatively long thumb and shorter fingers than other primates) was a typically human adaptation. But recent results (Almécija et al. 2015) suggest that the shape of human hands might not be an adaptation to tool use as has long been thought but rather that the shape is primitive rather than derived—that is, that the common ancestor of chimps and humans had hands more like humans than chimps and that it is the chimps’ hand with its elongated finger that is a derived trait (an adaptation). These two cases show why the use of history is necessary both to go beyond intuition and to constrain adaptationist explanations.

In principle, the historization of adaptationist explanations should not bother Wakefield greatly; after all, one cannot be against virtue. But the problem is the following: sometimes we have the information needed to establish that a human trait is an adaptation. For instance, we have access to traits that were present in our nonhuman primate ancestors or to traits that vary according to certain features of the evolutionary environment of adaptation (as in the case of malaria resistance, AIDS resistance, skin color, or lactose tolerance). But for many specifically universal human adaptations, the evidence necessary to establish that a trait is an adaptation does not exist. As Kaplan put it, “Such evidence is rarely available in the case of purported ‘universal’ human psychological adaptations. The very limited information available on the environments which key aspects of human evolution took place makes optimization techniques difficult to apply here. Further, while in some cases phylogenetic information about *Hominidea* may provide evidence relevant to adaptive hypotheses in humans, nature and history have ‘conspired’ to make the task more difficult with humans than it is in many other species” (2002, 297; for a similar conclusion, see Alden-Smith 2007, 253–254 and Thornhill 2007, 32). What Kaplan means by the latter is that our closest living relatives are the great apes (and we are not that close to them, as we diverged from a common ancestor about 6 million years ago). Because there is no other hominid alive, we cannot compare the fitness consequences of a trait that would have appeared somewhere in hominids’ evolution. Was that trait giving an adaptive advantage over others who did not have it (or did have it to a lesser degree)? We cannot answer this question for many universal human traits (like language). Therefore, it seems to be impossible to establish the designed function of these putative adaptive traits.

What this means for evolutionary psychiatry is that it is possible to establish evolutionary functional criteria for some mechanisms that evolved before the *Homo* genus. For instance, if Price and colleagues (1994) are right about depression, that is, if it is the

result of an adaptation to life in social groups (a form of de-escalation strategy) that we inherited from group-living ancestors common to primates, we should be able to use the comparative method and establish that the mechanism is indeed an adaptation. Similarly, in principle, it should be possible to establish the adaptive character of traits that vary inside the human population too. But for some mechanisms (the number of which has to be empirically determined but might include language, reasoning, inhibitory control, some forms of learning, imagination, and many others), we might just never know the facts necessary to establish that they are adaptations (for similar claims, see Richardson 2007, 38). Therefore, judgments about their dysfunction will be based on hunches about what is normal or abnormal, and as we saw, the past demonstrates that hunches such as these are unreliable, as they are especially open to the influence of values and norms.

What is there to conclude from this? Surely not that the epistemological objection cannot be met for some of the mental mechanisms that populate our mind. For instance, Stephen Downes (2009) recently observed that there is no reason to think that our species-typical mechanisms (if we understand “species-typical” mechanisms as meaning the mechanisms that are possessed by members of our species and not those that are exclusive to our species) were all produced in the Pleistocene. Some might predate that period, and some might be more recent. For both groups of mechanism, there is hope that we could find the facts necessary to establish with certainty that they are adaptations.¹⁸ But as Kaplan and others have observed, for the mechanisms that are supposed to have appeared in the Pleistocene, the task might be forever beyond our reach. If such is the case, the picture of the normal mind on which psychiatry is supposed to lean will be forever incomplete.

Conclusion

I have shown that HD analysis could take advantage of X-Phi methods. It could turn out that HDA is not what laypeople or professionals have in mind. If this is the case, Wakefield might want to argue that it is what they should have in mind, and thus in that case, his argument would be a normative argument. It could also turn out that people’s concepts of disorder vary as a function of different cues. In that case, it would be important to identify these cues in order to understand and prevent misunderstandings caused by these different concepts. Finally, Wakefield could be right on with his analysis. In that case, using X-Phi will only strengthen his position.

I also proposed a version of the epistemic objection. As I demonstrated, it applies only for a subclass of mental capacities, those that are human specific and universal. For these, the prospects of establishing that they are adaptations are slim. If such is the case, and depending on the number of mental capacities belonging to that subclass, some of the attributions of mental disorder might be desperately speculative. The

epistemic objection can thus become an objection to HDA. Not that it is not a good description of the concept that we are using, but given that it will leave a (large?) part of our mental attribution subject to value judgments or norms, it cannot play a normative role.

Notes

1. As Wakefield writes elsewhere, “A manual will be coherent and conceptually valid (i.e., valid in discriminating disorder from non disorder) only if its construction is guided by an adequate definition of disorder. In addition to determining which conditions are identified as disorders, such definition provides a framework for constructing diagnostic criteria for specific disorders” (1993, 160).

2. Recently, Wakefield has defined the “mental” in “mental disorder” a bit more precisely: “Mental dysfunctions are not specific mental states but rather dysfunctions in the brain mechanisms designed to produce or regulate mental states, and the dysfunction emerges in irregularities in the production and the regulation of mental states” (2007b, 127). This definition has important implications for psychiatry that I am not sure Wakefield endorsed. For instance, one can suppose that one psychological function of vision is to produce “mental representations” of the visual scene. If such is the case, then blindness should be considered a mental dysfunction, which is not the case. To which extent would Wakefield would want to reform psychiatry based on this definition is an unanswered question (see Murphy 2006 for a reformative position on the matter).

3. This surely does not mean that it is without problems (see Cooper, this volume; De Block and Sholl, this volume). For instance, it is not clear for whom the dysfunction has to be harmful for it to be judged a full-fledged disorder. Does it have to be harmful for an individual, their genes, family, or society in general?

4. As Wakefield puts it, “Mechanisms are naturally selected because they confer greater fitness on the organism *in a particular range of environments*. It is not the sheer number of environments in which there is harm, *but whether there is harm in the kinds of environments for which the organism was designed to operate harm free, that determines whether there is a dysfunction*” (1993, 166, my emphasis).

5. Neander (1995) rightly notes that this notion of normativity that is associated with evolutionary function or normal function is not evaluative. She writes, “Teenage fertility is biologically normal, but it does not follow that teenage fertility is a good thing; on the contrary, if we could induce (temporary and reversible) infertility in all girls under the age of twenty, that would probably be better [Boorse, 1975]. *Judging that something is functioning properly is not the same as judging that its functioning is good*” (111, my emphasis).

6. Wakefield makes such claims throughout his papers. For instance, he recently wrote that “every claim about a concept can be considered *an implicit empirical claim* about how individuals in some linguistic community use a term, so it is possible to empirically study whether the proposed conceptual analysis accurately portrays the community’s linguistic practices” (2007c, 41, my emphasis).

7. Poland (2003) makes similar criticisms to Wakefield's method: "there is good reason to be suspicious of just exactly how reliable such intuitions are, what they do and do not tap, and hence just how indicative they are of the actual employment of the concept of mental disorder in mental health practice. Regarding reliability, it is an important area of inquiry to determine just what kinds of factors influence and shape intuitive judgments, and what kinds of factors can support or undermine consistency of intuitive judgment across individuals or context. ... there are serious concerns that consultation of linguistic intuitions is far too impoverished an evidential basis for inferring the conceptual commitments of mental practitioners. Mental health practice takes place in a variety of very different contexts, and it concerns more than just talk. ... *Information about the actual diagnostic decisions that are made in clinical contexts, the ways in which the concept of mental disorder is reflected in relevant scientific research, as well as the role of the concept in other practical contexts (e.g., legal, public policy) is very likely a far more accurate and probative measure of conceptual commitments of practitioners than their intuitive response to hypothetical cases*" (34, my emphasis). I will return later to the latter idea in this chapter.

8. Nichols and colleagues take the results of this experiment and others as showing that some epistemic intuitions are not universal, as they vary with culture, socioeconomic status, and educational background. They take this as motivating a form of skepticism concerning the possibility of providing a conceptual analysis that would be accepted in every culture.

9. Some other items were also tested, for instance, the need for professional assistance or the possible duration of the condition, but for the sake of space, I will mention the results only briefly. In short, clinicians think that individuals meeting *DSM-IV* criteria for antisocial personality disorder while not being disordered should still be treated by a professional and believe that their problems could continue into their adult life (see Wakefield et al. 2002).

10. A difference that might show up in the understanding of mental disorders.

11. Studies from Pottick and colleagues (2007) suggest that certain features of the situation (occupation of the professional, race or ethnicity of the patient, etc.) might affect judgments of mental disorder. It seems that Wakefield considers these effects as performance distortions (following Chomsky's distinction, someone can be a competent user of a concept, even if they err in applying it to a situation or a thing, due to fatigue, for instance) or difference due to various particular theories about the same concept. An alternative account could be that people have (sometimes slightly) different concepts of mental disorders.

12. Wakefield recognizes this: for instance, he writes that "an evolutionary approach to natural versus disordered anxiety can offer a conceptual basis to help restrain such excesses [which consist in pathologizing natural emotions]" (Horwitz and Wakefield 2012, 19).

13. In his reply to McNally, Wakefield (2001) is making an additional point. As he says, "Often the evolutionary conclusions are themselves so speculative and unreliable that they can distort rather than solidify the evidential process. So, as an epistemological necessity, of course inquiry into function and dysfunction will continue to rely mainly on the study of current causal relationships—with the critical proviso...that current causal relationships are often taken as proxies for past design" (350). I am not convinced at all by this line of argument. Take your

standard paper in cognitive science; for instance, Munakata et al. (2011) describe two different pathways that are responsible for the inhibition of thoughts, actions, or emotions. The phenomenon of interest is cognitive inhibition, and what researchers want to know is how inhibition is produced. Researchers are making no assumptions concerning the evolutionary origin of the phenomenon nor would they think that the phenomenon is less in need of explanation if it was learned that inhibition is a spandrel or is not the result of natural selection.

14. Wakefield writes that, as it is clearly shown from historical and anthropological accounts of psychopathology, “values, norms and ideologies deeply influence what people take to be natural functions, in particular when scientific understanding of what is functional and dysfunctional is lacking (as it is the case for numerous aspects of the mental life)” (2006, my translation).

15. There is still a debate about this question, but the fact that the debate even exists proves that it is far from obvious that the dinosaur had wings designed for flying.

16. On the basis of an analysis of claw curvature, Naish (2011) remarks that different *Archaeopteryx* species might have a different behavioral lifestyle.

17. *Adaptive thinking* starts with a consideration of the adaptive problems (say choosing a place to live) that a creature has to solve in its evolutionary environment of adaptation to infer or to hypothesize the presence of mechanisms that are designed to solve them (e.g., aesthetics preferences for certain types of habitat; Orians and Heerwagen 1992). *Reverse engineering* starts with the presence of a mechanism (e.g., color vision) and tries to infer the problem it is designed to solve (e.g., making fruit more perceptually salient in a darker foliage background; Mollon 1996). In other words, “reverse engineering infers the adaptive problem from the solution which was adopted. Adaptive thinking infers the solution from the adaptive problem” (Griffiths 1996, 514).

18. This is if the human cultural environment does not profoundly modify the functioning of these older structures (Buller 2009, 79). For instance, work by Dehaene and colleagues (2010) shows that learning to read profoundly modifies the organization of the cortex. According to these researchers, “during education, reading processes must invade and ‘recycle’ cortical space devoted to evolutionary older functions, opening the possibility that these functions suffer as reading expertise sets in” (1359).

References

Alden-Smith, E. 2007. Reconstructing the evolution of the human mind. In *The Evolution of the Mind: Fundamental Questions and Controversies*, S. W. Gangestad and J. A. Simpson (eds.), 53–59. Guilford.

Almécija, S., J. B. Smaers, and W. L. Jungers. 2015. The evolution of human and ape hand proportions. *Nature Communications* (6): 7717.

Bergner, R. M. 1997. What is psychopathology? And so what? *Clinical Psychology: Science and Practice* 4(3): 235–248.

Boorse, C. 1975. On the distinction between disease and illness. *Philosophy and Public Affairs* 5(1): 49–68.

- Buller, D. 2009. Four fallacies of pop evolutionary psychology. *Scientific American*, January, 74–81.
- Buss, D. 2006. *The Murder Next Door: Why the Mind Is Designed to Kill*. Penguin Books.
- Colombo, A., G. Bendelow, B. Fulford, and S. Williams. 2003. Evaluating the influence of implicit models of mental disorder on processes of shared decision making within community-based multi-disciplinary teams. *Social Science and Medicine* 56(7): 1557–1570.
- Cosmides, L., and J. Tooby. 1994. Beyond intuition and instinct blindness: Toward an evolutionary rigorous cognitive science. *Cognition* 50(1–3): 41–77.
- Cullen, S. 2010. Survey-driven romanticism. *Review of Philosophy and Psychology* 1(2): 275–296.
- De Block, A. 2008. Why mental disorders are just mental dysfunctions (and nothing more): Some Darwinian arguments. *Studies in History and Philosophy of Science Part C* 39(3): 338–346.
- Dehaene, S., F. Pegado, L. W. Braga, P. Ventura, G. Nunes Filho, A. Jobert, G. Dehaene-Lambertz, R. Kolinski, J. Morais, and L. Cohen. 2010. How learning to read changes the cortical networks for vision and language. *Science* 330(6009): 1359–1364.
- Dennett, D. C. 1995. *Darwin's Dangerous Idea: Evolution and the Meaning of Life*. Simon & Schuster.
- Downes, S. M. 2009. The basic components of the human mind were not solidified during the Pleistocene Epoch. In *Contemporary Debates in Philosophy of Biology*, F. Ayala and R. Arp (eds.), 243–252. Blackwell.
- Fitch, W. T., and D. Reby. 2001. The descended larynx is not uniquely human. *Proceedings of the Royal Society B* 268: 1669–1675.
- Fulford, K. W. M., and A. Colombo. 2004. Six models of mental disorder: A study combining linguistic-analytic and empirical methods. *Philosophy, Psychiatry, and Psychology* 11(2): 129–144.
- Gelman, S. 2010. Modules, theories, or islands of expertise? Domain-specificity in socialization. *Child Development* 81(3): 715–719.
- Gould, S. J., and E. S. Vrba. 1982. Exaptation: A missing term in the science of form. *Paleobiology* 8(1): 4–15.
- Griffiths, P. E. 1996. The historical turn in the study of adaptation. *British Journal for the Philosophy of Science* 47(4): 511–532.
- Griffiths, P. E., and K. Stotz. 2008. Experimental philosophy of science. *Philosophy Compass* 3(3): 507–521.
- Harland, R., E. Antonova, G. S. Owen, M. Broome, S. Landau, Q. Deeley, and R. Murray. 2009. A study of psychiatrist's concepts of mental illness. *Psychological Medicine* 39(6): 967–976.
- Heinrich, J., S. J. Heine, and A. Norenzayan. 2010. The weirdest people in the world. *Behavioral and Brain Sciences* 33(2–3): 1–23.
- Horwitz, A. V., and J. C. Wakefield. 2007. *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford University Press.

- Horwitz, A. V., and J. C. Wakefield. 2012. *All We Have to Fear: Psychiatry's Transformation of Natural Anxieties into Mental Disorders*. Oxford University Press.
- Kaplan, J. M. 2002. Historical evidence and human adaptations. *Philosophy of Science* 69(53): S294–S304.
- Kauppinen, A. 2007. The rise and fall of experimental philosophy. *Philosophical Explorations* 10(2): 95–118.
- 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(1): 82–104.
- Knobe, J. 2007. Experimental philosophy. *Philosophy Compass* 2(1): 81–92.
- Knobe, J., and J. Doris. 2010. Responsibility. In *The Moral Psychology Handbook*, J. Doris and the Moral Psychology Research Group (dir. publ.), 321–354. Oxford University Press.
- Kupfer, D. J., M. B. First, and D. A. Regier, eds. 2002. *A Research Agenda for the DSM-V*. Washington Press Association.
- Lieberman, P. 1998. *Eve Spoke: Human Language and Human Evolution*. Norton.
- McCloskey, M. 1983. Intuitive physics. *Scientific American* 248(4): 122–130.
- McNally, R. J. 2001. On Wakefield's harmful dysfunction analysis of mental disorder. *Behaviour Research and Therapy* 39(3): 309–314.
- Millikan, R. 2002. Biofunctions: Two paradigms. In *Functions: New Essays in the Philosophy of Psychology and Biology*, A. Ariew, R. Cummins, and M. Perlman (eds.), 113–143. Clarendon.
- Mollon, J. D. 1996. The evolution of trichromacy: An essay to mark the bicentennial of Thomas Young's graduation in Göttingen. In *Brain and Evolution*, N. Elsner and H.-U. Schnitzler (eds.), 125–139. Springer.
- Munakata, Y., S. A. Herd, C. H. Chatham, B. E. Depue, M. T. Banich, and R. C. O'Reilly. 2011. A unified framework for inhibitory control. *Trends in Cognitive Sciences* 15(10): 453–459.
- Murphy, D. 2006. *Psychiatry in the Scientific Image*. MIT Press.
- Nahmias, E. 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research* 73(1): 28–53.
- Nahmias, E., S. G. Morris, T. Nadelhoffer, and J. Turner. 2005. Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology* 18(5): 561–584.
- Naish, D. 2011. [Review of] Glorified dinosaurs: The origin and evolution of birds. *Historical Biology* 23: 435–438.
- Neander, K. 1991. Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science* 58(2): 168–184.
- Neander, K. 1995. Misrepresenting and malfunctioning. *Philosophical Studies* 79(2): 109–141.

- Nichols, S., and J. Knobe. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs* 41(4): 663–685.
- Nichols, S., S. Stich, and J. M. Weinberg. 2003. Metaskepticism: Meditations in ethnoepistemology. In *The Sceptics*, S. Luper (ed.), 227–247. Ashgate.
- Nudds, R. L., and G. J. Dyke. 2010. Narrow primary feather rachises in *Confuciusornis* and *Archaeopteryx* suggest poor flight ability. *Science* 328(5980): 887–889.
- Orians, G. H. and J. H. Heerwagen. 1992. Evolved responses to landscapes. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. J. H. Barkow, L. Cosmides, and J. Tooby (eds.), 555–579. Oxford University Press.
- Panksepp, J. 2007. Can PLAY diminish ADHD and facilitate the construction of the social brain? *Journal of the Canadian Academia of Adolescent Psychiatry* 16(2): 57–66.
- Poland, J. 2003. *Whither Mental Disorder.* Unpublished manuscript.
- Pottick, K. J., S. A. Kirk, D. K. Hsieh, and X. Tian. 2007. Judging mental disorder: Effects of client, clinician, and contextual differences. *Journal of Consulting and Clinical Psychology* 75(1): 1–8.
- Price, J., L. Sloman, R. Gardner, P. Gilbert, and P. Rohde. 1994. The social competition hypothesis of depression. *British Journal of Psychiatry* 164(3): 309–315.
- Richardson, R. 2007. *Evolutionary Psychology as Maladapted Psychology*. MIT Press.
- Roe, K., and D. Murphy. 2011. Function, dysfunction, and adaptation? In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory*, P. R. Adriaens and A. De Block (eds.), 216–237. Oxford University Press.
- Sacks, O. 1997. *The Island of the Colorblind*. Vintage.
- Thornhill, R. 2007. Comprehensive knowledge of human evolutionary history requires both adaptationism and phylogenetics. In *The Evolution of the Mind: Fundamental Questions and Controversies*, S. W. Gangestad and J. A. Simpson (eds.), 31–37. Guilford.
- Thornhill, R., and C. T. Palmer. 2000. *A Natural History of Rape: Biological Bases of Sexual Coercion*. MIT Press.
- Wakefield, J. C. 1992a. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47(3): 373–388.
- Wakefield, J. C. 1992b. Disorder as harmful dysfunction: A conceptual critique of *DSM-III-R*'s definition of mental disorder. *Psychological Review* 99(2): 232–247.
- Wakefield, J. C. 1993. Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. *Journal of Abnormal Psychology* 102(1): 160–172.
- Wakefield, J. C. 1997. 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. 1999. Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology* 108(3): 374–399.
- Wakefield, J. C. 2000. Aristotle as sociobiologist: The ‘Function of a human being’ argument, black box essentialism, and the concept of mental Disorder. *Philosophy, Psychiatry, and Psychology* 7(1): 17–44.
- Wakefield, J. C. 2001. Evolutionary history versus current causal role in the definition of disorder: Reply to McNally. *Behavior Research and Therapy* 39(3): 347–366.
- Wakefield, J. C. 2005. Biological function and dysfunction. In *Handbook of Evolutionary Psychology*, D. Buss (ed.), 878–902. Oxford University Press.
- Wakefield, J. C. 2006. Fait et valeur dans le concept de trouble mental: le trouble en tant que dysfonction préjudiciable. *Philosophiques* 33(1): 37–64.
- Wakefield, J. C. 2007a. The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry* 6(3): 149–156.
- Wakefield, J. C. 2007b. What makes a mental disorder *mental*? *Philosophy, Psychiatry, and Psychology* 13(2): 123–131.
- Wakefield, J. C. 2007c. Why psychology needs conceptual analysts: Wachtel’s ‘discontents’ revisited. *Applied and Preventive Psychology* 12(1): 39–43.
- Wakefield, J. C., S. A. Kirk, K. J. Pottick, X. Tian, and D. K. Hsieh. 2006. The lay concept of conduct disorder: Do non-professional use syndromal symptoms or internal dysfunction to distinguish disorder from delinquency? *Canadian Journal of Psychiatry* 51(4): 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(3): 380–386.
- Woolfolk, R. L. 1999. Malfunction and mental illness. *The Monist* 82(4): 658–670.
- Woolfolk, R. L. 2013. Experimental philosophy: A methodological critique. *Metaphilosophy* 44(1–2): 79–87.