Defining Mental Disorder

Jerome Wakefield and His Critics

The title-level DOI for this work is:

doi:10.7551/mitpress/9949.001.0001

16 The Developmental Plasticity Challenge to Wakefield's View

Justin Garson

Introduction

According to Jerome Wakefield's influential analysis of "disorder," part of what makes something a mental disorder is that it stems from an inner dysfunction. A trait is dysfunctional, in turn, when it cannot do what natural selection designed it for. Many of Wakefield's critics have raised the possibility that there could, in principle, be mental disorders that do not involve inner dysfunctions in this sense. Along these lines, I'm going to argue that some mental disorders might result from "developmental mismatches." This takes place when the environment that the fetus or child encounters is very different from its adult environment, and the kinds of strategies (physical traits, behaviors, or psychological dispositions) that the fetus or child used to master the early environment are maladaptive in the later environment. I argue that this would be a case of disorder without dysfunction, and I give some empirically plausible examples. To begin, however, I will discuss a concrete example of a developmental mismatch in the biological world.

The tiny crustacean of the genus *Daphnia* provides a remarkable example of developmental plasticity. One species, *Daphnia cucullata*, is about 3 millimeters in length and inhabits lakes across Europe. It has several invertebrate predators. If a *Daphnia* is raised in the vicinity of predators, it grows a tough, helmet-shaped head. This "helmet" is a boon as it makes it difficult for predators to swallow it. The helmet, I take it, is an adaptation designed by natural selection to protect the *Daphnia* in perilous waters. Other species of *Daphnia* have evolved equally impressive defenses, such as tail spines, "neckteeth," and crests (a kind of pointed head shape) (Tollrian and Dodson 1999). These are called "inducible" defenses as their appearance is triggered by the presence of kairomones, a kind of hormone released by the predator. They can also be epigenetically transmitted (Agrawal et al. 1999). If a female *Daphnia* is exposed to kairomones, her offspring are more likely to grow the helmet-shaped head even in the absence of predators.

There are some drawbacks to the "helmet" phenotype. First, helmets are metabolically expensive; they require more calories to maintain. Second, the large head reduces the *Daphnia*'s mobility. That is why natural selection gave the *Daphnia* a certain degree of morphological flexibility. It only grows the helmet if it needs to. Once a phenotype is selected—"helmet" versus "normal"—reversibility is limited. (Different defenses show different degrees of reversibility, from completely reversible to completely irreversible; the helmet phenotype is closer to the latter.) This is an example of developmental plasticity. Something inside the *Daphnia* encodes a conditional rule: "if predators, then helmet; if no predators, then no helmet."

Developmental plasticity is a subtype of phenotypic plasticity, and there are many different mechanisms for it (Pigliucci 2001). Imprinting is another such mechanism (see section I). At the most general level, developmental plasticity takes place when there are at least two distinct adult phenotypes the juvenile organism can grow into, and which phenotype it assumes depends on the contingencies of its formative context. When biologists talk about developmental plasticity, moreover, they typically imply that each adult phenotype represents an *adaptation* to that specific developmental context, that is, that the phenotype is "appropriate" to that context. Getting a permanent scar as a result of a playground accident is not an example of developmental plasticity.

Developmental plasticity has its risks. The big risk is a "developmental mismatch." Suppose we hatch some *Daphnia* eggs in a tank swarming with predators, and they grow the helmet-shaped head. Suppose we then remove the predators from the tank. Then the *Daphnia* experience only the disadvantages of the helmet phenotype and none of its perks. Their condition becomes *chronically maladaptive*. It would be acutely troublesome for them if we forced them to compete over limited resources with their "normal"-shaped counterparts, who need less food and can get to it faster.

Some biologists like to describe the risks inherent in developmental plasticity by using a gambling metaphor. The developing organism can be seen as making a kind of "prediction" about what its future world will be like, on the basis of its present conditions. In other words, it "samples" its present environment and extrapolates into the future. It then "selects" a phenotype that would maximize its fitness in that anticipated future environment. This is called a "predictive adaptive response" (Gluckman et al. 2009; Glover 2011). If its "prediction" turns out to be correct, it is rewarded with enhanced fitness; if it is incorrect, its fitness is reduced. The latter scenario is a "developmental mismatch."

Here is an intuition that I have. (Fortunately, I need not rely exclusively on intuition here, because I have a specific theory of function, one with strong independent credentials, that yields precisely this result. In the next section, I'll describe that theory.) It seems to me that talk of "dysfunction" is out of place when it comes to developmental mismatches. Let me clarify. Suppose there is a member of *Daphnia* that chose the "wrong" phenotype; that is, suppose it was raised in a tank with predators,

it grew the helmet-shaped head, and later, the predators were removed. It exhibits a developmental mismatch and takes a fitness loss as a result. In my opinion, this does not represent an inner "dysfunction." Put metaphorically, nothing "went wrong" inside that *Daphnia*. Its developmental machinery is operating exactly as it is "supposed to." It is neither defective nor diseased; it's just unlucky. (Of course, the mismatch can *cause* a dysfunction, for example, if the *Daphnia* dies of malnutrition. But that sort of dysfunction is incidental to the mismatch; having a mismatch does not logically imply dysfunction.)

This brings me to the central question of the chapter. What if some of our current psychiatric ailments result from developmental plasticity, *rather than* dysfunction? In other words, what if, in certain individuals with bona fide mental disorders, the disorder represents a developmental mismatch, much like a helmet-shaped *Daphnia* in a predator-free environment (see Garson 2015, chap. 8)? Some psychiatric researchers take this possibility quite seriously (for a recent review, see Glover 2011). For example, they argue that some anxiety disorders, such as generalized anxiety disorder, might arise from a contrast between one's formative environment and one's adult environment. The "formative" environment can include both the prenatal "environment," as well as the postnatal environment of the infant or young child. Such mismatches can be chronically maladaptive for the individuals that possess them. This is not a bit of philosophers' speculation but an empirically plausible conjecture, one that should be accepted, or refuted, on empirical grounds. I do not know whether this budding research program—sometimes known as Developmental Origins of Health and Disease (DOHaD)—will ultimately be vindicated (for an overview, see Gluckman and Hanson 2006). But I think it represents an exciting new avenue for exploring the roots of major mental disorders.

This conjecture—that *some mental disorders are developmental mismatches*—raises a significant problem for Wakefield's "harmful dysfunction" (HD) analysis of mental disorder, which holds that all mental disorders stem from biological dysfunctions (e.g., Wakefield 1992, 1999a, 2011). I will lay out the developmental plasticity challenge in three stages. First, I will set out the underlying theory of biological function and dysfunction that Wakefield and I accept. An important upshot of this discussion is that there is a distinction between saying that a trait is *dysfunctional* and saying that it is *functioning normally in an unsuitable environment* (or, perhaps better, that it is unable to function *due to* an unsuitable environment). I will also raise the thorny issue of "function indeterminacy" and explain its relevance to my argument. Second, I will describe a line of criticism that Wakefield's opponents have repeatedly raised. It is called the "evolved mismatch critique." I raise it here because the logical structure of that argument mirrors, in some important ways, the logical structure of my own. I happen to believe that the evolved mismatch critique, when properly understood, undermines Wakefield's analysis, but I will not lean too heavily on that argument here. In the third

part, I will explain the developmental plasticity challenge. I will also respond to potential objections against this challenge.

To give credit where credit is due, I should point out that two psychologists, John Richters and Stephen Hinshaw (1999, 442–443), raised a version of this "developmental plasticity challenge" against Wakefield's analysis. They ask us to consider a hypothetical example of a young boy who grows up in an abusive home and develops symptoms of conduct disorder as a result (e.g., aggressive behavior). Suppose that those symptoms were functional and adaptive in the abusive home and that he had those symptoms *because* they were useful to him. Suppose, finally, that when the boy moves to a new, nurturing home, he retains the symptoms of conduct disorder. The boy's "hostile world orientation" is now a fixed part of his young personality, much like the *Daphnia*'s helmet-shaped head. Richters and Hinshaw believe that, in this case, the boy has a mental disorder with no underlying dysfunction. In my terminology, this would be a "developmental mismatch." Richters and Hinshaw, however, raise the example in passing and not entirely persuasively (in section III, I will address Wakefield's response). Moreover, their paper makes numerous general claims about function and development that I do not accept. One way to think about my project is that it represents an attempt to illuminate, more clearly than they did, the logical structure of the argument underlying their example and to give a more plausible example to bolster their case.

I. What Are Functions? What Are Dysfunctions?

One of the perennial questions of the philosophy of psychiatry is, what are mental disorders? What do all of these diverse conditions, such as schizophrenia, bipolar disorder, and personality disorders, have in common—if anything—that makes them mental disorders? In Wakefield's view, disorders generally, whether mental or physical, are *harmful dysfunctions* (the "HD analysis"). This definition has two parts: *harm* and *dysfunction*. First, in order for something to be a disorder, it must be harmful as judged by prevailing social norms. Second, for something to be a disorder, it must stem from an inner dysfunction on the part of the individual. It can be a dysfunction on the part of the brain or nervous system, for example, or it can be a dysfunction on the part of the mind. The notion of *dysfunction* is what distinguishes disorders, per se, from socially deviant behavior (such as belonging to the American Nazi Party) and from psychological states that are merely unpleasant to have (such as stress about a job interview). The idea that the notion of *dysfunction* is somehow implicated at the core of mental disorders has been raised before (e.g., Spitzer et al. 1977; Klein 1978; American Psychiatric Association 1980, 6). Unfortunately, those authors did not provide a clear explanation of what "dysfunctions" were, so their attempts at definition left much to be desired.

What is it, then, for something to be a dysfunction? What are functions? Here, Wakefield relies on a certain conception of function that philosophers of science

developed in the 1970s and 1980s, the *selected effects* theory of function (see Wright 1973; Millikan 1984; Neander 1991). There are several nuances here when it comes to distinguishing various forms of the theory, but roughly, the theory holds that *the function of a biological trait is whatever it was selected for by natural selection*. In other words, the function of a trait is the reason it evolved by natural selection. The reason my nose has the *function* of helping me breathe, and not the *function* of holding up glasses, is that the former benefit explains why people have noses (via appeal to natural selection). Traits that did not evolve by natural selection, such as birthmarks or freckles, do not have functions at all, even if they happen to benefit us from time to time. I accept the selected effects theory, and in the remainder of this chapter, I will take it for granted, and I urge the reader to do the same, at least for the sake of argument. (Garson [2016] defends the theory from several common criticisms.)

Now that we know what functions are, it would seem that defining "dysfunction" would be fairly easy. But we quickly encounter conceptual obstacles. In the remainder of this section, I will note some of those obstacles and develop a definition of "dysfunction" that resolves them. At first pass, it would seem that *something is dysfunctional just when it cannot perform its function*. The eyes of a person who is congenitally blind are dysfunctional because they cannot perform their function of seeing.

There are some details that are not entirely resolved, but it is not necessary to resolve them here. For example, most traits can exhibit various rates of functioning, and it is difficult to "draw a line" between those instances when a trait functions at a low but intuitively "acceptable" rate and those instances when it is dysfunctional (Wakefield 1999a, 379; also see Schwartz 2007; Garson and Piccinini 2014). Some of Wakefield's critics have raised this "line drawing problem" as a significant objection (e.g., Lilienfeld and Marino 1995, 414), but I do not consider it a deep problem. Rather, it reflects the standard sort of vagueness that most philosophical concepts possess. In my usage, to say a trait is "dysfunctional" implies either that it cannot perform its function at all (e.g., cardiac arrest) or that it can only perform its function at an unacceptably low rate. A trait is "nonfunctional" when it does not have a function (birthmarks or freckles).

Here is a more serious problem with our simplistic definition of dysfunction. Just because a trait cannot perform its function, that alone does not make it dysfunctional. In other words, the inability to perform a function (at an acceptable rate) is necessary, but not sufficient, for dysfunction. A simple example will prove the point. Suppose I am blindfolded. My eyes cannot perform their function of seeing. They are not, however, dysfunctional. So dysfunction is not just the inability to perform a function but something more. To say that something is dysfunctional implies that it cannot perform its function for intrinsic or constitutional reasons and not just because of an unsuitable environment (Dretske 1986; Neander 1995; Wakefield 1999a, 385). There are various ways we can describe the blindfold situation. One way is to say that my eyes are "functioning normally in an unsuitable environment." Another way is to say that my eyes

are "unable to function *because* of an unsuitable environment." It does not really matter which way we describe the case. The important point is that it is not a dysfunction.

These considerations suggest that we should amend our definition to say that *something is dysfunctional just when it cannot perform its function, for "inner" or "constitutional" reasons, rather than because it's in an unsuitable environment*. But we are not quite out of the woods yet. There is a certain conceptual puzzle that has plagued the selected effects theory for decades and creates new trouble for defining "dysfunction." (In fact, it plagues most theories of function that tie function to evolution.) It is called the problem of function indeterminacy. (There are various forms of indeterminacy; here, I will describe the "hierarchical" form of indeterminacy.) The problem of indeterminacy, at the most general level, is that there are many ways of describing a trait's function, all of which are allowed by the selected effects theory. Is the function of the heart simply to *beat*? To *circulate blood*? To *bring nutrients to cells*? To *keep the organism alive*? All of these descriptions are acceptable because they are all effects that explain why the heart evolved by natural selection. (The issue is more complex, because the heart has other functions than merely beating; for example, it also regulates its rate of pumping in order to maintain a stable ratio of carbon dioxide to oxygen. I will set aside such details at present and simply focus on the heart's beating.)

Fortunately, in most cases, it does not matter how we describe the heart's function. In some cases, however, it matters quite a bit, particularly when we are trying to specify, in a rigorous and precise way, when something dysfunctions. That is because different ways of describing a trait's function may be more or less suggestive of "dysfunction." Suppose somebody's heart is beating at a relatively normal rate, but that person has a ruptured artery in his or her brain that prevents blood from circulating effectively. Is that person's *heart* dysfunctional? If we say that the function of the heart is simply to *beat* (i.e., engage in systole and diastole), then it is *not* dysfunctional. It is doing exactly what it is "supposed to" do. If we say that the function of the heart is to *circulate blood*, it *might* be dysfunctional (since, after all, it cannot do what it is "supposed to" do). It would depend on the details of our definition of dysfunction.

Different theorists have proposed different solutions to the problem, and I will not attempt to survey them all. I endorse a simple and plausible solution developed by Neander (1995), who also explores conceptual nuances that I do not have the space to explore here. Her solution stems from the following observation. In our example, the different descriptions of the heart form a certain series, that is, a hierarchy defined by cause and effect. By beating, the heart circulates blood. By circulating blood, the heart brings nutrients to cells. By bringing nutrients to cells, the heart keeps us alive. When we say that the function of the heart is simply *to beat*, we are describing the most "proximal" member of that chain. When we say that the function of the heart is to *keep us alive*, we are describing its most "distal" member. When we say that the function of

the heart is to *bring nutrients to cells*, we are describing an "intermediate" member of that chain, somewhere between the most proximal and the most distal.

Her view is that a trait is dysfunctional only when it cannot perform its most proximal function. The heart is dysfunctional only when it cannot *beat*. There are two good reasons for accepting her solution. The first is that it is intuitively appealing. I think it gives the "correct" verdict in the example of the heart. The second is that it is, from a biomedical perspective, the most sensible solution. In the biomedical context, when we say that something is dysfunctional, we are indicating, in a pragmatic way, that it is an appropriate target of medical intervention (Buller 1997). But presumably, if the heart cannot circulate blood because of a ruptured brain artery, we should target the artery and not the heart! To say that the heart is dysfunctional seems contrary to good medical practice. Wakefield (1999a, 386) also describes the problem of indeterminacy (see his bacterium example), and he seems to accept the same solution.

My viewpoint about function indeterminacy informs my intuition about the *Daphnia* case. Let's assume, for the sake of simplicity, that there is a mechanism (*M*) in the *Daphnia* that obeys the following rule: "if predators, then helmet; if no predators, then no helmet." There are many ways of describing *M*'s function. We could say that *M*'s function is to trigger a certain developmental sequence (one that yields the helmet head) in response to kairomones. Perhaps it does this by releasing a certain hormone into the bloodstream. Then the function of *M* is to *release hormone H in response to kairomones*. Alternatively, we could say that *M*'s function is to *protect the individual from predators*. The former is a more proximal way of describing its function and the latter a more distal way of describing its function. That is because the two descriptions form a series: the mechanism typically *protects the individual from predators* by *releasing hormone H in response to kairomones*.

Now, suppose we have a developmental mismatch. Is there any dysfunction? Described in the most proximal way, the answer is no. *M* fully and adequately discharged its function when it released hormone *H* in response to kairomones. Described in a more "distal" way, there could be a dysfunction. After all, if there are no predators around, *M* certainly cannot perform its function of protecting the individual from them. So, whether or not it is "dysfunctional" depends, in part, on how we describe it. I prefer the more proximal description for the reasons I gave above.

My discussion may seem to belabor the point, but there are cases where I think Wakefield is potentially inconsistent in his approach to function indeterminacy, and the way we describe the *Daphnia*-type case is pivotal to my argument. On one hand, Wakefield's (1999a, 386) explicit comments about indeterminacy seem to agree with my own, namely, that we should prefer the most "proximal" description of an item's function (as in the bacterium case). On the other hand, some of his specific examples seem to run contrary to that point. He discusses an example of filial imprinting gone

awry, that is, where a gosling imprints on a passing porcupine (Wakefield 1999b, 468; 2000, 263). (Imprinting refers to a developmental "window" of time in which a juvenile organism forms a strong, lifelong preference. The function of filial imprinting in goslings is to cause them to form an attachment to their own mothers. The mechanism by which this works is that they form an attachment to the first large, suitably moving object that they encounter. Imprinting goes awry when the mechanism causes a gosling to imprint on an object that is not its mother.) The gosling now has an enduring inner disposition to follow around a porcupine. Wakefield says that this disposition is a dysfunction. I do not consider it a dysfunction (Murphy and Woolfolk [2000b, 279] have similar reservations about Wakefield's imprinting case). I think there would be a dysfunction if the gosling failed to imprint on a passing porcupine, so long as that porcupine moved about in the right sort of way and if that porcupine entered the gosling's visual field at just the right time. The gosling's disposition would be chronically maladaptive but not dysfunctional.

I suspect that the difference of opinion between Wakefield and myself traces back to the problem of function indeterminacy. For there are two ways of describing the function of the imprinting mechanism in the gosling's brain, and one is more proximal than the next. The first, most proximal description is to say *M*'s function is to cause the gosling to form a strong attachment to the first large, suitably moving object it sees. The second, more distal description is to say *M*'s function is to cause the gosling to have a disposition to follow its mother. The first is more proximal than the second because the mechanism typically causes the gosling to follow its mother *by* causing the gosling to form a strong attachment to the first large, suitably moving object it sees. If we stick with the first description in the porcupine case, we see there is no dysfunction. *M* has performed its job admirably. If we stick to the second description, we have some evidence of dysfunction (after all, *M* cannot discharge its function). I have given reasons for my preference for the more proximal description. I will come back to this issue in section III.

To summarize this rather abstract discussion: the function of a trait is the reason it evolved by natural selection, and *a trait is dysfunctional just when it cannot carry out its most proximal function, for constitutional reasons*.

II. The Evolved Mismatch Criticism of the HD Analysis

In this section, I present one long-standing objection against Wakefield's view, the "evolved mismatch" criticism. I raise it here because it forms a backdrop to my own argument (next section), and it has a very similar logical structure. What if some of our devastating psychiatric ailments, such as major depression, anxiety disorders, psychopathy, and so on, actually *benefited* our Pleistocene ancestors? What if, moreover, the *fact* that they benefited those ancestors partly explains why they are around today? Then, if we accept the selected effects theory of function, we would have to say that those

disorders do not arise from "dysfunctions." They would be adaptations. Furthermore, if we accept the HD analysis, we would be forced to conclude that depression (say) is not actually a mental disorder. That strikes me as deeply counterintuitive. It seems to me that depression, particularly when severe enough to lead to hospitalization or suicide attempts, constitutes a paradigmatic mental disorder, regardless of how it happened to evolve. People who have raised this mismatch critique against Wakefield include Lilienfeld and Marino (1995, 416; 1999, 406), Richters and Hinshaw (1999, 442), Woolfolk (1999, 662), Bolton (2001, 194), Murphy and Stich (2000, 81), and Murphy and Woolfolk (2000a, 244).

Let me clarify what I take to be the strongest form of the evolved mismatch argument. I am *not* claiming that any particular mismatch hypothesis is true. Rather, the best argument is a modal one, and I will summarize it in four sentences: it is empirically plausible that some mental disorders represent mismatches, not dysfunctions. Therefore, it is logically possible that the same is true. But the HD analysis implies that this claim is logically impossible. So, the HD analysis is wrong.

Here is another way of putting the point, one that Wakefield sometimes opts for. To the extent that the HD analysis is a conceptual analysis of clinical usage, then Wakefield is committed to the following prediction: *if* researchers and clinicians were to generally accept that a certain condition (say, antisocial personality disorder) is an evolved mismatch, *then* they would stop labeling it a "disorder." Wakefield uses the example of fever to bolster his point. Medical researchers once considered fever to be a "disorder"; when they came to grasp its adaptive significance, they stopped calling it that (Wakefield 2000, 260; also see Wakefield 1999b, 468). I am not entirely convinced by this example, since what we discovered about fever is that it is beneficial for us. It is not a mismatch at all. So, I don't think we can use the fever example to draw inferences about an evolved mismatch case. Moreover, as I indicated above, it would be surprising to me if Wakefield's prediction were correct, because several researchers have endorsed mismatch hypotheses for various disorders, and they seem to believe, judging by their terminology, that the conditions they study are, in fact, "disorders" (or "pathologies," "diseases," etc.) (e.g., McGuire and Troisi 1998; Gluckman and Hanson 2006; Glover 2011). Nonetheless, I applaud Wakefield for being willing to make a risky prediction, and I wish more philosophers would do the same.

I will give a simple example to convey the style of an "evolved mismatch" explanation. One theory of depression is known as the "social competition" hypothesis (Price et al. 1994). In the Pleistocene era, when many of our cognitive mechanisms were being formed, there were numerous male-to-male conflicts over food, shelter, and sexual partners. Occasionally, one of the "disputants" must have been severely outcompeted by the other. Now, zoom in on a particular such conflict. Suppose the "underdog" had some gene mutation that caused him to feel depressed, rather than aggressive. Suppose his depressed mood made him bow out of the fight and to accept a lower status within the

social hierarchy. Then, the depressed feelings would have conferred a fitness advantage (over individuals with no mutation), since it would have prevented him from getting killed or wounded in a pointless fight. The proponents of the social competition hypothesis believe that it explains various cognitive, behavioral, and neurochemical features of depression and that it has important implications for therapy.

Mismatch explanations are a subtype of adaptationist explanation. As such, they can, in the best-case scenario, be rejected or revised as fresh evidence surfaces. For example, Murphy (2005, 756) notes that the social competition hypothesis fails to explain a range of somatic symptoms associated with depression, but he acknowledges that other mismatch hypotheses may be superior. For example, some mismatch theorists have argued that depression does not result from competition, but it has a kind of signaling function to elicit help from parents or partners (Watson and Andrews 2002). Others have tried to synthesize the competition and signaling theories (see the "social risk" hypothesis of Allen and Badcock 2003; for a recent overview, also see Rottenberg 2014). (Of course, if there are different types of depression, then one would not expect a single mismatch theory to cover them all.) Others have criticized mismatch hypotheses for phobias (Murphy 2005; Faucher and Blanchette 2011). My point is that, even if one or another specific mismatch hypothesis fails, that does not undermine the credibility or coherence of mismatch hypotheses generally. By the same token, the failure of one or another adaptationist hypothesis does not undermine adaptationism as a research program. If anything, I think the general research program is gaining new momentum, particularly in light of the DOHaD.

Note that it is one thing to be critical of the way that clinicians overdiagnose depression, and it is another thing to deny that depression is a disorder. Horwitz and Wakefield (2007) eloquently argued that clinicians overdiagnose depression. But in saying this, they acknowledge that there is a genuine mental disorder (which they sometimes call "depressive disorder") and that, sometimes, psychiatrists correctly identify it. The claim I am making is this: what if, say, what they call "depressive disorder" turns out to be an evolved mismatch? Then, Wakefield would have to acknowledge that it is not a mental disorder, which is contrary to the view that he and Horwitz staked out in that book.

In the next section, I will explore a variant on the evolved mismatch argument, one that has been neglected in the literature surrounding the HD analysis. I call it the developmental plasticity challenge, and I think it raises additional problems for Wakefield's view.

III. Developmental Mismatches and Dysfunctional Mechanisms

Suppose there were some condition (one that psychiatrists study) that proved to be a developmental mismatch, rather than an evolved mismatch. That is, suppose it were the outcome of developmental plasticity, but the environment changed, like the *Daphnia*'s

helmet-shaped head in predator-free waters. Then, according to the HD analysis, it would not stem from a "dysfunction," and therefore, it would not be a mental disorder. I think there are empirically plausible cases of such developmental mismatches, and they constitute a problem for Wakefield's view.

Perhaps the most plausible example of a developmental mismatch is for the anxiety disorders, such as generalized anxiety disorder. When we talk about anxiety disorders, we are not talking about transient states of disturbance that accompany common stressors such as a job loss or a move. Those emotions, I take it, are normal, nondisordered responses to the vicissitudes of life. What I am talking about are more or less chronic, maladaptive conditions that seem disproportionate to any external "triggers." Generalized anxiety disorder is described as a chronic and uncontrollable state of worry that is disproportionate to external stressors and that can easily shift from one concern to another. It can be associated with restlessness, fatigue, and the inability to concentrate. Panic disorder is the condition of having recurring panic attacks. These are typically short-lived but extremely intense episodes of distress and discomfort (American Psychiatric Association 2013). If anything has a right to be called a mental disorder, these do.

Is it plausible to think that anxiety disorders, at least in some individuals, represent developmental mismatches—more or less irreversible adaptations to prenatal or early postnatal experiences? There is some suggestive evidence for the theory. The empirical basis for the theory is that fetuses, infants, and young children who are exposed to highly stressful environments are susceptible to anxiety disorders as adults (e.g., Heim and Nemeroff 2001; McGowan et al. 2009; Glover 2011). Perhaps that should not seem very surprising, but it leads to an intriguing conjecture: what if susceptibility to anxiety disorders is an example of developmental plasticity? That is, what if susceptibility represents a kind of adaptation to a high-stress developmental context? Then, those disorders would not, in fact, represent dysfunctions. They would have the exact same ontological status as the helmet-shaped head of the *Daphnia*. They would be adaptations. If that were the case, then, according to the HD analysis, they would not be disorders. That seems wrong.

One might wonder how having an anxiety disorder, like generalized anxiety disorder, could be beneficial or useful. In what context might those help us survive? The specific adaptationist hypothesis on offer is that anxiety is associated with enhanced vigilance to potential threats in one's environment (Glover 2011). People who are anxious tend to watch out for things that might hurt them. As a consequence, if there are real, genuine threats in one's formative environment, then it could very well pay to be anxious as an adult, in a kind of chronic, intense, way, not in a run-of-the-mill way. (I presume that Wakefield would agree that some level of anxiety is adaptive, but that extreme levels represent a dysfunction of those anxiety-generating mechanisms. But my point is that it is empirically plausible that "extreme" levels of anxiety could represent adaptations, too. That is the possibility I wish to explore here.) Similar sorts of

arguments have been offered for other disorders such as attention-deficit/hyperactivity disorder and conduct disorder.

As a matter of logic, there are two ways that Wakefield could defend his theory, on the assumption that some mental disorders represent developmental mismatches. First, he could say that there is no dysfunction in those individuals, but there is no disorder either. Second, he could accept that, of those anxiety disorders that result from developmental plasticity, they are genuine mental disorders, but there *is* an underlying dysfunction. I will briefly summarize why both responses strike me as unsatisfying.

Wakefield (1999b, 468) explores both lines of argument in response to Richters and Hinshaw's (1999) example. Recall that, in their argument, we are asked to imagine a young boy who grows up in an abusive family environment and responds by developing symptoms of conduct disorder. They ask us to suppose that those symptoms are, at least initially, adaptations and that they are useful and valuable to him in that abusive environment. Later, the boy moves to a nurturing, nonabusive environment, but the symptoms of conduct disorder do not abate. They are a "fixed" part of his character, like the *Daphnia*'s helmet-shaped head. They argue that this would be an example of a disorder without a dysfunction. Wakefield entertains two different responses: the first that there is no dysfunction and no disorder either, and the second that there is both a disorder and a dysfunction.

3.1 No Dysfunctions, No Disorders

It seems to me that Wakefield's first line of response, which would deny that anxiety disorders are actually mental disorders, strays perilously close to resolving the issue by definitional fiat or stipulation. At the very least, I would like to be given an independent reason for thinking that, if some anxiety disorders result from plasticity, then they are not real disorders at all. Without any good independent reason, I worry that the HD analysis is something like a stipulative definition, rather than a conceptual analysis or a theoretical definition (e.g., Millikan 1989), and Wakefield clearly does not intend the HD analysis to be a piece of stipulation. Moreover, as I indicated above, the few people who have endorsed the claim that some mental disorders are developmental mismatches seem to describe those conditions as "disorders," "pathologies," and so on (as I indicated in the previous section; e.g., Gluckman and Hanson 2006). So I do not think this first line of response is entirely promising.

One line of evidence that Wakefield might adduce to support this move is to say that clinical intuitions are, in fact, overwhelmingly on his side and that my own linguistic intuitions are idiosyncratic. Interestingly, Wakefield and his colleagues have conducted some experiments that suggest that clinical intuitions tend to be consistent with his HD analysis of mental disorder (Wakefield et al. 2002; Wakefield et al. 2006). Those experiments went as follows. A large number of graduate students in mental health were presented with a series of vignettes. In one version of the story (the "negative

environment" vignette), a teenager exhibits various symptoms of conduct disorder, but these symptoms are portrayed as having current usefulness (e.g., in response to current family abuse or gang activity). In another version of the story (the "internal dysfunction" vignette), a teenager exhibits the same symptoms, but nothing in the story would suggest that they are appropriate or useful responses to some current life situation. Graduate students tended to judge that the individual in the first vignette did *not* have a genuine mental disorder, but the individual in the second did. Wakefield interpreted this result to support his HD analysis, because there is clearly no "dysfunction" in the first.

I do not believe that these vignettes are relevant to this discussion, although they are fascinating in their own right. That is because they do not contrast "dysfunction" scenarios with "mismatch" scenarios. Rather, they contrast a scenario in which the condition has obvious current utility with one where it does not. Technically, a "mismatch" scenario would fall under the second type of vignette, where the trait in question does not have current utility. It would be interesting, however, to extend those sorts of experiments to investigate clinical intuitions regarding mismatch cases.

3.2 Both Dysfunctions and Disorders

The second way to respond is to say that, in the case I've described (where anxiety disorders result from developmental plasticity), the affected individuals do have mental disorders, but there *are* underlying dysfunctions. But where is the dysfunction? Wakefield entertains this response to the conduct disorder case. As he writes (1999b, 468), "If the mechanism's function is to shape personality specifically in response to the early broader environment (not the family environment, which evolutionarily is expected to be reasonably benign) to prepare for the later broader environment, then the 'accidental' setting of personality parameters by extreme (evolutionarily unexpectable) family abuse is a dysfunction." The idea is this. In the conduct disorder case, the child has a certain cognitive mechanism, *M*. The function of *M* is to sample the threat level in his formative environment and to shape his personality as a result, so as to prepare him for the types of encounters he might face in the future. Unfortunately, because of the abusive family environment, *M* is unable to adequately prepare him for his future environment. After all, the future environment (let us suppose) is relatively benign, which, of course, he couldn't have guessed from his abusive family life. So, *M* cannot fully discharge its function, and it is dysfunctional.

If this is the sort of response that Wakefield is committed to, then there is a substantive disagreement between Wakefield and myself regarding how to describe the conduct disorder case. Like the blindfold case or the *Daphnia* case, I would say that this is a situation where *M* is functioning normally in an unsuitable environment (or, perhaps, *M* is unable to function because of the unsuitable environment) and it is not dysfunctional. It seems to me that the function of this hypothetical cognitive mechanism is to "sample" the ambient level of threat in the formative environment and to

adapt the child's personality as a result. It seems to me that the mechanism in question has discharged its function flawlessly, just like the *Daphnia*'s helmet head or like the gosling's imprinting on a passing porcupine. (Keep in mind, of course, that as a result of the anxiety disorders, various dysfunctions may ensue from time to time. Someone who is prone to panic attacks might form the false belief that heavy exercise will promote fresh panic attacks and refrain from exercise on that account, thereby increasing his or her risk for cardiovascular disease. But that sort of dysfunction would be incidental to the disorder; it would not be constitutive of it. Additionally, note that I am not claiming that, e.g., conduct disorder is actually an adaptation but that the empirical plausibility of that conjecture suffices to undermine Wakefield's analysis.)

As I indicated in section I, this disagreement has its root, I think, in the problem of function indeterminacy, which I described earlier. There are always two different ways to describe the function of some cognitive mechanism. First, we can describe it in "proximal" terms. Here, in the case of conduct disorders, the mechanism has the function of (say) sampling the threat level in the early environment and calibrating the lifetime level of aggressiveness accordingly. Second, we can describe it in "distal" terms. Here, the mechanism has the function of protecting the child from future threats. How should we classify the developmental mismatches? If we describe the mechanism in terms of its proximal function, it seems to me that there is no dysfunction. If we describe the mechanism in terms of its distal function, there can be. I believe that Wakefield describes the conduct disorder case in terms of "dysfunction" because he has latched onto the more "distal" description of the function, and I think this is inconsistent with his explicit remarks on indeterminacy, where he states that we should prefer the most "proximal" description of an item's function (again, see Wakefield's [1999a, 386] bacterium example). As a consequence, I do not believe that this particular line of response is available to him.

In the foregoing, I have raised a fairly novel critique of Wakefield's HD analysis, the developmental mismatch challenge. I have pointed out that my argument is, in essence, a modal one. It is empirically plausible, and hence logically possible, that *some mental disorders result from developmental mismatches,* but the HD implies that this is logically impossible, so the HD analysis is incorrect. The HD analysis also makes a prediction about clinical usage and I have given reasons for my skepticism about that prediction. I have explored two sorts of responses that Wakefield might give and discussed why I think they are unsatisfying.

Acknowledgments

I'm grateful to Robyn Bluhm, David Frank, Ginger Hoffman, Brent Kious, and Jonathan Tsou for comments and feedback on an earlier draft. I'm also grateful to the audience members of the 2015 meeting of the Association for the Advancement of Philosophy and Psychiatry in Toronto, where some of this material was presented.

References

Agrawal, A. A., C. Laforsch, and R. Tollrian. 1999. Transgenerational induction of defences in animals and plants. *Nature* 401: 60–63.

Allen, N. B., and P. B. T. Badcock. 2003. The social risk hypothesis of depressed mood: Evolutionary, psychosocial, and neurobiological perspectives. *Psychological Bulletin* 129(6): 887–913.

American Psychiatric Association. 1980. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. American Psychiatric Association.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association.

Bolton, D. 2001. Problems in the definition of 'mental disorder.' *The Philosophical Quarterly* 51(203): 182–199.

Buller, D. 1997. Individualism and evolutionary psychology (or, in defense of "narrow" functions). *Philosophy of Science* 64(1): 74–95.

Dretske, F. 1986. Misrepresentation. In *Belief: Form, Content, and Function,* R. Bogdan (ed.), 17–36. Clarendon.

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.

Garson, J. 2015. *The Biological Mind: A Philosophical Introduction*. Routledge.

Garson, J. 2016. *A Critical Overview of Biological Functions*. Springer.

Garson, J., and G. Piccinini. 2014. Functions must be performed at appropriate rates in appropriate situations. *British Journal for the Philosophy of Science* 65(1):1–20.

Glover, V. 2011. Prenatal stress and the origins of psychopathology: An evolutionary perspective. *Journal of Child Psychology and Psychiatry* 52(4): 356–367.

Gluckman, P., A. Beedle, and M. Hanson. 2009. *Principles of Evolutionary Medicine*. Cambridge University Press.

Gluckman, P., and M. Hanson (eds.). 2006. *Developmental Origins of Health and Disease*. Oxford University Press.

Heim, C., and C. B. Nemeroff. 2001. The role of childhood trauma in the neurobiology of mood and anxiety disorders: Preclinical and clinical studies. *Biological Psychiatry* 49(12): 1023–1039.

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

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.

Lilienfeld, S. O., and L. Marino. 1995. Mental disorder as a Roschian concept: A critique of Wakefield's "harmful dysfunction" analysis. *Journal of Abnormal Psychology* 104: 411–420.

Lilienfeld, S. O., and L. Marino. 1999. Essentialism revisited: Evolutionary theory and the concept of mental disorder. *Journal of Abnormal Psychology* 108(3): 400–411.

McGowan, P. O., A. Sasaki, A. C. D'Alessio, S. Dymov, B. Labonte, M. Szyf, G. Turecki, and M. J. Meaney. 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience* 12(3): 342–348.

McGuire, M., and A. Troisi. 1998. *Darwinian Psychiatry*. Oxford University Press.

Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories*. MIT Press.

Millikan, R. G. 1989. In defense of proper functions. *Philosophy of Science* 56(2): 288–302.

Murphy, D. 2005. Can evolution explain insanity? *Biology and Philosophy* 20(4): 745–766.

Murphy, D., and S. Stich. 2000. Darwin in the madhouse: Evolutionary psychology and the classification of mental disorders. In *Evolution and the Human Mind: Modularity, Language, and Meta-Cognition,* P. Carruthers and A. Chamberlain (eds.), 62–92. Cambridge University Press.

Murphy, D., and R. L. Woolfolk. 2000a. The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, and Psychology* 7: 21–252.

Murphy, D., and R. L. Woolfolk. 2000b. Conceptual analysis versus scientific understanding: An assessment of Wakefield's folk psychiatry. *Philosophy, Psychiatry, and Psychology* 7(4): 271–293.

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.

Pigliucci, M. 2001. *Phenotypic Plasticity: Beyond Nature and Nurture*. Johns Hopkins University Press.

Price, J., L. Sloman, R. Jr. Gardner, P. Gilbert, and P. Rohde. 1994. The social competition hypothesis of depression. *British Journal of Psychiatry* 164(3): 309–315.

Richters, J., and S. Hinshaw. 1999. The abduction of disorder in psychiatry. *Journal of Abnormal Psychiatry* 108(3): 438–445.

Rottenberg, J. 2014. *The Depths: The Evolutionary Origins of the Depression Epidemic*. Basic Books.

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

Spitzer, R. L., M. Sheehy, and J. Endicott. 1977. *DSM-III*: Guiding principles. In *Psychiatric Diagnosis,* V. M. Rakoff, H. C. Stancer, and H. B. Kedward (eds.), 1–24. Brunner/Mazel.

Tollrian, R., and S. I. Dodson. 1999. Inducible defenses in Cladocera: Constraints, costs, and multiple environments. In *The Ecology and Evolution of Inducible Defenses,* R. Tollrian and C. D. Harvell (eds.), 177–202. Princeton University Press.

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

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. 2000. 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. 2011. Darwin, functional explanation, and the philosophy of psychiatry. In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory,* P. R. Adriaens and A. De Block (eds.), 143–172. Oxford University Press.

Wakefield, J. C., S. A. Kirk, K. J. Pottick, D. K. Hsieh, and X. Tian. 2006. The lay concept of conduct disorder: Do nonprofessionals use syndromal symptoms or internal dysfunction to distinguish disorder from delinquency? *Canadian Journal of Psychiatry* 51: 33–39.

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.

Watson, P. J., and P. W. Andrews. 2002. Toward a revised evolutionary adaptation analysis of depression: The social navigation hypothesis. *Journal of Affective Disorders* 72(1): 1–14.

Woolfolk, R. L. 1999. Malfunction and mental disorder. *The Monist* 82(4): 658–670.

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