CHAPTER 11

What Are Mental Disorders?

A middle-aged government employee becomes convinced that the government inserted a tracking device in his rectum while he slept. A college freshman gets so paralyzed by anxiety that she fails all of her classes and drops out of school. A teenager squeezes a mouse to death for entertainment. All of these individuals have symptoms of mental disorders, including schizophrenia, panic disorder, conduct disorder, and perhaps others. What are mental disorders? What do these constellations of thinking, feeling, and acting have in common that makes them all the same kind of thing?

This question feeds into a bigger one: What are disorders more generally, mental or physical? Here, I'll focus on the more specific question of what mental disorders are. One reason to raise the question is that mental disorders are fascinating and they challenge some of our basic ideas about human nature and the mind. Philosophers of mind are rightly intrigued by mental disorders because they promise to shed light on other questions, such as the nature of thought, rationality, and consciousness (Graham 2010). Closer to the topic of this book, I want to know whether mental disorders should always be understood in terms of biological dysfunction.

Knowing what mental disorders are has practical benefits, too. Since the 1970s, American psychiatry has been rocked by a series of controversies about whether or not certain conditions are mental disorders. In the early 1970s, psychiatrists in the American Psychiatric Association (APA) engaged in heated debates about whether or not homosexuality was a mental disorder. The APA famously put the matter up for a vote; its members voted 58 percent to 37 percent to drop the diagnosis (Kirk and Kutchins 1992, 88). More recently, members of the APA have argued with each other about categories like premenstrual dysphoric disorder, gender identity disorder, and Asperger syndrome. They also argue about the proper boundaries of these categories. For example, can a person grieving the loss of a loved one be diagnosed with depression?

Knowing what mental disorders are can help to resolve just those kinds of debates. If we know what mental disorders are, we know what evidence we'd have to muster to decide whether a particular condition really is one. One reason psychiatrists in the early 1970s couldn't agree about whether homosexuality should belong in their classification system was that they couldn't agree on what kind of evidence is needed to prove that *anything* is a mental disorder. Some psychiatrists argued that it is a disorder because, they claimed, it stems from bad genes or bad parenting. They thought you must look at genetics and childhood to know whether there's a real disorder. Other psychiatrists argued that history is irrelevant. To decide whether something is a disorder, they thought, you only need to consider whether people with the alleged condition can be well adjusted and happy (see Garson 2015, chapter 8).

Psychiatrists still disagree about what sorts of evidence is needed to show something is a disorder. I once asked the chair of the *DSM-5* task force, David Kupfer, why the APA included premenstrual dysphoric disorder (PMDD) in the most recent edition of their manual. He said PMDD is included because it has a known biochemical mechanism and it can be treated with selective serotonin reuptake inhibitors. For him, that was enough. People can differ about what sorts of evidence you need to decide.

In this chapter, I'm going to try to make some progress on this debate about what mental disorders are. Specifically, I'm going to consider a very influential theory, Jerome Wakefield's harmful dysfunction account (1992; 1999a; 1999b). He thinks mental disorders always, logically speaking, involve inner dysfunctions. My conclusion will be largely negative. While Wakefield and I agree, in the main, about what functions and dysfunctions are, I don't think mental disorders always spring from dysfunctions. One of the most interesting implications of GSE is that some mental disorders are probably functional, not dysfunctional. It's worth taking on Wakefield's theory in particular because so many psychiatrists have explicitly endorsed it (see below).

Here's the plan for the chapter. I'll first introduce Wakefield's view, and position it in relation to more long-standing debates about mental disorders (Section 11.1). Then I'll introduce a common critique of Wakefield, known as the "evolutionary mismatch" problem (Section 11.2). Section 11.3 will present a novel version of the mismatch objection, which I call the *developmental mismatch* problem. Then I'll discuss how GSE further erodes the foundations of Wakefield's hybrid view (Section 11.4). If GSE is right, there are probably many more functions,

and possibly mismatches, than we thought. Finally, I'll consider, and respond to, two potential objections (Section 11.5).

11.1 Mental Disorders and Biological Dysfunctions

Here's one way of approaching Wakefield's position. On the face of it, when we say that someone has a mental disorder, we're saying that the condition is undesirable, or it's unusual, or that it violates commonly held standards of what's good or proper in society. There are, however, many conditions that are undesirable, uncommon, or otherwise violate widely shared norms of what's proper but still aren't mental disorders. Being a neo-Nazi, or any other "home-grown terrorist," arguably meets all three, but most psychiatrists don't think it's a mental disorder. What's this something more that mental disorders have?

Here's an idea that emerged in the 1970s. What distinguishes a mental disorder from any socially disvalued condition is that when someone has a mental disorder, something has gone wrong inside of them. Consider the difference between someone who murders another because of a religious ideology – say, a man murders a doctor for performing abortions – and someone who kills because he believes that the victim was a disguised and malevolent alien. The second person surely has a mental disorder. That's because something went wrong inside of him, in his thought life, that caused him to act that way. The first might just be in the grips of a dangerous ideology; in the second, something in his brain isn't working the way it's supposed to.

That puts us in the position of having to spell out this idea of "there being something wrong inside of you" or there being something inside of you that "isn't working right." A tempting way to do so is in terms of function and dysfunction (Klein 1978; Spitzer and Endicott 1978). What distinguishes the second murderer from the first is that his mind, or his brain, wasn't functioning well. It was dysfunctional. The only thing that could make one think others are disguised aliens is a dysfunctional brain. This leads us to the idea that in order for something to be a mental disorder, it has to stem from an inner dysfunction. That isn't enough for being a mental disorder, but it's a necessary condition.

Wakefield articulated a theory of disorder along just these lines. He thinks, simply put, that disorders (whether mental or physical) are "harmful dysfunctions." Wakefield holds a hybrid view about disorders; it mixes facts and values. There are two aspects to his definition, harm and dysfunction. The "harm" part just means that the disorder is deemed

harmful by the person's culture. That's the value part. The "dysfunction" part means that the disorder is caused by an inner mechanism that fails to perform its function. That's the fact part. Wakefield's view is initially attractive because it acknowledges some truth in both "normativist" and "naturalistic" views – that is, it sees the role of value judgments while nonetheless insisting on a solid natural foundation.

What are dysfunctions? For Wakefield, as for myself, functions are selected effects. We differ, however, in details. Wakefield often writes as if the only kind of selection process relevant for functions is natural selection in the evolutionary sense. That means, for him, for something to be a mental disorder it must stem from the failure of a mechanism to perform its evolved function. This move purports to place psychiatry on firm evolutionary footing. In Wakefield's view, psychiatrists should resolve debates about classification, in part, by considering evolutionary history. It's not surprising that many psychiatrists and psychologists with a biological orientation have explicitly endorsed Wakefield's theory (e.g., Klein 1999; Spitzer 1999; First 2007; Nesse 2007).

I find Wakefield's view problematic. When we dig into the empirical literature, we find some well-recognized mental disorders that don't necessarily involve dysfunctions. As I'll explain, there are reasons to suspect that disorders as varied as generalized anxiety disorder, antisocial personality disorder, and even the delusions associated with schizophrenia might actually have functions, just like viral conjunctivitis. Maybe they're adaptations, not dysfunctions.

I'm not saying that some mental disorders are actually adaptations. I am saying they might be. It is *logically* possible (it is not a contradiction in terms), *nomologically* possible (it is consistent with the laws of nature), and *epistemically* possible (consistent with what we know) that some mental disorders are adaptations. Since Wakefield's theory implies that it isn't logically possible for a mental disorder to be an adaptation, then, ipso facto, it's not possible in any other way. Why exclude this prospect out of hand? Although we might decide, after much more research, that mental disorders generally involve dysfunctions, nobody's in a position to say that yet. Moreover, there are promising new research programs premised on the idea that some mental disorders are adaptations, such as the Developmental Origins of Health and Disease (DOHaD) project (see Gluckman and Hanson 2006; Glover 2011). I'll return to DOHaD later in the chapter.

I think it's important to recognize that mental disorders don't always involve dysfunctions. To take it a step further, it's good to see that we actually don't have a clue what mental disorders are. When we have a

friend or a loved one with a severe mental disorder, it can be extremely scary, perplexing, or sad. We want to understand what's going on because that understanding gives us a sense of mastery, even if a false one. It gives us a chance to pretend that we know what's happening and that, in principle, we know how to fix it. It's better to stop pretending that we really understand what mental disorders are. In fact, there are some mental health advocacy groups, like the Icarus Project, that don't even like using the term "mental disorder," because they think it conveys a false sense of understanding. Instead, they use the term "madness." "Madness," although quaint, is a term that doesn't pretend.

11.2 Mismatch or Dysfunction?

I'm not the first to argue that mental disorders don't necessarily involve dysfunctions. Many others have done the same. This is the basis for the "evolutionary mismatch" critique of Wakefield's view – see Lilienfeld and Marino (1995, 416; 1999, 406–407); Richters and Hinshaw (1999, 442); Woolfolk (1999, 662); Murphy and Stich (2000, 81–84). The idea is that some disorders, such as psychopathy, depression, or anxiety, are adaptations, not dysfunctions; they've just lost their value in our current environments.

Consider evolutionary psychology – specifically the version of evolutionary psychology that hit the academic scene in a big way in the early 1990s (Buss 2008; see Garson 2015, chapter 3). Proponents of this view often assume, as a working hypothesis, that many of our current-day mental capacities evolved in our shared Pleistocene past, from about two million years ago until about 10,000 years ago. (There are many different evolutionary approaches to psychology; the so-called Santa Barbara school is just one of them, and even this is a somewhat simplistic characterization of their ideas – see Confer et al. 2010 for a more up-to-date synopsis.)

Many evolutionary psychologists actually think some of our major mental disorders, like depression, the anxiety disorders, and even obsessive-compulsive disorder, might be adaptations, not dysfunctions. Nesse (2000), in a now-famous paper, reviews many different adaptationist hypotheses about depression. Brüne (2008) is a large survey of adaptationist hypotheses about mental disorders. If they're right – and that's a big if – then they wouldn't be dysfunctions. At worst, they'd be mismatches, like getting startled by a rubber snake. That's not a dysfunction; it's an adaptation. It's an example of a system performing its function perfectly well, just in the wrong environment.

This is such a pivotal point – that mismatches aren't dysfunctions – that it's worth lingering on for a moment. In Chapter 8, we decided that in order for a mechanism to be dysfunctional, it's not enough that it's in the wrong environment. Instead, there has to be something about its inner constitution that prevents it from performing its function. Remember the blindfold example: If being dysfunctional just amounts to being unable to perform one's function, then my eyes would be dysfunctional just because I'm blindfolded. To make sense of this, I urged the view that if something's dysfunctional, then, even if it were in its normal environment, it still wouldn't perform its function. I then identified the normal environment with the trait's selective environment.

We also saw that people describe a trait's function in different ways. Some describe a trait's function in terms of its proximal effects (like the heart's beating) and some in terms of its distal effects (bringing nutrients to cells). Following Neander's lead, I said that we should restrict functions to those effects that are the most proximal to the trait. I went beyond Neander by saying that proper functions are just proximal functions. It's not as if a trait has different functions, some proximal and some distal. Strictly speaking, it's wrong to say the function of the heart is to bring nutrients to cells. It is fallacy of division: attributing a property of a whole (the circulatory system) to a part.

Now, go back to the snake detector reflex. (I'm supposing, for the sake of argument, that there's a mechanism in the human brain that has the function of detecting snakes; see Isbell 2009 for an interesting discussion.) Technically, this is a response function. There is a mechanism that has the function of releasing adrenalin in response to a certain configuration of visual stimuli (an S-shape of a certain size and thickness, say). If I jump back at a life-like rubber snake, there's no dysfunction. Everything is working just as it should. (In fact, it would be dysfunctional if I didn't startle at the life-like snake.) My snake detector mechanism just isn't in its normal environment. Its normal environment is one in which there are real snakes slithering around, not a bunch of rubber toys. If you put the mechanism in its normal environment, with real snakes slithering around, it would respond just fine.

If evolutionary psychologists like Nesse are right, the mechanism that causes us to be depressed, for example, is the same sort of thing. It makes us to feel low when we're confronted with certain life situations, such as being attached to an unattainable goal. It has the function of helping us to detach from those goals, and thereby helping us to adjust to the reality of our situation. That's how it helped our Pleistocene ancestors. The problem

is that, nowadays, getting depressed generally isn't a useful response to the situation one is facing. Struggling to find a job can cause depression. If the struggle is due to something you have no control over, like a recession that causes a low (but not exceedingly low) demand for a certain specialty, then it's not time to get depressed; it's time to redouble your efforts to get a job. There's a mismatch between the way someone is designed to respond, and the response that would be adaptive in the current situation, but there's no dysfunction. Being mismatched and being dysfunctional aren't merely two different ways of saying the same thing. They're mutually exclusive.

Mismatch hypotheses aren't universally accepted. Like any scientific hypothesis, they can be revised or rejected in light of new data. Murphy (2005) and Faucher and Blanchette (2011), for example, criticize mismatch hypotheses for phobias; Murphy (2005) criticizes a certain mismatch theory of depression, the social competition hypothesis. This process of criticism and revision is crucial for healthy science and it's to be unconditionally welcomed. Note, however, that even if we decide to drop one or two specific mismatch hypotheses, we're not entitled to throw out the whole research program. That's just not how research programs work (Lakatos 1970).

Some people don't take evolutionary psychology as a whole very seriously because they think that it's nothing more than "just-so stories." Gould and Lewontin (1979) famously chided their colleagues for inventing cheap and often unverifiable evolutionary explanations for various traits. Lewontin (1998) specifically took on the evolutionary psychologists, and a number of philosophers followed suit, most notably Buller (2005) and Richardson (2007). If you are one of those people, then you might as well move on to the next chapter now, because you're not going to accept Wakefield's view anyway. In other words, if you think we're rarely licensed to make any sorts of claims in this direction, you'll agree that Wakefield's proposal is practically useless for any real diagnostic purposes – you're already on my side, even if not for the same reasons.

11.3 Developmental Mismatches

In earlier work, I proposed a novel criticism of Wakefield's view, based on what I called a "developmental mismatch" (Garson 2010; 2015, chapter 8; forthcoming b). A developmental mismatch is a special sort of evolutionary mismatch. Often, natural selection doesn't confer a fixed phenotype on an organism, but bestows upon it a certain capacity to "select" between adult phenotypes during its development. This is called

developmental plasticity and it's a kind of phenotypic plasticity (Pigliucci 2001).

There are two kinds of developmental plasticity relevant here. The first can be dubbed "switching." This takes place when a small handful of adult phenotypes is available to the developing organism, and the organism "selects" the most appropriate one. A famous example is the tiny crustacean *Daphnia*. If *Daphnia* eggs are released in waters swarming with predators, they grow a tough, helmet-shaped head. This is good for *Daphnia* because it helps it avoid getting eaten, though it makes it swim more slowly. If there are no predators around, it grows a normal, streamlined head. To keep things simple, we'll treat these two forms as being innate; we'll imagine that the genome encodes a program that says, roughly, "If predators, grow helmet; otherwise, no helmet." Chemical signals from the predator just trigger the right developmental path.

We can refer to the second kind of developmental plasticity as "imprinting." Here, there are a vast number of potential adult phenotypes available, depending on the nuances of the formative environment. In contrast to "switching," the local environment plays a more powerful role in shaping the form of the adult phenotype. An example is filial imprinting in goslings. Goslings can imprint onto pretty much any large, moving object that they encounter during the developmental window. There are potentially billions of phenotypes available to them depending on their environment. Being disposed to follow one's mother is one phenotype. Being disposed to follow a pair of boots is another. Being disposed to follow a porcupine is a third.

Developmental plasticity creates a new sort of mismatch in addition to the evolutionary kind. I'll stay with the "switching" case for the sake of illustration. Suppose an organism such as *Daphnia* is programmed to adopt phenotype P1 in environment E1, and P2 in E2. Suppose it is placed in environment E1, then it develops P1. Then, suppose the animal is moved to E2. Now, there is a mismatch between its phenotype and its environment, <P1, E2>, but it's not an evolved mismatch. The mismatch isn't due to a transition from the environment its distant ancestors faced to its current-day environment. The mismatch is due to the transition from its formative environment to its mature environment.

Here is the question: Is there any analogy to *Daphnia* in the realm of mental disorders? Is it possible that some mental disorders represent developmental plasticity? Consider generalized anxiety disorder (GAD). The *DSM* characterizes GAD as an excessive, diffuse, and disproportionate worry that disrupts normal life activities like work or school (American

Psychiatric Association 2013, 222). Some psychiatrists believe that GAD, and anxiety disorders more generally, result from developmental plasticity. They're not dysfunctions.

The psychiatrist Vivette Glover (2011) points out that children who grow up in high-stress formative environments are more prone to anxiety disorders as adults. This is a well-established correlation (e.g., Heim and Nemeroff 2001; McGowan et al. 2009). She goes on to speculate about the reasons for this correlation. She hypothesizes that the high-anxiety phenotype is an adaptation designed to cope with high-stress environments. The idea is that high-anxiety individuals tend to be more vigilant to potential dangers in their environments. They're more sensitive to potential danger signals around them. That would be very helpful if the level of danger in the formative environment is a good indicator of the level of danger in the mature environment. If she's right, then GAD is an example of developmental plasticity.

If GAD results from developmental plasticity, then it'd be very easy for mismatches to come about. Suppose someone is raised in a high-stress environment, and then is moved to a low-stress environment right after the developmental "window" has shut. There is now a permanent mismatch between the individual's proneness to anxiety and the likelihood that the person will actually confront a dangerous situation. It's just like the *Daphnia* with the helmet-shaped head that is moved to safe waters.

My point isn't that GAD definitely is a mismatch. It might or might not be. My point is that GAD is a mental disorder regardless of the truth or falsity of that hypothesis. As noted above, I think it's logically, nomologically, and epistemically possible that GAD is a developmental mismatch, and also a bona fide mental disorder. Wakefield thinks it's logically impossible that GAD is both a mismatch and a mental disorder. If it's a mental disorder, it's not a mismatch, and vice versa. More on this momentarily.

11.4 Generalized Selection Processes and Mental Disorders

I haven't said anything yet about how GSE can be brought to bear on this problem. I talked about the possibility of evolutionary mismatches, and then expanded this idea to include developmental mismatches, too. Still, the significance of GSE should be obvious. GSE expands the domain of entities that can possess direct proper functions, thereby increasing the likelihood of that a given condition is functional, not dysfunctional. Any sort of selection process, including operant conditioning, antibody selection, and neural

selection, can create functional traits, and also potentially result in mismatches. In short, the more function-bestowing selection processes there are, the greater the probability that any given mental disorder has a function.

The psychologists John Richters and Stephen Hinshaw (1999, 442–443) gave a simple but true-to-life illustration of the point I'm after. Consider conduct disorder, marked by "a repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated" (American Psychiatric Association 2013, 469). Unlike antisocial personality disorder, conduct disorder is typically diagnosed in children. These violations might include frequently starting fights, using weapons in fights such as broken bottles, being cruel to animals, breaking into homes, or theft. Suppose a young boy is raised in an abusive home (a known risk factor; ibid. 473). Suppose that, in the abusive context, aggressive and violent behavior tendencies are implicitly reinforced. For example, imagine it's a place where unexpected outbursts of violence are a great way of signaling, "keep a safe distance from me," and they come to be reinforced because they signal this. Then, those angry outbursts, and the psychological dispositions that support them, have functions. They have the function of warning would-be abusers to stay away. To be more accurate, we should say that the underlying *disposition*, which Richters and Hinshaw call a "hostile-world orientation," has a function – namely, the function of generating violent behavior on demand.

Now, suppose the child moves to a nonabusive home, where the violent behavior isn't needed anymore. The underlying disposition – this "hostile-world orientation" – keeps generating the behavior, but there are no more benefits, only drawbacks. Again, it's just like the helmet-shaped *Daphnia* in safe waters. The psychological disposition still performs its function just fine, but it's not in the normal environment for its fulfillment. It's mismatched.

Another example of a potential function, this time involving neural selection, is substance abuse (American Psychological Association 2013, 561). Scientists who study addiction believe that drugs like cocaine effectively hijack or coopt the neural mechanisms involved in the reinforcement of behavior (Kauer and Malenka 2007). Specifically, there is a tract of dopamine neurons that runs from the midbrain to the limbic system's nucleus accumbens. This is the mesolimbic dopamine tract. Any action that causes this tract to release dopamine is reinforced. Though the exact mechanisms are still a matter of controversy, it's likely that synapse selection is involved. The release of dopamine acts as a global signal that selectively strengthens just those

neural connections that caused the behaviors that led to pleasure (Schultz and Dickinson 2000, 490). Sadly, this mechanism can have terrible outcomes. If you give rats the opportunity to self-stimulate mesolimbic dopamine neurons, they'll do that to the exclusion of everything else (Witten et al. 2011). This mechanism is likely involved in drug addiction, too.

Once a synapse is selectively strengthened this way, it comes to acquire a new function. It has the function of causing the behavior that led to its differential retention. If that behavior included seeking out and using drugs, then that becomes its (direct proper) function (Garson and Papineau in prep). Some of the conditions that the *DSM*-5 lists are probably adaptations, not dysfunctions.

A fascinating final example of a potential function comes from the delusions associated with the paranoid subtype of schizophrenia. One somewhat adventurous theory holds that the core dysfunction involved in paranoid-type schizophrenia is a fairly low-level perceptual one. I see a coffee mug sitting next to a coffee machine and a bowl of sugar. The perception of the coffee mug is "decoupled" from the context that lets me see a meaningful connection between the three objects. I also form connections between objects in new and unusual ways: for example, the floral design on the coffee mug with the floral print of my colleague's jacket. Clearly, the fragmentation of the perceptual scene – my inability to form meaningful connections between things – is going to make it hard to navigate my normal responsibilities (Uhlhaas and Mishara 2007; also see Bovet and Parnas 1993).

In response to this perceptual disruption, I start to form delusions. These are beliefs that society considers bizarre but that might actually help me make sense of this strange new world. They strike me with the force of an epiphany or religious conversion. For example, I might form the belief that someone is conducting experiments on me, or that people are playing games with me, or that I've been appointed for a divine mission. Everything clicks into focus in a new way. If that's right, then the delusions are learned responses to the perceptual abnormalities. The delusional system of beliefs gets "fixed" over others because of its useful consequences in the way that Papineau (1984) describes (see chapter 4). It's an adaptation, like the "hostile-world orientation" invoked to explain conduct disorder. This theory of delusion, together with GSE, implies that delusions have functions; they're not dysfunctions.

Now, it must be said, consistent with Wakefield, that even if this theory of schizophrenia is right, schizophrenia still stems from a dysfunction – namely, the low-level perceptual dysfunction that skews my ability to

connect perceptions correctly. The delusions would be a secondary, adaptive, response to that dysfunction, like having a fever when you're fighting the flu. The fever is part of the constellation of symptoms that make up the flu, but it's an adaptive response to the infection. The point is that to the extent that we were inclined to think that the delusions themselves were dysfunctions, we were wrong – they're adaptations. This example shows us how easy it is to be led astray when we label something a "dysfunction." It also shows us how easy it is to go wrong in how we treat schizophrenia. If the delusions represent a secondary, adaptive response to perceptual abnormalities, then it'd be useless at best – and harmful at worst – to target the delusions as if they were the core problem (Bovet and Parnas 1993, 595).

The idea that schizophrenia might represent an adaptation isn't new. Several researchers over the last sixty years have conjectured that some aspects of schizophrenia might have adaptive benefits and might be retained precisely for those benefits. One of the most famous psychoanalytic theories of schizophrenia in the 1950s was the "double-bind" theory, which was developed by Gregory Bateson and his colleagues but inspired by the work of psychoanalyst Frieda Fromm-Reichmann (Fromm-Reichmann 1948; Bateson et al. 1956). This view said that, as a child the person with schizophrenia must have been repeatedly confronted with a kind of "lose-lose" situation (typically imposed by the mother) where any coherent response would be penalized, and which forced the child to adopt radical solutions, such as delusions and incoherent speech – although I'm not endorsing this theory. Others, like the neuroscientist and psychiatrist Solomon Snyder, suggested that the delusions of schizophrenia might have some adaptive role in coping with disorganized thoughts (Snyder 1973, 66). The idea that schizophrenia or its components could be adaptive is not some fringe theory but one that mainstream psychiatrists have entertained for decades. While some of its incarnations have little going for them – like double-bind – we can't get rid of the bigger picture.

11.5 Objections and Replies

I'll consider two criticisms in closing. I said we should reject Wakefield's theory. He says mental disorders require inner dysfunctions; I say something can be a mental disorder even when there's no dysfunction. Logically speaking, Wakefield has two ways to respond to any of the particular counterexamples I raise. First, he can say that in the case I describe there's no dysfunction, but there's no mental disorder either. Second, he could say

that, in the case I describe, there is indeed a mental disorder, and there's a dysfunction, too. I'll take both responses in turn.

No Dysfunction, No Disorder

Wakefield (2000, 260) has considered the possibility of evolved mismatches before. For example, he considered the idea that antisocial personality disorder might be an adaptation to ancestral environments, not a dysfunction. If that were right, he says, we should stop calling it a "disorder," because it wouldn't be one. It would represent, "normal (though problematic) variation. . .Antisocials would be different and at odds with others in the way that, say, men and woman are different and sometimes at odds, without either being disordered." I suppose he'd be inclined to describe anxiety disorders in the same way if it were shown that some anxiety disorders are adaptive responses to developmental stress.

There are two problems with this rejoinder. First, it strikes me not only as counterintuitive but as contrary to normal medical judgment. For example, the researchers who say that disorders like psychopathy and depression are adaptations still label them as "disorders," or "diseases" (e.g., Gluckman and Hanson 2006; Glover 2011). The medical professionals who think that depression, anxiety, and psychopathy are evolved adaptations, also think that they are bona fide mental disorders. Nesse (2000), for example, thinks that depression is an adaptation, but he still thinks it's a disorder. He says it's a "medical disease" (p. 15). Thus, Wakefield's view doesn't capture normal medical usage.

Wakefield sometimes uses the analogy of fever to bolster his idea that if we were to discover that depression (say) is an adaptation, we shouldn't call it a disorder (Wakefield 2000, 259). Medical professionals believed at one time that fever was a disease or disorder. When they learned that fever is an adaptation, and that it's part of the body's concerted defense mechanisms, they stopped calling it a "disorder," and they were right to do so. He thinks that if we discovered that antisocial personality disorder, for example, is an adaptation, then we should regard it just as we regard fever.

I don't find the analogy between antisocial personality disorder and fever very convincing. I agree that fever isn't a disorder, but I think that even if we discovered that antisocial personality disorder were an adaptation, we shouldn't conclude that it isn't really a disorder. There's a deep asymmetry between these two conditions. Fever has immediate and ongoing benefits for the flu sufferer; it seems to help the immune system

work better, although it's not entirely clear why. Even if antisocial personality disorder did have a benefit in Pleistocene times, it doesn't have any now, as far as we know. There's an obvious reason that medical professionals think fever isn't a disorder, quite independently of how it evolved.

There is a much deeper issue here, and it touches on the question of what the point of a theory of disorder is. (This is my second critique.) Wakefield, I believe, would be prepared to say, given compelling empirical data, that antisocial personality disorder isn't a disorder, that GAD isn't a disorder, and perhaps that schizophrenia isn't a disorder either. That seems like a bad consequence for a theory of disorder. If you want to frame a definition of "mental disorder," and your definition, combined with some empirical data, implies that antisocial personality disorder, GAD, and schizophrenia aren't really disorders, then you ought at least to acknowledge how thoroughly revisionary and stipulative the project is. It shouldn't be presented as a conceptual analysis or a theoretical definition of mental disorder. Reasons should also be given for why we should accept such a revisionary and stipulative definition.

Philosophers have tacitly accepted this principle for millennia. If a preferred analysis of a certain term implies that many of the *paradigm* cases for the term's application might not be in the term's extension, then it's time to rethink that analysis (granting that to rethink it isn't to dismiss it out of hand). If I offer an analysis of what numbers are, and it turns out that 42 might not be a number, then that's a strong sign that something's wrong with my analysis. If I offer a theory of what justice is, and it turns out that knowingly framing and executing an innocent person might not be unjust, then it's likely that something's wrong with my theory. Antisocial personality disorder, schizophrenia, and GAD are paradigm mental disorders. If Wakefield's theory implies that those might not really be mental disorders, then short of some very good reason to radically revise our ontology, we should reject that theory.

A Disorder and a Dysfunction

Now I'll move on to the second part of Wakefield's reply. Wakefield could say, of some of the cases I mention, that there is a legitimate disorder. He could also say that there's an underlying dysfunction, too, and that I've therefore failed to present a solid counterexample.

Take the hypothesis that the delusions of schizophrenia represent adaptations to some low-level perceptual abnormalities (Uhlhaas and Mishara 2007). Wakefield could accept all that but still say that schizophrenia is a

disorder. It's just that the dysfunction turns out to be in a different place than where we first thought it was. We might have thought, prior to this research, that schizophrenia is a disorder and the delusions represent the dysfunction. It turns out (let's suppose) that the delusion is actually an adaptation, but there is a dysfunction involved, only at a lower, perceptual level. That is entirely consistent with his theory. Wakefield would have to hope all the other examples work out in a similar way.

Wakefield has used just this line of reasoning in the conduct disorder case. He considers Richters' and Hinshaw's thought experiment, where a child is exposed to an abusive environment and develops a disposition to engage in violent outbursts but later is moved to an environment where those outbursts are useless or harmful. Wakefield is willing to entertain the hypothesis that, in this case, there is a real mental disorder – conduct disorder. Since he thinks there's a dysfunction, too, for him, it's not a counterexample.

According to Wakefield, where exactly is the dysfunction? What, specifically, has gone wrong? Wakefield (1999b, 468) says:

> 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.

Here's the idea. Let's suppose that in human beings there is a universal, innate mechanism, *M*. *M* is an adaptation. Its function is to gauge the "baseline" level of violence and hostility in an individual's formative environment, and then use this information to shape that individual's personality. It works according to a simple if-then rule: If there is a high level of baseline violence in my environment, then I should adopt a hostile-world orientation and behave accordingly; otherwise, I shouldn't. But the mechanism can go awry. It's designed to detect the baseline level of violence in my broader environment – not just my home but my community and natural environment. The problem is that a home environment is a limited and distorted sample of what the broader environment is like. Maybe my parents are abusive, but nobody else in my community is and there aren't any ecological threats either. *M*, however, doesn't "know" that the broader environment is safe; it just uses the home environment as a proxy for the broader environment, and creates a hostile-world orientation. *M* is dysfunctional because it triggers a hostile-world orientation when it's not supposed to.

I think Wakefield is wrong in his assessment of dysfunction. Our disagreement traces back to the issue of indeterminacy. The case I described with the hostile-world orientation is structurally identical to the case of *Daphnia*. It seems to me that if the *Daphnia* is in an environment swarming with predators, and it grows a helmet-shaped head, and is then moved to a safe environment, there's no dysfunction. Even if it's generally in safe waters, but due to a statistical fluke, there are one or two predators nearby – enough to trigger the "helmet" phenotype – there's no dysfunction either. The mechanism responsible for sampling the immediate environment and producing the corresponding phenotype has discharged its function admirably. First, I think that's intuitive, but I can go beyond intuition here. I gave reasons, in Chapter 7, for saying that proper functions are proximal functions. If we stick to the level of proximal functions in the *Daphnia* case, we can see that there is no dysfunction. Proximally described, the function of *Daphnia*'s switching mechanism can be read as a command: If hormone H is present in the immediate vicinity, in quantities Q, trigger the developmental sequence that normally results in helmet phenotype. The same point holds in the conduct disorder case. The only way one can get the result that there's a dysfunction here is if one has incorrectly latched on to a quite "distal" description of the function in question.

Nor am I contradicting Wakefield's own commitments here. Wakefield, too, seems to agree that proper functions are proximal functions. At one point, he considers Dretske's famous magnetosome case. If you remember, Dretske (1986, 26) discusses an anaerobic bacterium with an inner magnet. The magnet aligns the bacterium with magnetic north. In normal environments, that's the direction of oxygen-free water. It needs to be in oxygen-free water or it will die. What's the function of the magnetosome: to align the bacterium with magnetic north, or to move that bacterium to oxygen-free water? Dretske says it's the former because, if you hold a bar magnet over the bacterium, it'll swim up to it (and hence to the oxygen-rich water) and die. Nothing's malfunctioning; the magnetosome performed its function perfectly well.

Wakefield says he agrees with Dretske about the magnetosome. He says that if we hold a bar magnet above the bacterium, thereby leading the bacterium to deadly waters, the magnetosome isn't dysfunctional. As he puts it, "one is inclined to judge the bacterium disordered when and only when its magnetosome does not successfully orient the bacterium to prevailing north. If one holds a magnet near the bacterium to fool it. . .one does not cause a disorder" (Wakefield 1999a, 386). Wakefield should draw a similar conclusion about the case of conduct disorder.

In short, neither of Wakefield's preferred responses work out well. I've shown that it's possible (in any of the standard senses) for a mental disorder to be an adaptation, not a dysfunction. GSE supports this position by significantly expanding the range of potential functions.

The funny thing is that Wakefield's account is probably the best we have going. Perhaps it's even the "only game in town" – but that doesn't make it right. On reflection, maybe it's not such a bad thing if we don't have an analysis of mental disorder. As I said at the beginning, it's probably better to throw up our hands and admit that we don't really know what mental disorders are then to falsely believe that we do. I'm inclined to think, as Foucault (2006 [1961]) did, that the possibility of mental illness is a kind of shadow side of reason, and that reason just isn't equipped to put a familiar face on it. Maybe we should learn how to live in that unknowingness, rather than plaster it over with tidy definitions.

CHAPTER 12

A New Kind of Teleosemantics

I lean back in my chair in my office and prepare for an upcoming class. I worry about whether I'll finish the book in time. I wonder how my kids are doing in school. My thoughts are straying this way and that. What is it for a thought to be about something? How can a thought, which is inside my head, be about something outside my head? Is there a special sort of relationship between the thought and the thing? There's surely no invisible chain that links my brain to my (future) book or my children, who are on the other side of the East River.

Intentionality – the *aboutness* of thought – is both extremely familiar and deeply mysterious. It's familiar because it's woven into every aspect of thinking and wanting. It's mysterious because it doesn't correspond obviously to any known physical relationship. As Fodor (1987, 97) famously puts it:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep.

So what is intentionality? How can we understand it in terms of more familiar physical properties? Can we understand it in this way at all?

It's worth taking a moment to flesh out the mystery. Some representations, like beliefs, have truth conditions. They can be true or false, correct or incorrect, right or wrong. Perceptual representations can have "veridicality" conditions. They can be accurate or inaccurate. While desires don't have truth conditions, they have satisfaction conditions. They can be satisfied or unsatisfied, fulfilled or unfulfilled, successful or unsuccessful. Normal physical objects, like rocks and trees, don't have truth or satisfaction conditions. They can't be true or false, accurate or inaccurate, satisfied