Defining (Mental) Disorder

A fairly large literature has accumulated on the notion of 'disorder' and related notions like 'disability', 'illness', 'sickness', 'disease', 'distress' and so forth. Different theorists enumerate these notions and the relationship between them a bit differently. In the attempt to avoid a purely verbal dispute I'll follow what seems to be a recent convention of using the term 'disorder' though I will say a bit later to reassure people that I am getting to the heart of the debate.

Dominic Murphy states that 'mental disorder' is a term that is used in a variety of different contexts. In his book 'Psychiatry in the Scientific Image' he maintains that the scientific concerns can be carved off from the extra-scientific concerns. He then proceeds to focus on the scientific concerns. He identifies the scientific concerns as the project of finding out the nature and causes of mental disorder. The extra-scientific concerns include the legal notion of insanity, issues of moral responsibility, and therapeutic concerns such as that of involuntary treatment. While I'm not completely convinced that the scientific concerns can be isolated from the extra-scientific concerns at the end of the day I will attempt to focus on the scientific concerns in this talk. This is related to my greater project of trying to offer a foundation for a science of mental disorder.

The official definition of mental disorder is provided by the American Psychiatric Association in the clinician's handbook the Diagnostic and Statistical Manual of Mental Disorders. The APA states that:

'*…each of the mental disorders is conceptualised as a clinically significant behavioural or psychological syndrome or pattern that occurs in an individual and that is associated with present distress (e.g., a painful symptom) or disability (i.e., an impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a*

behavioural, psychological, or biological dysfunction in the individual. Neither deviant behaviour (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above'.

The DSM view exemplifies a general position that Murphy has dubbed the 'two-stage view'. According to the two-stage view there are two kinds of facts that jointly determine whether an individual is mentally disordered. The first set of facts are characterised as non-evaluative or objective facts about malfunction or dysfunction. The second set of facts are characterised as evaluative facts about harms to persons. The DSM definition can be broken down into three parts and the first two seem to correspond to the two stages of the two-stage view. The first part of the definition talks about 'present distress', 'painful symptoms', 'disabilities', and 'sufferings'. The notion that seems to be relevant here is that of harm to persons. The second part talks about 'behavioural, psychological, or biological dysfunction in the individual'. Facts about dysfunction are thought to be objective in the sense that they are not dependent on or determined by our values.

The rest of the definition, while not fully transparent, makes more sense when viewed in its historical context. Despite there being two editions of the Diagnostic and Statistical Manual prior to 1980 it wasn't until the third edition that there was an attempt to define mental disorder. Coopers discusses how the APA felt the need to define mental disorder in response to sustained pressure and critique from gay rights activists and the anti-psychiatry movement. The third part of the definition is an attempt to state that moral deviance is not sufficient for mental disorder unless the previous two conditions are met. It is basically intended as a direct denial of the claim made by some anti-psychiatrists that 'mental disorder' is nothing more than a label given to people who deviate from social or moral norms.

One thing that is notably lacking from the APA's account of disorder is an account of what makes a disorder psychiatric as opposed to neurological or ‘mental’ as opposed to ‘non-mental’. All that is said about this in the introduction to the manual is that 'the term general medical condition is used merely as shorthand to refer to conditions and disorders that are listed outside the "mental and behavioural disorders" chapter of the ICD’. Since the International Classification of Diseases manual does not tell us the

difference between general medical conditions and psychiatric disorders either I think that it is fair to say that this is merely shifting the problem.

The distinction between psychiatric and neurological disorders is very controversial. The main issue is whether disorders that are currently regarded as psychiatric will become subsumed under neurology with advances in neuro-biology or whether there is something distinctively 'mental' about mental disorders. While this isn't an issue that I will focus on here it is important to note the assumption that psychiatry shares the notion of disorder with the rest of medicine. Mental disorders are thought to be kinds of medical disorders as is reflected in psychiatry's status as a specialist field within medicine. While some theorists have argued that mental disorder involves a different sense of disorder or a metaphoric use of the term 'disorder' the assumption of the two-stage view is that it is the same notion of disorder that is in play. This is important because problems for the two-stage view aren't thought to be problems merely for psychiatry, they are thought to be problems for the rest of medicine as well. This is also important as counter-examples to the two-stage view make use of examples from both psychiatry and general medicine and in what follows I too shall make free use of examples from psychiatry and general medicine.

I now want to turn to Wakefield's 'Harmful Dysfunction' analysis of disorder. His account is another example of the two-stage view and while it differs a little from the DSM definition it is a clearer statement than that provided by the DSM. Another advantage to focusing on Wakefield's account is that he offers arguments for his view and he has received much attention for his sustained and somewhat artful defence of his definition. The Harmful Dysfunction analysis states, in a nutshell, that a disorder is a failure of the evolutionary function of a mechanism where that failure results in harm to persons. While he has written a number of papers in which he states and restates and subtly alters his arguments I think that the following is a charitable reconstruction of why it is that he thinks that we should accept his definition of disorder.

P1) It is a-priori that disorder is due to an inner dysfunction that results in harm to persons. (At this stage he regards 'malfunction' to be a pre-theoretic, folk notion).

P2) a. It is a-priori that it is a matter of scientific discovery what causal process fixes the functions and dysfunctions.

b. It is a-priori that the relevant process is the causal-historical process that is the explanation for the mechanisms existence and maintenance in current populations.

P3) It is a-posteriori that the relevant causal-historical process for fixing which effects are functions and malfunctions (for explaining the existence of the relevant mechanisms) is evolution by natural selection.

C) Disorders are thus failures of an inner mechanisms evolutionary function that results in harm to persons.

Wakefield's general project is part of a strategy that Fullford dubs 'the naturalisation cascade'. The notion is that while a term such as 'disorder' contains both evaluative and non-evaluative components the non-evaluative component can be reduced to another term such as 'dysfunction'. Ultimately the naturalisation project is thought to bottom out in purely causal processes. While there are a variety of terms on offer (for example illness, disability, disease, etc) and while different theorists locate them at different points in the naturalisation cascade this seems to be a general approach to attempting to provide an account of the scientific foundation of medicine in general and psychiatry in particular. The notion is basically that psychiatry can be grounded in medicine and medicine can be grounded in biology. The scientific foundation of psychiatry is thus ensured. While Wakefield doesn't say anything about the naturalisation cascade that Fullford outlined he does draw a parallel between his analysis of disorder and the causal-historical analysis of terms like 'gold' and 'water'. It is important to note that instead of naturalising 'disorder' directly, however, Wakefield attempts to naturalise disorder by way of naturalising function and dysfunction.

While many theorists have objected to Wakefield's Harmful Dysfunction account the majority of the objections have focused on dysfunction rather than on harm. While I shall also focus my critique on dysfunction before I do this I do want to say a few words about the relevant notion of harm. The first thing to note is that the motivation for the harm component comes from the idea that it seems plausible that a person

could have a malfunctioning mechanism and yet not be harmed. An example of this would be someone who had a malfunctioning mechanism that resulted in a phobia of flying. If the person didn't need or desire to travel then the person wouldn't be harmed by their malfunction, however, and thus while they do have a malfunction they do not have a disorder. The main motivation for the inclusion of the harm component in the DSM was that irrespective of whether homosexuality was the result of malfunction it was not a mental disorder insofar as it did not result in harm. One issue that this example seems to raise, however, is whether harms to the individual that result from the prejudice of society count as harms. It seems clear that much more needs to be said about the relevant notion of harm.

Another thing to note about the harm component is that whether an individual is harmed or not is thought to be an evaluative or normative matter. This might seem surprising and yet I haven't found any arguments as to why the relevant notion of harm is thought to be evaluative or normative. While it is often noted that whether a person is harmed by a malfunction can be highly dependent on their social and cultural environment this wouldn't seem to rule out the possibility that there are non-normative or non-evaluative facts about whether or not an individual in a certain socio-cultural environment is harmed and facts about how much they are harmed. There does seem to be a lot of work to be done on the notion of harm before we have a satisfactory account of disorder. I won't tackle this issue here, however. What is important to note for my purposes today is that the two-stage view maintains that there is a normative or evaluative component to disorder but that that is completely separate from the non-evaluative notion of dysfunction. I now wish to turn to a critique of the dysfunction condition and I'll restrict my criticism to the dysfunction component of the two-stage view.

It was once thought that xx (or female) was a malfunctioning sex in the sense that females were malfunctioning males. This was a particularly common view in medicine where anatomy texts took masculine anatomy and symptoms to be standard. We also find this in Freud's theory of human development where females were thought to have the additional problem of coming to see that they were in fact malfunctioning males. We don't regard females are malfunctioning males anymore, however. Instead we regard xx (female) and xy (male) to be dif-functions in the sense

that they are two different (though equally functioning) ways of being. If someone is xxy or xxx, however, then we typically regard them to be malfunctioning. There is a current movement for these alternative genotypes to be regarded as dif-functions rather than dysfunctions, however. One biologist has argued that there are 5 sexes rather than 2 and it is hard to see how purely causal facts can settle these issues one way or another. The relevant issue is how purely causal historical processes determine whether something is a function or a dysfunction or a dif-function.

Murphy maintains that the two-stage view is committed to the idea that functions and malfunctions are independent of human interests.

'…on the two-stage view, the criteria for assessing adequate performance are supplied by nature rather than by a human practice… It is not the view that relative to human goals and interests, we can establish what psychological systems should be like and how they should be arranged to meet those goals and further those interests. Rather, it is the view that psychological normality imposes non human, natural functional standards. Those standards exist independently of what people think they should be'.

He doesn't think that this notion of function creates a problem with respect to non-normative causal processes on the one hand and a normative notion of function and malfunction on the other, however.

"Some people will say that since even this view licenses statements about what some biological system ought to be like, it is in fact normative in a fairly weak sense… All of medicine is normative in this sense – the problem is whether any science is not, though, because all sciences license expectations about what ought to happen in a normal system: stars, for example follow a reliable progression through developmental stages, so we can predict what ought to happen to them'. P 85

In the literature on disorder there has been a sustained critique of the notion that function can be used to ground psychiatry in non-normative facts where those non-normative facts consist in purely causal processes. The main line of argument comes from those who think that function of parts are determined by teleological functions of the whole. This line of criticism is largely inspired by Aristotle's teleological notion of function and it seems to rely on Aristotelian notions of a 'good person'. While I don't wish to engage with Aristotles view here I do think that there is

something to this line of criticism. We can see the problem as one of a complete description of the causal processes in our world failing to entail facts about function and malfunction. This is the familiar point that there seems to be a gap between purely causal processes on the one hand and normative facts about function and malfunction on the other.

With respect to Murphy's example of the stars developmental stages it would seem to me that models of **reliable** progression are models of statistically normal or average progression. If a star does not progress through the stages that the model describes because of intervening causes then I don't think that we describe the star as malfunctioning except insofar as malfunctioning is analysed as deviation from the norm or average that we have built into our model. Mozart was statistically abnormal or deviant with respect to his musical abilities but we don't usually want to say that he was malfunctioning in virtue of his statistical deviation. Similarly, if we attempt to say that xxx is a dysfunction because it is statistically abnormal this doesn't seem to be a satisfactory analysis of the relevant notion of dysfunction. It seems plausible that entire populations could be malfunctioning in the sense of having some medical condition like broken legs or infestation by parasites and it similarly seems plausible that mental disorders could turn out to be far more prevalent than we had supposed. The statistical notion of abnormality thus does not seem to be the relevant notion for an explication of the bio-medical notion of dysfunction. The statistical notion does seem to be the relevant notion in the case of the star, however. One thing that I find interesting here is that insofar as we can say that the star is malfunctioning it is malfunctioning in virtue of falling short of the statistically average process that we have built into our model.

Now, we might have intuitions that these alternative genotypes are appropriately regarded as dysfunctions rather than dif-functions in virtue of the individuals with those genotypes being harmed by them. There seem to be cases where individuals are harmed by things that aren't malfunctions, however, and we do not regard all harms to be mental or physical disorders and so we would need to know more about the relevant notion of harm.

One might maintain that these alternative genotypes are malfunctions because the individuals are unable to reproduce. If we consider things from the level of group selection rather than individual selection and individuals with other sexes were found to invest heavily in their kin, for example, then it isn't obvious that they are malfunctioning compared with dif-functioning, however. It might be the case that there is a fact of the matter as to whether individual selection or group selection is the relevant process for fixing the functions and malfunctions but this doesn't seem obvious to me.

Wakefield maintains that there is something special about natural selection with respect to fixing 'natural functions' that obtain independently from us. When he attempts to explain what is special about natural selection he appeals to our explanatory interests, however. **If** we are interested in knowing what it was that past tokens did that accounts for their survival and reproduction **then** evolution by natural selection is the relevant causal process. We first need to identify survival and reproduction as the effects that are relevant for fixing the functions, however. The next step is to identify the effects of mechanisms where those effects contribute towards survival and reproduction or away from survival and reproduction. The notion is (roughly) that if an effect contributes towards survival and reproduction then we have grounds for considering the effect to be a function of the mechanism whereas if the effect hinders survival and reproduction then we have grounds for considering the effect to be a malfunction of the mechanism. This is, of course, a very rough picture. There are issues to do with whether causal processes are enough to fix functions or whether we need to invoke counter-factuals as well. I'm not attempting to offer necessary and sufficient conditions for natural function here, however, I'm just trying to very roughly convey the line that Wakefield and others are trying to run. What is important to note is that the identification of survival and reproduction as the relevant standard seems comparable to the identification of statistical average as the relevant standard in Murphy's example of the star.

It seems that while one can't get normativity from purely causal processes one might be able to get normativity from a conjunction of our explanatory interests together with non-evaluative facts about statistical averages or causal processes. This conclusion is less disturbing for a science of psychiatry than the conclusion that

mental disorder is determined by our moral or social evaluations as the anti-psychiatrists maintained, however.

Murphy maintains that it is far from obvious that the relevant notion of function to ground psychiatry in causal facts is the evolutionary notion of function. In particular, Murphy and Woolfolk maintain that it seems possible that mental disorders could result from harmful failures of spandrels, or ex-aptations. One example could be that if the mechanisms that subserve language don't have the evolutionary function of enabling us to read this wouldn't undermine the status of dyslexia as a disorder. Murphy makes a case for science modelling Cummins functions rather than evolutionary functions in some instances and he maintains that Cummins functions seem more relevant for the medical sciences than the evolutionary notion of function. The notion of a Cummins function is the sense of function in which it is true to say that Harvey understood the function of the heart centuries before Darwin. It seems that Cummins notion of function may be more relevant for the medical notion of disorder.

Attributing a Cummins function to some mechanism (such as a heart valve) seems to similarly require us to identify or choose some output of the overall system that fixes the function of the parts, however. If we grant that the relevant effect of the heart is the pumping of blood then we can attribute functions to the parts of the heart with respect to what contribution they make to the hearts pumping blood. If we want to say why the function of the heart is to pump blood then we can appeal to the role that the heart plays with respect to the biological homeostasis of the organism (or something along those lines). The problem then becomes how we identify the biological homeostasis or survival as the relevant function of the organism. It seems that Cummins functions aren't able to ground function and malfunction in objective facts as we are required to identify what it is that the overall system is 'supposed' to do before we read off functions and malfunctions of the parts relative to what it is that we think the overall system to be 'supposed' to be doing.

Murphy maintains that the malfunction assumption does for psychiatry what the adaptationist assumption does for evolutionary biology. He goes on 'which is to say that sometimes the malfunction assumption is false, sometimes we don't know whether it is true or false but that does not impugn diagnosis'. One thing that

concerns me about the malfunction assumption, however, is that it is supposed to be what grounds psychiatry as a non-evaluative science and that it seems to recommend a methodology for modelling mental disorders. The methodology seems to be that we model 'normal' or 'functional' biological or psychological processes and then we explain disorders by appealing to breakdowns in the model. Much work in the cognitive neuro-sciences and the bio-medical sciences has been done utilising this approach. We have explanations that characterise delusions as being the result of some kind of breakdown in belief formation and / or retention mechanisms; we have explanations of autism as a theory of mind deficit and so forth. The malfunction assumption can't make much sense of other projects that have been done, however. Instead of working with the malfunction assumption some theorists have worked with a function or adaptationist assumption where certain traits (such as histrionic or psychopathic) may be modelled as evolutionary adaptive strategies. Some theorists have attempted to characterise disorders such as depression, schizophrenia, and anxiety as evolutionary adaptive strategies that result in harm in present environments because environmental circumstances are far removed from those in savannah life.

While I'm not going to look at the plausibility of particular theories that have been offered my main point here is that the malfunction assumption does not seem to be required in order for us to study mental disorders scientifically. Instead of attempting to model mental disorders as deviations from some standard one could simply describe the causal processes that seem relevant for some behavioural output while remaining neutral on whether that behavioural output is adaptive or maladaptive. Science can thus model the causes of certain kinds of behavioural symptoms even in the absence of the malfunction assumption. What seems harder to do in the absence of the malfunction assumption, however, is to say what it is about certain conditions or people that means that they are disordered.

My last objection to Wakefield is something that I shall just touch on briefly though it is something that I want to develop. The objection is that he commits himself to too much a-priori when he maintains that malfunction is necessary for mental disorder (and in particular when he identifies causal-historical processes as being necessary for mental disorder). A lot of theorists have attempted to construct counter-examples where we intuitively regard the person to have a mental disorder and yet where it is

stipulated that they do not have an inner malfunction. Wakefield then responds to these objections by maintaining that either there is malfunction after all (and thus the alledged counter-examples actually provide support for his view) or that he is not inclined to regard the individual to be mentally disordered since there is no failure of inner function. Wakefield thus maintains that our intuitions about who is and who is not mentally disordered should be revised to be in keeping with the malfunction assumption. His critics maintain that conversely our intuitions about malfunction should be revised to be in keeping with our intuitions about which individuals are and are not mentally disordered. There seems to be a bit of a stand-off with this tactic.

If we take a step back from the debate it seems to me that what is going on here is that we have three main intuitions about mental disorder.

- The first intuition (or set of intuitions) are around which people are appropriately regarded as mentally disordered and which conditions are appropriately regarded as mental disorders. This intuition is important with respect to the role of prototypical cases helping us fix the reference for our term.

- The second intuition is that mental disorder is a natural kind term. We think that science will discover whatever it is that the prototypical cases have in common that people without mental disorders lack.

- The third intuition is the dysfunction assumption or the notion that people with a mental disorder have an inner malfunction. This basically captures our intuition that there is something wrong with these people.

In maintaining that mental disorder involves inner dysfunction a-priori Wakefield makes the third condition essential and thus non-negotiable. There seems to be a tension between his maintaining that the relevant dysfunctions are to be determined by science on the one hand and his stipulating that science must discover a relevant dysfunction on the other. If scientists succeeded in offering an evolutionary adaptive account of mental disorder, however, and did not characterise them as the result of inner dysfunctions then Wakefield would be left having to conclude that there aren't any mental disorders. Some anti-psychiatrists maintain that mental disorders do require inner malfunction then go on to argue that prototypical cases of mental

disorder do not involve inner malfunction. They thus seem to agree that inner malfunction is necessary for mental disorder and their disagreement comes down to whether scientists will discover that the prototypical cases of mental disorder have an inner malfunction or not. I think that it would be unwise to make any of the above intuitions essential to an account of mental disorder. While they are strong intuitions that go some way towards helping us fix the reference it might be that we need to revise our assumptions depending on how the world turns out. If the facts about function and malfunction can't be read off purely causal facts then it would be hard to see what sense to make of the notion that scientists 'discover' functions and malfunctions by investigating purely causal processes, however.

A problem remains with respect to defining mental disorder. If malfunction isn't necessary for mental disorder or if malfunction isn't a matter for science to discover then what is it that grounds psychiatry or medicine as a scientific discipline? The anti-psychiatrists often maintain that prototypical cases of mental disorder don't involve inner malfunction so much as their behaviour violating certain kinds of social or moral norms. They don't say much more about what kinds of social or moral norm violation are relevant, however, and it seems clear that there are many kinds of social and moral norm violations (such as laziness or strangeness or moral 'badness') where we don't regard the person as being mentally ill. Without more of an account of what it is about their behaviour that we regard to be indicative of disorder their view seems implausible as it stands.

While survival and reproduction are fairly obvious standards for fixing functions if we are interested in evolutionary biology it does seem that our explanatory interests play a crucial role in allowing us to get normativity from the notion of function and malfunction. With respect to medicine there is widespread agreement that disorders that threaten a persons survival are disorders and the reasonableness of this view seems to be inherited from the reasonableness of survival as something that we are interested in promoting. Despite widespread agreement that some conditions are disorders there are controversial cases in medicine, however. There are some conditions where it is unclear whether they are disorders or mere problems in living or whether surgery is a medical requirement or merely elective. The further away one gets from issues of survival and the more expensive the treatment the more

controversy there is as to the status of the individual or the condition as appropriately being regarded as disordered. Psychiatry doesn't seem to be concerned with survival of persons in quite the way that biology and medicine are.

So in my talk today my conclusion is largely negative. The claim is that the malfunction assumption can't do the work that is required of it. The biological notion of function can't ground psychiatry in facts about purely causal processes because the biological notion of function requires us to identify survival and reproduction as the relevant features for fixing functions. Given the explanatory interests of evolutionary biology this is a reasonable thing to do and I have no problem with the scientific status of biology. It is unclear, however, that survival and reproduction are the relevant features for fixing functions and malfunctions for psychiatry. I think that much more work needs to be done on the harm component with respect to understanding what disorders have in common. Coopers has stated that mental disorders might be a little like the notion of 'weeds'. That is to say that our values might well be crucial for determining the class of things that we are interested in. While there is no objective science of weeds because weeds don't have non-evaluative properties in common that differentiate them from non-weeds there can still be a scientific classification of plants, however. It could similarly be the case that individuals with certain kinds of mental disorders share certain causal processes in common though our values are an important part of how we distinguish the mentally disordered from the non-mentally disordered. While the anti-psychiatrists maintain that the relevant values are social or moral it is not the case that any social or moral norm violation is indicative of mental disorder. More work needs to be done on what kinds of norms are relevant. The notion of malfunction isn't very explanatory if it is merely an assumption that we have built into our model where we could redescribe the casual processes that we have discovered as dif-functions instead of dysfunctions.