Philosophical Communications, Web Series, No. 29
Dept. of Philosophy, Göteborg University, Sweden
ISSN 1652-0459

Bengt Brülde

THE CONCEPT OF MENTAL DISORDER

Filosofiska institutionen
Göteborgs universitet

CONTENTS

Chapter One. Setting the stage[1]

This is an essay about the concept of mental disorder, or more specifically, about how this concept should be defined (and why). This question can be formulated in several different ways, depending on in what broader category one thinks that the mental disorders belong. For example, if one assumes (with Svensson 1990) that the mental disorders belong to the broader class "abnormal behaviour and mental afflictions", the question is basically what (if any) abnormal behaviours and mental afflictions that should be classified as mental disorders, or alternatively, where we should draw the line between those abnormal behaviours (etc.) that are pathological, on the one hand, and those that are not, on the other. In a similar way, we can ask what (if any) "problems in living" that it is appropriate to view as mental disorders, and so on.

In this essay, I will conceive of the mental disorders as *conditions* rather than as, for example, behaviours, afflictions, or problems. The question can then be formulated as follows: What conditions (if any) should be categorized as mental disorders? I will also assume that mental disorders are disorders, i.e. that they belong to a wider category disorder (malady, or pathological condition), a category which also includes the somatic disorders. It can then be asked how we should draw the line between pathological and non-pathological conditions, and how we should distinguish those pathological conditions that are mental from those that are physical or somatic.

A related question that I will also touch upon is the more radical question whether there should be a concept of mental disorder at all, or more specifically, whether it is appropriate or legitimate to categorize *any* conditions, afflictions (etc.) as mental disorders, i.e. to group them together under a common heading in this way. Or as Svensson (1990) puts it, "[i]s it correct to conceptualize certain [any] abnormal behaviour and/or mental afflictions in terms of mental illness?" (p. 15) Are "mental illnesses" disease-type problems, on a par with somatic or "ordinary" disease-type problems? (ibid., p. 84)[2] And if the conditions that are currently classified as mental disorders should not (e.g. for some extra-theoretical reasons) be conceptualized in this way, how should they be conceptualized instead?

The main reason why this essay is about mental *disorder* rather than e.g. mental *illness* or mental *disease* is that the most practically relevant category is a broader category that

[1] The work on this essay has been supported by the Bank of Sweden Tercentenary Foundation, in connection with the project Relativism. I also want to thank Frank Lorentzon and Filip Radovic for having read and commented on chapter four.

[2] A closely related question is whether there any good reasons for viewing any mental afflictions as medical problems at all, e.g. as "problems that medicine as an art and science holds the legitimate expertise to deal with" (Svensson 1990, p. 123), problems that can or should be solved with medical means, that belongs to the area of medical responsibility, or the like.

also includes injury, retardation, and so on.[3] The practically important thing is obviously how we distinguish disorder from non-disorder, and not how we draw the line between e.g. disease and injury, or between illness and disease. For example, it seems plausible to argue that people who suffer from disorders are entitled to health care whereas people who do not suffer from disorders are not, but it would be strange to argue that people who suffer from diseases are entitled to health care whereas people who suffer from e.g. illnesses or injuries are not. It is not just of little or no practical importance how we distinguish e.g. disease from injury, distinctions like these also seem rather arbitrary.

Before we look at how these questions might be answered (in chapters two, three, and four), let us first ask ourselves whether it is important how these questions are answered, and if so, why. Why should we care about how the concept of mental disorder is defined? In relation to this question, there will also be a brief discussion of the more radical question whether we should have a concept of mental disorder at all. When we have looked at the possible purposes of a definition, it is time to shift our attention to the question of what constitutes a good definition, i.e. what criteria we should use for assessing different tentative definitions of mental disorder.

What's the point? Some possible purposes of a definition

Is it important to arrive at a well-founded definition of "mental disorder", and if so, why? What is the point, why should we care about how the concept of mental disorder is defined? What purposes do we want the concept to serve?

It is not likely that we need such a definition for any theoretical or scientific purposes. For example, we don't need a well-founded definition of "mental disorder" to arrive at a more correct view of the world, i.e. mental disorder is no natural kind, and it is highly unlikely that there is such a thing as a true definition of the concept (cf. the section on constructionism on pp. 42-43 below). Or alternatively put, there is little or no reason to believe that any of the medical sciences have any need for a category of mental disorder, i.e. that such a concept belongs in any mature explanatory scientific theory about anything, e.g. in the way some diagnostic categories seem to do.

Not everyone agrees with this "anti-theoretical" idea, however. For example, Murphy and Woolfolk seem to believe that we need a concept of mental disorder in scientific contexts (cf. Murphy and Woolfolk 2000b, p. 290). Or more specifically, they seem to

[3] As Culver and Gert (1982) observe, "[i]t would be very useful to have a general term which includes disease, illness, injury, headache, allergy, and so on. We believe that all illnesses, injuries, diseases, headaches, hernias, and even asymptomatic allergies do have something in common. We propose 'malady' as the general term that includes them all." (p. 66) They also add that by "malady", they mean roughly "any condition in which there is something wrong with the person" (ibid., p. 66). In this essay, I will use the term "disorder" instead of "malady".

believe that a concept of mental disorder should provide a way of integrating research on psychopathology into other sciences of the mind, and to further our understanding of phenomena labelled pathological (Murphy and Woolfolk 2000a, p. 242). They don't tell us how this is supposed to happen, however, so there is really no reason why we should accept the view that it is a purpose of a definition to contribute to better explanations. As Wakefield (2000b) points out, direct substantive scientific payoff is *not* the function of a conceptual analysis (p. 268).

Another possible theoretical purpose of a definition is that it should help us construct better classifications. For example, Murphy and Woolfolk (2000a) claim that a concept of mental disorder should produce a parsimonious and consistent nosology, that it should underlie a heuristically fruitful taxonomy of mental disorders (p. 242), and Wakefield (1992) argues that a "correct understanding of the concept [of disorder] is essential for constructing 'conceptually valid' […] diagnostic criteria that are good discriminators between disorder and nondisorder" (pp. 373-374). But how is a definition of "mental disorder" supposed to help us construct a valid or fruitful taxonomy of mental disorders, apart from the trivial idea that it can help us determine what conditions should be included in such a taxonomy? In my view, it seems clear that a definition of "mental disorder" cannot help us distinguish different disorders from each other, i.e. to draw lines *within* the category of mental disorder. However, in those cases where there is a fuzzy boundary or graded transition between some disorder (e.g. a personality disorder) and some "normal" condition (cf. pp. 20-21 below), a definition of "mental disorder" may well help us to distinguish between the two. It is highly doubtful whether such (somewhat arbitrary) distinctions between the pathological and the "normal" have any explanatory value, however, i.e. whether the diagnostic categories arrived at in this way belong in any explanatory theory.

To conclude, we should reject the idea that a conceptual analysis can and should be scientifically useful (directly or indirectly). This is not to deny that the *phrase* "mental disorder" may well be indispensable in some of the medical humanities (like history of medicine or medical anthropology), where e.g. our cultural beliefs about mental disorder are sometimes made objects of study. It is hardly necessary to give a plausible definition of the concept to engage in this kind of endeavour, however.

The alternative view is that if it has any purpose at all to define "mental disorder", then this purpose is *practical*. This view is based on the observation that it has important practical consequences how the concept is defined, primarily for the people who are classified as disordered. We all have a lot of normative and other beliefs about mental disorder, e.g. beliefs about how different disorders are most effectively treated, how one should relate to people with mental disorders, and so on, and these beliefs influence our attitudes and actions in a number of ways. Moreover, there are a substantial number of laws and regulations which contain the concept of mental disorder, and which have im-

portant practical consequences. In short, it makes a difference how the concept is defined.[4]

The primary practical purpose of a definition is that it should help us make better decisions, e.g. about who is entitled to publicly funded health care or to sick leave with compensation. What we ask for in this case is a definition that makes certain reasonable norms and regulations (e.g. that the severely mentally ill has a right to health care) as reasonable as possible. This is not the only way in which a definition can (if accepted) make a difference, however. Its effects can also be mediated by dubious norms and beliefs, e.g. that the mentally ill are not fully human, or that they have less rights than the rest of us. It might be argued that the practical purposes of a definition should somehow take this type of case into account as well, e.g. that we should define the concept of mental disorder in such a way that it helps make the world a better place (if accepted), even in those cases when the definition affects people by being incorporated into implausible norms and regulations. As we will soon see, it is far from certain that this wish can be fulfilled, however.

In any case, for the time being, we can (depending on whether the relevant norms are reasonable or not) distinguish between two kinds of practical purposes of a definition. Let us first look at the first type of practical purpose, i.e. the cases where a definition might help us make better decisions, viz. by making certain (already) reasonable norms and regulations as reasonable as possible.

A list of possible practical purposes (related to plausible norms)

If the primary practical purpose of a definition is that it should (ideally) help us make better decisions, we first have to ask ourselves what kinds of decisions a definition can possibly help us make. Or alternatively put, what types of normative problems can a definition of "mental disorder" help us solve, and what plausible beliefs are there that contain an implicit or explicit reference to mental disorder? Here is a list of practical problems where a well-founded definition might offer some guidance, i.e. a list of different ways in which a definition might be of normative relevance.

1. A definition might help us decide who is entitled to publicly funded health care or medical insurance reimbursement. For example, we tend to believe that people with mental disorders have (at least when the disorder is severe enough) rights to special mental health services, rights which they would not have if they were "merely distressed". This suggests that a definition of "mental disorder" might help us determine

[4] It has been pointed out to me by Eve Garrard (in conversation) that there might be other good reasons for wanting a well-founded definition, i.e. reasons that are neither scientific nor practical. For example, we might (somehow) need a plausible concept of disorder to orient ourselves in the world. I will not investigate this possibility any further in this essay, however.

what services there should be for people who suffer from a certain condition, or whether a certain individual should be denied insurance benefits for mental health services. A well-founded definition cannot settle these questions all by itself, however. To have a disorder may well be necessary for health care, but it is hardly sufficient.

2. A definition should (ideally) help us determine who is entitled to sick leave with compensation. In many countries, people with mental or other disorders have a right to compensation for their "mental injuries", a right that they would not have if they were not disordered. This suggests that a definition might help us determine what compensation arrangements there should be for people who suffer from a certain condition. The presence of a disorder is not sufficient for compensation, however, e.g. in the Swedish case, it is also necessary that the disorder has a detrimental effect on one's ability to work.[5]

3. It might be argued that a definition of "mental disorder" might help us settle certain normative (legal) issues in forensic psychiatry, e.g. that it should help us determine what criminals should be sentenced to psychiatric care rather than to prison (in the Swedish system), or what criminals that should be legally excused from criminal responsibility (e.g. in almost all European countries).

Sometimes people commit crimes influenced by mental illness (or better: while in a condition commonly regarded as a mental disorder). Different societies react to these people in different ways. For example, in most countries (e.g. all European countries except Sweden), what Tännsjö (1999) calls the Excuse Model has been adopted. In these countries, mentally disordered criminals are sometimes legally excused, i.e. they are regarded as not guilty, and thus not punished for what they did. However, this does not mean that these people cannot be detained or subjected to coercive psychiatric treatment on other grounds. In Sweden (where "the Mixed Model" has been adopted), these people are never legally excused. Instead, they are sometimes sentenced to psychiatric treatment (as a form of punishment).

Now, this obviously gives rise to the more general question of how society *should* react to these people, i.e. what general model that is most appropriate. Is it the Excuse Model, the Mixed Model, or what Tännsjö (1999) calls the Full Responsibility Model? It is highly unlikely that a definition of "mental disorder" can help us settle this question, however, i.e. that it can help us determine what type of model that is most appropriate.

However, there are also a number of more specific questions that arise *given* that a certain model is accepted. For example, if the Excuse Model is accepted, we have to ask (a) when (under what conditions) a person should be legally excused, and (b) when he should be detained or given compulsory treatment. And if the Mixed Model is accepted,

[5] In reality, the presence of a disorder isn't even necessary for compensation, since people are sometimes "sick-listed" due to bereavement or other personal crises.

we have to ask (c) when a person should be sentenced to psychiatric treatment. So, can a definition of mental disorder help us answer any of these questions, and if so, *how* can a definition be of help?[6]

(a) On the Excuse Model, the presence of mental disorder is neither necessary nor sufficient for the person to be legally excused. The reason why it is not necessary is that there are other conditions (e.g. dementia or mental retardation) that have a similar status. The reason why it is not sufficient is that there are a number of other criteria that also must be satisfied, e.g. the disorder must have as a consequence that the agent did not know or understand what he was doing, that he did not know that what he was doing is wrong or illegal, or that he could not help doing what he did (that he could not control his acts). It is therefore somewhat doubtful to what extent, if any, a definition of "mental disorder" can help us determine when someone should be legally excused.

(b) Can a definition of mental disorder help a proponent of the Excuse Model to determine when a "criminal" should be detained or given compulsory psychiatric treatment? In this case, mental disorder is certainly a *necessary* condition for detainment (etc), and it might therefore be of some relevance how the concept is defined. Again, it is far from sufficient, however. A number of other criteria must also be satisfied, e.g. the person must be dangerous to himself or others, or the condition must be possible to treat (in the UK). It is therefore unclear to what extent a definition of mental disorder can help us determine when compulsory treatment is appropriate in the forensic case.

(c) Let us now assume that the Mixed (Swedish) Model is accepted. In this case, can a definition of mental disorder help us determine when (under what conditions) a person should be sentenced to psychiatric treatment rather than to prison? Well, the presence of a mental disorder certainly constitutes a *necessary* condition for such a punishment, but again, the mere presence of a mental disorder is far from sufficient. A number of other criteria must also be satisfied, e.g. the mental disorder must be serious, and the person must be dangerous to others. It is therefore somewhat doubtful whether a definition of mental disorder can help us determine when this form of punishment is appropriate.

So far, I have assumed that we can get little normative guidance from a definition if the presence of a mental disorder is "merely" a necessary condition for e.g. compulsory treatment. But is this really the case? Are there no examples of conditions which satisfy the other relevant criteria (e.g. where the person is confused, has difficulty in controlling his acts, is dangerous, etc), but where there is disagreement on whether the condition is a mental disorder? In fact, there seems to exist at least one such case, e.g. the case of antisocial personality disorder (psychopathy). Some people (preferably men) who are categorized in this way are obviously potentially dangerous, and this is partly due to the fact that they sometimes have great difficulty to control their impulses. But are they "mentally ill"? How this question is answered clearly has important consequences.

---

[6] It is worth pointing out that on some views, e.g. like Tännsjö's Full Responsibility Model, the question who is mentally ill or not does not arise at all.

To conclude, it is, in most cases, of little relevance to normative issues in forensic psychiatry how the concept of mental disorder is defined, and it is doubtful whether this ″purpose″ can help us choose between competing definitions. There is one possible exception, however, viz. it is clearly of practical importance whether we regard so-called personality disorders as mental disorders.

4. Another possible purpose of a definition is that it might help us to determine when a mentally ill person might be detained or subjected to psychiatric treatment against her will, e.g. whether a certain individual should be involuntary committed to a mental institution. Sometimes, mentally ill people (or better: people who are regarded as mentally ill) are compulsorily admitted and subjected to coercive psychiatric treatment. This gives rise to several questions, e.g. when, if ever, involuntary hospitalisation is appropriate. Can a definition of mental disorder help us answer this question, i.e. help us decide when compulsory treatment is appropriate? And if so, *how* can a definition be of help?

The presence of a mental disorder is certainly *necessary* for coercive psychiatric care, and it might therefore be of some relevance how the concept is defined. It is far from sufficient, however, i.e. a number of other criteria must also be satisfied, e.g. the mental disorder must be of a serious nature, there must be a need for treatment, the person must (because of his condition) be dangerous to self or others, or the person must be incompetent, i.e. incapable of making an autonomous decision about the treatment.

So again, we have to ask whether we can get any normative guidance from a definition if the presence of a mental disorder is "merely" necessary for compulsory treatment. And again, the central issue seems to be whether there are any conditions that satisfy other relevant criteria (e.g. where the person is dangerous, etc), but where there is at least some disagreement on whether the condition is a mental disorder. In fact, there seems to be at least one case, viz. ″severe personality disorder″ (including ″psychopathy″). Consider the following proposal. In 1999, "the UK government made it clear that it intended to introduce legislation in England and Wales for the compulsory and potentially indefinite detention of people with what is called 'dangerous severe personality disorder', whether or not they had been convicted of a serious criminal offence." (Kendell 2002, p. 110) This proposal seems to assume that personality disorders should be regarded as mental disorders. One reason for this is that the UK has incorporated the European Convention of Human Rights into its legislation. This convention ″prohibits the detention of anyone who has not been convicted by a competent court unless they are 'of unsound mind'″ (ibid., p. 110). And as Kendell points out, this means that ″the Government will have to argue that the potentially dangerous men it wishes to incarcerate are 'of unsound mind', and this means maintaining that they have personality disorders, and that personality disorders are mental disorders″ (ibid., p. 110).

To conclude, it is somewhat doubtful whether a definition of mental disorder can help us determine when compulsory treatment (or preventive detention) is appropriate. There is one possible exception, however, viz. it seems to be of importance, at least in the UK, whether or not the so-called personality disorders are regarded as mental disorders.

5. It can also be argued that a definition of "mental disorder" should help us specify the goals of medicine in general, and the goals of psychiatry in particular. Most of us believe that one of the central goals of the medical enterprise (and the health care system) is to cure or prevent diseases and other disorders, or to help "the sick" in other ways (e.g. by relieving their suffering). Or alternatively put, we normally think of disorders as conditions that require medical intervention, and for which medical intervention is appropriate. As Kendell (2002) points out, it seems reasonable to suggest that to regard a condition as a disorder has something to do with

> if it seems on balance that physicians (or health professionals in general) and their technologies are more likely to be able to deal with it effectively than any of the potential alternatives, such as the criminal justice system (treating it as a crime), the church (treating it as a sin) or social work (treating it as a social problem). (p. 112)7

In short, there is an intimate connection between being a disorder and being a condition that health professionals treat or should treat (cf. Wilkinson 2000). Mental disorders are of course no exceptions in this regard.

This suggests that a well-founded definition of "mental disorder" can help us specify the goals of medicine in general and the goals of psychiatry in particular. This idea is closely related to Wakefield's (1992) idea that a "correct" definition can help us "[a]t an institutional level, [to demarcate] […] the special responsibilities of mental health professionals from those of other professionals such as criminal justice lawyers, teachers, and social welfare workers", and that it can (in this way) help us settle jurisdictional disputes (p. 373).

However, it is worth noting that there is no necessary connection between a condition's being a disorder and its requiring medical intervention (cf. Wilkinson 2000). For example, psychiatry has other legitimate goals besides the goal of treating or preventing mental disorders, and it is far from certain that medical intervention is the most appropriate response to all disorders.

---

7 Kendell also observes that if an apparently effective treatment is introduced for a certain condition, this can produce a decisive change in medical opinion. If a condition responds to a treatment that is not simply a disciplined environment, it is likely that it will (in fact) be regarded as a disorder (Kendell 2002, p 114). This is as likely in the mental case as it is in the somatic case: "The introduction of effective treatments would probably have a decisive influence on psychiatrists' attitudes" (ibid., p. 114).

6. It might also be argued that a good definition of "mental disorder" might help us relate to people with problems in more appropriate ways. For example, we tend to believe that sick people are sometimes entitled to sympathy and support, and that illness might constitute a valid excuse for normally criticisable behaviour.[8] This suggests that a definition should (ideally) help us decide who is (so to speak) "entitled" to more support than the average person and when it is appropriate to excuse or tolerate people for what they do. However, this is not to say that the presence of a mental disorder is in any way sufficient for special treatment (it is also important how the person is diagnosed), or that people cannot be entitled to special treatment on other grounds.

Some more possible practical purposes

Apart from the rather obvious idea that a definition should ideally (if commonly accepted) improve or facilitate communication between different groups and individuals (by establishing a common language), the remaining practical purposes can all be regarded as versions of the more general idea that the concept of mental disorder should (ideally) be defined in a way that makes the world a better place, e.g. in a way which reduces the negative consequences that are sometimes associated with being labelled as mentally ill.

We have already seen how a definition can make the world better by helping us to make better decisions. However, we have also indicated (on p. 4 above) that there are other ways in which a definition can (if accepted) affect the world. For example, a definition of "mental disorder" can affect people by being incorporated into implausible norms and regulations, e.g. the idea that the mentally ill do not have the same rights as others, or that they are not entitled to the same respect as others. That we draw the line between mental disorders and other problems in a certain place can also have a number of unintended side effects, regardless of whether the definition is incorporated into plausible or implausible norms and regulations, e.g. what conditions we regard as mental disorders can affect the large-scale distribution of resources in different ways.

Now, here is a list of possible consequences of how we draw the line between mental disorders and other conditions, consequences that are either more or less unintended or mediated by implausible norms and regulations.

The most immediate consequence of how we define "mental disorder" is of course who is classified as mentally disordered, and how many. This gives rise to further effects. First, that a certain person is classified as e.g. mentally ill (as opposed to "troubled", "afflicted", "mad" or "different") can benefit or harm the person thus classified in different ways, but it can also bring advantages or disadvantages to the person's rela-

[8] The idea that illness can constitute an excuse can partly be explained by Reznek's suggestion that diseases are, by definition, involuntary conditions which cannot be eliminated by a decision, and that we are (therefore) not immediately responsible for our disorders.

tives. And second, what conditions are regarded as mental disorders can also have large-scale social effects, e.g. depending on how large a part of the population that is classified as suffering from a mental disorder.

To be classified as somatically ill or sick is sometimes associated with certain benefits on the interpersonal level, e.g. support and sympathy from friends and relatives. Many benefits of this type are not consequences of being classified as ill as such, however, but (rather) consequences of getting a specific diagnosis, or of getting some diagnosis or other. For example, to get a diagnosis can sometimes mean that one's suffering is socially recognized and accepted, the world might become a more orderly place, and so on.[9] However, it is doubtful whether these benefits are as extensive in the case of mental disorder as they are in the case of somatic disorder. It is true that we probably give mentally ill people some support and sympathy, and that they are more easily excused from responsibility than others, but in the case of mental disorder, it is not unlikely that the harms outweigh the benefits (at least on the interpersonal level). For example, people who suffer from a mental disorder (as such) are often stigmatized, especially people who suffer from classical mental illnesses like schizophrenia.[10] The stigma that is associated with certain disorders can take different forms, e.g. "intrapsychologically", it can take the form of shame, "interpersonally", it can take the form of harassment or social exclusion, and "institutionally", it can take the form of discrimination.[11]

How we draw the line between mental disorders and other "human problems" also has a number of large-scale social effects. For example, how large a part of the population that is categorized as mentally disturbed has an effect on how much we spend on e.g. psychiatric care, compensation, and medical research.[12] Group interests are also affected by how "mental disorder" is defined, e.g. a broader notion of mental disorder will probably give more power and income to the various mental health professions. (The more problems that are included in its area of expertise, the higher the power and

---

[9] That someone gets a certain diagnosis can also be beneficial to relatives, e.g. it can be beneficial to parents if their troubled child is given a diagnosis like ADHD. The reason for this is partly that it opens up for a somatic explanation ("It wasn't our fault after all!"), and partly that the child gains access to extra resources e.g. in school. I don't think this phenomenon is very common in the mental disorder area, however.

[10] At this point, it is worth noting that conditions that are classified as mental disorders may well be associated with stigma for other reasons than that they are classified in this way. For example, we would most probably not remove the stigma associated with pedophilia if we stopped classifying it as a disorder.

[11] As we have seen above, there are other possible "harms" besides stigmatizarion that are associated with being categorized as mentally ill or getting a psychiatric diagnosis, e.g. that it makes involuntary detainment and compulsory treatment possible.

[12] To simply medicalize certain conditions (i.e. to regard them as medical problems) might have similar effects, regardless of whether or not this medicalization is accompanied by pathologization (i.e. that the problem is not just regarded as a medical problem, but also as a pathological condition, i.e. as a disorder).

status of psychiatry.)[13] And if such a broad concept is connected with the idea that the causes of mental disorders are mainly biological, then the pharmaceutical industry will most probably benefit as well.

Another large-scale social consequence is related to the norm setting and controlling function of medicine that Foucault and others have written about. Medical sociologists often think of medicine as a powerful institution of "social control". This is a much stronger claim than the rather trivial idea that the medical professions exercise *medical control*, and that the pathologization or medicalization of conditions like pregnancy or alcoholism gives rise to medical supervision, monitoring and surveillance, on the one hand, and medical regulation and intervention, on the other. After all, we normally think of disorders as conditions that require medical intervention, and for which medical intervention is appropriate (see point 5 on p. 8 above). The idea that medicine operates as an institution of *social control* seems to include the further claim that the purpose or function of medical interventions is not just to benefit people (e.g. by curing or preventing diseases), but also to "reinforce existing social structures" or to maintain "traditional social values". A standard of normality is established and imposed, and this does not just affect the people who are corrected, but may also have a regulatory and disciplinary effect on the rest of us.

In this context, it is not really necessary to take a stand on whether "the real function" of medicalized discourse is that society (or some powerful group) can control certain individuals. Nor do we have to know whether this controlling function explains why the concept of mental illness was once introduced. We don't have to adopt such conspiratorial views in order to accept the idea that pathologization or medicalization can have these effects (intended or not), i.e. that it can make people more dependent on medical expertise, and that it can set or reinforce certain standards of normality that might, in turn, "reinforce existing social structures".

In short, there are at least three (salient) types of possible harms that are somehow associated with (e.g. dependent on) how "mental disorder" is defined. First, people who are classified as mentally disordered are sometimes stigmatized as a result. Second, the number of conditions included in the category of mental disorder has an effect on how powerful the mental health professions (and the pharmaceutical industry) are, and to what extent people rely on medical expertise rather than e.g. on themselves. And third, ascriptions of mental illness have sometimes been used for purposes of social control,

[13] A favorite example among medical sociologists is the case of pregnancy and childbirth. When these phenomena came to be treated as medical problems in need of "medical regulation and supervision", this made the doctors (the new experts) more powerful, whereas other groups became more powerless and helpless as a result. The midwives became subordinate, and the pregnant women became more dependent on medical expertise, which (supposedly) made them more helpless, viz. by stripping them of the ability to cope with their own problems (cf. Nettleton 1995).

e.g. to justify the use of medical power to impose certain standards of normality, viz. by intervening in socially disapproved behaviour.

So, do these possible harms have anything to do with the purposes of defining the concept of mental disorder? Well, it might be argued that these possible consequences should somehow be taken into account when defining the concept, or more specifically, that we should define the concept in a way that minimizes these harms. For example, one might argue that the concept of mental disorder should be defined in a way that makes abuse of psychotropic medication and involuntary confinement more difficult (cf. Wakefield 2000a, p. 41), or that we should have a concept of disorder that cannot be manipulated by the authorities (cf. Szasz 2000).

Is this a plausible view? For example, to what extent (if any) should we take the possibility of stigmatization into account when we define the concept of mental disorder? Let us first investigate whether there is any truth in the most extreme version of the "harm minimization view", viz. the idea that "[f]or any type of condition X, X is a disorder only if classifying it as a disorder has no significantly harmful effects" (Wilkinson 2000, p. 298).[14]

Wilkinson (2000) argues convincingly that this is an implausible view. He asks us to imagine a world in which there is a widespread extreme and irrational fear of others' ill-health, coupled with the false belief that all unhealthy states are highly infectious. In this world, the standard practice for dealing with illness is to kill those who are ill by burning them, thereby (it is believed) destroying the relevant infection and preventing it from spreading. In this world, classifying a condition as a disorder will almost always significantly harm people with that condition (by causing them to be killed). This does not mean that there are almost no disorders in this world, however (p. 298).

From this, Wilkinson concludes that "[w]hether a condition is a disorder or not does not depend on what consequences classifying it as a disorder would have for those with that condition." (p. 299) However, it can objected that this sort of moral consideration has at least *some* relevance to the issue of whether or not a certain condition should be classified as a disorder.[15]

The sceptical view: "there is no point!"

This concludes the section on why and how it might be of (practical) importance how the concept of mental disorder is defined. Not everyone thinks that it is important how we define this concept, however, so let us now look at the idea that we have little or no reason to care about how we define the concept of disorder. Is this a plausible claim?

---

[14] Wilkinson attributes this view to Kopelman.

[15] For example, it seems plausible to argue that part of the reason why we should avoid a purely evaluative concept of disorder is that such a concept would make it too easy for the authorities to justify the use (or abuse) of medical power.

The strongest argument that can be given to support this sceptical claim is probably the following three-step argument:

(1) The only possible reason why it might be important how we define "mental disorder" is that a well-founded definition can help us make better decisions, e.g. about who is entitled to health care or compensation. That is, we don't need such a definition for theoretical or scientific purposes (cf. pp. 2-3 above), and we can ignore those effects of a definition that are unintended or mediated by implausible norms (cf. Wilkinson's position on p. 12 above).

(2) The concept of mental disorder can never do this normative job alone, however. For example, a person is not entitled to publicly funded health care or compensation *merely* because he suffers from a mental disorder, and a person cannot be sentenced to compulsory psychiatric treatment *merely* because he committed a crime under the influence of a mental disorder. No matter what practical question we have in mind, other conditions must also be satisfied, e.g. the disorder has to be severe enough, the person must be unable to work, the person must be dangerous to others, or the like (cf. pp. 4-9 above).

(3) This suggests that we might as well attack the relevant normative questions directly, without using the concept of mental disorder at all. As Wilkinson (2000) suggests, "answers to questions such as 'What services, or compensation arrangements, should there be for grieving people?' ought not to be determined by the health status of grief, but rather by the needs and the suffering of grieving people." (p. 304) That is, if we want to decide whether people with a certain condition are entitled to health care or not, it is not really necessary to determine whether or not the condition is a disorder. Instead, we should focus on the relevant empirical questions, e.g. how much suffering or disability that is associated with the condition, to what extent it responds to medical interventions, and so on (cf. also Malmgren 1984). The idea that we don't really need a definition of "mental disorder" to make good decisions can also be formulated as follows: There are no plausible normative beliefs which contain an essential reference to the concept of disorder, i.e. none of the "plausible normative beliefs" mentioned on pp. 4-9 above are *really* (or maximally) plausible.

To give further support to (3), it might also be argued that it is positively misleading (that it might even be harmful) to believe that we can find a single concept of mental disorder that will help us deal with *all* the different practical questions on pp. 4-9. On this view, there is no single definition of "mental disorder" that fits all practical purposes listed above, e.g. that can both help us decide who is entitled to health care and who should be excused from criminal responsibility.

This is a good argument, but it can still be argued that it is practically important how the concept of mental disorder is defined. First, it is not likely that we can do without con-

cepts like "disease", "health" or "mental disorder" in all normative contexts. For example, the goals of medicine can to a large extent be formulated in terms of e.g. well-being, suffering, life expectancy, disability and functioning. But to determine e.g. what kind of well-being and functioning that should be promoted, it seems necessary refer to concepts like "health" and "disease", viz. because it is e.g. health-related and/or disease-related functioning that medicine ought to promote. Second, it is important to consider that concepts like "disease", "illness", and "injury" are deeply rooted in our culture, and that people think, feel and act in these terms. This suggests that we should not ignore those effects of a definition that are e.g. unintended or mediated by implausible norms, as it is assumed in step (1) above. It is not impossible that we would do better without these concepts, e.g. that we should abolish the concept of mental disorder totally (like Szasz suggests), but since such a revolutionary conceptual change is not very likely to happen, the best thing we can do is probably to settle for reforms, i.e. to try to influence things for the better by providing better definitions.

Should there be a concept of mental disorder at all?

Some of the practical purposes listed above suggest that we might have good reasons to adopt a rather narrow or restrictive definition, e.g. that we should define the concept in such a way so as to minimize stigma or other harmful effects. This might, in turn, give rise to the idea that we should not just make the class of mental disorder rather narrow, but that we should make it totally empty, i.e. that we should reject the concept altogether. On this radical view, it is always inappropriate and illegitimate to pathologize people's "abnormal behaviour and mental afflictions", i.e. to categorize them as mental disorders. This idea is often accompanied by an even stronger claim, viz. that we shouldn't even medicalize these conditions, i.e. view them as medical problems. So, what's the alternative? If we must not view a certain behaviour or affliction as a disorder or medical problem, how should we view it instead? Here, it is likely that the proponents of the radical view want us to conceive of the relevant conditions as "social problems", "problems in living", or "deviations from social norms" rather than, for example, in terms of madness, crime, sin, or obsession.

Is this a plausible view, or should we stick to the idea that there should be a concept of mental disorder, i.e. that we have good reasons to conceptualize at least some afflictions and behaviours as mental disorders rather than as, say, "problems in living"?

Let us first see whether there are any *conceptual* considerations that can help us decide between the two views. First, we can safely ignore the fact that we actually use the phrase "mental disorder" in our everyday speech. The fact that we actually classify certain conditions as mental disorders does not imply that we ought to do so. This does not mean that all conceptual reasons are irrelevant, however. As Svensson (1990) suggests, "it is through showing that phenomena called mental illnesses are sufficiently alike bod-

ily illnesses (or that they are not) that one can sustain (or refute) the reasonableness of the concept of mental illness." (p. 18)[16] In my view, there are at least some mental afflictions that are sufficiently like somatic disorders to be classified as mental disorders (e.g. schizophrenia or bipolar disorder), i.e. Svensson's conceptual postulate seems to support the common view rather than the radical view.

The next question is whether there are any *"theoretical"* considerations that can help us decide between the two views. For example, are there any conditions that are best conceived of as mental disorders (medical problems) for theoretical reasons, i.e. because this contributes to our understanding of these conditions, or because it helps us to explain these conditions? It is tempting to give an affirmative answer to this question, e.g. because there seem to be at least some mental afflictions that are best explained in biochemical terms. However, if a condition can be explained in this way, it is doubtful whether it can still be regarded as a genuine *mental* disorder (cf. e.g. p. 47 below). If we restrict ourselves to those disturbing conditions that need to be explained in mental rather than biological terms, it is far from certain classifying these conditions as disorders helps us understand them any better.[17] As far as I can tell, there may well be a large number of conditions that are better understood and explained if they are not conceptualized in medical terms. In short, it seems that "theoretical" considerations of this kind cannot really help us decide between the common view and the radical view.[18]

There are also a rather large number of *practical considerations* that are of relevance in this context. These considerations are all reflections on what *consequences* it has to categorize e.g. our "mental afflictions and abnormal behaviours" as mental disorders rather than as e.g. "problems in living", and whether it would have *better* consequences if these conditions were (instead) categorized in some other way, e.g. as social problems, crimes or sins. Now, most of these consequences are already implicit in the list of possible practical purposes of a definition (cf. pp. 4-12 above), so to avoid repetition I'll just offer a brief list designed for this slightly different purpose.

If certain conditions are conceptualized as mental disorders, the most immediate effect is that the people who suffer from these conditions will be categorized as mentally ill or disordered. This can be both beneficial and harmful for the people thus classified. First, there are a number of apparent *benefits* associated with this label, e.g. the person might be entitled to publicly funded health care, reimbursement or compensation, and he might reap some of the advantages that are associated with the sick role, like support

---

[16] He also points out that this "comparison postulate" (as he calls it) is based on "the notion that the concept of 'ordinary' or bodily illness is the more basic, the paradigmatic and the far more well-entrenched concept" (ibid.).

[17] This is not to say that most conditions that are *currently* classified as mental disorders are best understood in psychological terms, however.

[18] In my view, this somewhat "neutral" view is fully consistent with the "anti-theoretical" view expressed on pp. 2-3 above.

and sympathy from others, or reduced responsibility. However, the by far most important benefit is that medicine as an art and science is sometimes good at dealing with the problem, e.g. to actually cure the person, or to reduce his suffering. In short, medical-type interventions might sometimes be the most effective way to deal with the problem.

However, there are also a number of possible disadvantages associated with being classified as mentally ill. We are already familiar with some of these apparent harms, e.g. that attributions of mental disorder can be used to justify involuntary mental hospitalization and compulsory treatment, and that the people who are classified as mentally ill are often stigmatized as a result. Some of these apparent harms might be more beneficial than they seem, however, e.g. compulsory treatment isn't always a bad thing. As far as the issue of stigmatization is concerned, it is true that the concept of mental illness is often a stigmatizing concept (as e.g. Szasz (2000) claims), and that the attribution of mental illness might have dehumanizing and degrading effects. The idea that "a diagnosis of mental illness automatically removes the 'patient' from the class of human beings called 'persons'" (p. 13) is probably somewhat exaggerated, however. It should also be noted that successful attributions of mental disorder might help certain people avoid an even worse kind of stigma, e.g. the horrible treatment they would perhaps get if they were viewed as obsessed or bewitched rather than as ill. This is not to deny that mental illness is sometimes associated with stigma, however, and that it would be better for some people to be viewed as e.g. "deeply troubled" rather than as disordered.

Other possible harms associated with being classified as mentally ill are less apparent. For example, if medicine does *not* hold the legitimate expertise to deal with a certain problem or condition, medical-type interventions are not just ineffective, but they might also be positively damaging. Attributions of mental disorder can also make the individual more helpless or powerless, e.g. by letting him enter the sick role, and thus remove some of his responsibility, or by causing him to rely too much on medical expertise rather than on himself.

We have also seen that classifying certain "mental afflictions and abnormal behaviours" as mental disorders rather than as e.g. "problems in living" has large-scale social effects. For example, how large a part of the population that is categorized as mentally disturbed has an effect on how much resources we spend on psychiatric care, compensation, research, and the like. Moreover, what conditions that are classified as disorders has an effect on how powerful the medical professions are, and to what extent people rely on medical expertise rather than e.g. on themselves. It also serves the interest of the pharmaceutical industry, and it if Szasz (2000) and others are right, it may also serve the interest of the conservative forces in the society at large, viz. by making various social-control measures possible. However, it seems somewhat exaggerated to argue that the primary function of "the mental hospital has always been, and continues to be, the psychiatric segregation and control of socially undesirable persons, typically because they are deemed to pose a 'danger' to the 'health of society.'" (Szasz 2000, p. 11)

Do any of these large-scale effects give us a reason for rejecting the concept of mental disorder altogether? Well, this seems to depend on what the alternative would be. One possible scenario is that far more resources would be spent on various kinds of social work or psychotherapy, and that this would increase the power and status of the social worker, benefit the psychotherapeutic industry, and so on. It is also possible that the people who currently frequent our mental hospitals would be controlled or excluded in other ways, ways which would not necessarily be more humane.

To conclude, it seems to me that as far as the practical considerations are concerned, we should keep our concept of mental disorder. However, it remains to be seen to what extent the conditions currently classified as mental disorders ought to be classified in this way. In any case, I will assume that we should have a concept of mental disorder, and that it is of at least some importance how this concept is defined. But before we turn to the question of how it should be defined, it might be appropriate to reflect on what we want a such definition to be like, i.e. what constitutes a "good" definition of "mental disorder".

Some tentative desiderata for a "good" definition

So, what kind of answer are we looking for, what constitutes a "good" definition of "mental disorder"? Or alternatively put, what desiderata (requirements, or conditions of adequacy) should a definition satisfy, according to what criteria should we assess how "good" or "bad" a certain proposed definition is? Before we take a closer look at the different conditions that a definition of "mental disorder" should ideally satisfy, it is important to note that these conditions determine what kind of arguments that can be given for or against a given analysis of the concept. It should also be noted that some of these desiderata are closely related to why we want a definition in the first place, i.e. to the purposes of a definition, whereas other conditions are more or less independent of these purposes.

In the following, I will assume that our present category of mental disorder is "socially constructed", i.e. that it is a human invention that does not correspond to any natural kind or category. I will also assume that there is no natural kind that even remotely coincides with our present category of disorder, and that there is (for this reason) no such thing as the correct or true definition of the concept.[19] This suggests that we cannot require from a definition that it captures some real or natural category, i.e. that we should look elsewhere for our conditions of adequacy. Given that we want such a definition mainly for certain practical purposes, this is just how it should be. Even if there were a real definition of the concept (i.e. some natural category which has some affinity with our present category), it can be argued that there is no need to look for it.

[19] These constructionist claims will be further elaborated on pp. 42-43 below.

Here are a number of different conditions that a definition of "mental disorder" should ideally satisfy.

1. The ordinary language condition ("descriptive adequacy"). There are two aspects of this condition. First, a definition of "mental disorder" should be consistent with how the phrase is ordinarily used,[20] particularly how it is used (not defined!) by the medical professions, and second, a definition should (at least to some extent) explain our intuitive judgments of disorder and non-disorder (cf. Wakefield 2000a, p. 17).

The idea that a definition should be consistent with ordinary usage has many facets. First, a definition should include those conditions we intuitively regard as pathological, and it should exclude what we intuitively regard as non-pathological, e.g. normal grief, unhappy love, and other "problems in living". In particular, a definition should reflect our generally agreed upon, uncontroversial judgements about what conditions are disorders, e.g. that schizophrenia and major depression are disorders. Second, a definition should be consistent with our current diagnostic systems (which is not to say that the concept of disorder has to be defined in the same type of terms as the different disorders in the plural, e.g. symptomatically). Third, a definition should also be consistent with the fact that attributions of disorder are attempts to partially explain people's behaviour and/or symptoms (cf. Wakefield 1992, p. 377).

Fourth, a definition of "mental disorder" should be consistent with the fact that we tend to regard the category of mental disorder as a sub-category of the more general category disorder (malady, or pathological condition), a category which also includes somatic disorders. This suggests that we want an analysis of "mental disorder" that is a special case of a general theory of disorder, i.e. that our definition should be consistent with such a theory. If we combine this with the idea that "the concept of 'ordinary' or bodily illness is the more basic, the paradigmatic and the far more well-entrenched concept" (Svensson 1990, p. 18), and that the concept of mental disorder is some kind of extension of this concept, we get the requirement that a certain condition cannot be classified as a mental disorder unless it is sufficiently or relevantly similar to the different somatic disorders.[21]

---

[20] At this point, it is worth noting that the term "disorder" is probably not very common in everyday speech (terms like e.g. "malady" or "pathological" might be more common) . This suggests that what we should really require of a definition is (rather) that it is consistent with how terms like "disease", "illness" and "injury" (terms which denote the different sub-categories of disorder) are ordinarily used.

[21] This is what Svensson (1990) calls "the comparison postulate" (cf. pp. 14-15 above). This postulate says that "the feasibility (or non-feasibility) of conceptualizing certain forms of human behaviour and mental afflictions in terms of mental illness must be the outcome of a comparison of these 'mental-illness' problems with 'ordinary' or bodily illnesses. That is, it is through showing that phenomena called mental illnesses are sufficiently alike bodily illnesses (or that they are not) that one can sustain (or refute) the reasonableness of the concept of mental illness." (p. 18)

That is, it is desirable to arrive at a definition which does not classify a condition as a mental disorder unless it is sufficiently like a bodily disorder for the two categories to be subsumed under a common head-category (cf. ibid., pp. 12-13). It is also desirable, however, and this is the fifth point, that the two types of conditions are sufficiently *dissimilar* to motivate the separation of them into two distinguishable sub-categories (ibid., p. 13). That is, a definition of "mental disorder" should help us draw a line between mental and somatic disorders. One may ask whether this is of any practical importance, however, and if so, why. A possible answer to this question is that such a distinction can help us demarcate the area of responsibility for psychiatry.

The idea that a definition should (at least to some extent) *explain* why we think and talk about mental disorders the way we do also has several facets. For example, a definition should (ideally) explain why almost all of us regard conditions like schizophrenia as pathological, whereas we tend to disagree about whether e.g. certain "personality disorders", alcoholism, or learning difficulties should be classified as disorders. A definition should also explain our judgements about severity, e.g. why we conceive of certain disorders as more severe than others. It would also be an advantage if a definition could explain why so many people tend to believe that the category of mental disorder is an objective category, e.g. that it was eventually *discovered* that homosexuality is not really a disorder, or that it has not yet been discovered whether "burnout syndrome" is really a disorder.

2. The value condition. What we think of as disorders are typically undesirable conditions that we think we ought to control and avoid. In particular, we tend to regard a disorder as something bad or harmful for the person who suffers from it. A definition of "mental disorder" should not just be consistent with these facts. Ideally, it should also explain why it is that we regard most or all disorders as e.g. harmful. And if disorders are *necessarily* undesirable, as e.g. Wakefield seems to think,[22] a definition should explain this too, for example by containing an explicit value component. It is worth noting that the value condition is but a special case of the ordinary language condition, but because of its central importance, I've decided to make it a category of its own.

3. The theory condition. The idea that a definition should (to some extent) explain why we regard certain conditions as disorders, or why we think of certain disorders as worse (or more severe) than others, strongly suggests that a definition of "mental disorder" should ideally take the form of a general and coherent conceptual *theory*. A mere list of diagnostic categories will not do. For example, the idea that something is a mental disorder if and only if it is included in e.g. DSM-IV or ICD-10 is not a good definition. What we want is a category of mental disorder that is based on a coherent, explicit set of de-

[22] "To have a disease, an illness, or a disorder is *necessarily* to have a (*prima facie*) negative condition" (Wakefield 2000a, p. 19, my italics).

fining features, i.e. a category which (in psychometric terms) exhibits a high degree of construct validity (cf. Jablensky and Kendell 2002, p. 10). Such a category doesn't just have explanatory value, it can also (if commonly accepted) facilitate communication between different groups or individuals.

4. The precision condition. A definition of "mental disorder" should be sufficiently clear and precise so that there is, *in principle*, no doubt whether or not a certain condition belongs to the category of mental disorder. That is, a definition should draw a sharp conceptual boundary between mental disorders and related non-pathological conditions, and between mental and somatic disorders. Svensson (1990) calls this condition the "conceptual stringency postulate". According to this postulate, "it is desirable, worthwhile and to some extent urgent to strive for conceptual stringency and clarity" (p. 18), and accepting the postulate "simply requires that one should regard a higher degree of conceptual stringency as preferable to a lower degree, and therefore worth striving for" (ibid., p. 19).

It is worth noting that precise definitions and sharp conceptual boundaries sometimes give rise to the idea that there are equally sharp boundaries in nature. We should reject this idea, however. In reality, the boundary between disorders and other conditions are often fuzzy, perhaps especially in the case of personality disorders. This suggests that a "dimensional approach" is sometimes better than a "categorical model", i.e. that it is (at least sometimes) preferable to think of the difference between "normality" and pathology as a matter of degree rather than as a "categorical" (or qualitative) difference.[23]

So, are there any good reasons to adopt a more dimensional approach in this context? Jablensky and Kendell (2002) seem to think so. On their view, "[t]he cardinal disadvantage of the categorical model is its propensity to encourage a 'discrete entity' view of the nature of psychiatric disorders. […] Dimensional models, on the other hand, have the major conceptual advantage of introducing explicitly quantitative variation and graded transition between […] 'normality' and pathology. […] This is important not only in areas of classification where the units of observation are traits. […] There are clear advantages, too, for the diagnosis of 'sub-threshold' conditions such as minor degrees of mood disorder and the specific 'complaints' which constitute the bulk of the mental ill-health seen in primary care settings." (p. 15)[24] Other examples of conditions that seem to exist along a continuum are voyeurism and other "paraphilias" (cf. Culver and Gert 1982, p. 106).

---

[23] My own conceptual theory of health is another example of a dimensional approach. Cf. e.g. Brülde 2000a, 2000b, Brülde and Tengland 2003.

[24] These "sub-threshold" conditions do not just include the cases which barely meet the diagnostic criteria, and that are associated with only mild distress or impairment in functioning. It has also been shown that there are cases (e.g. of depression) that "fail to fulfil the criteria for a disorder in the present diagnostic classifications" but that are nevertheless associated with "significant distress and disability and with clinically significant signs or symptoms" (cf. Üstün et al 2002, pp. 30-31).

The dimensional approach might seem particularly attractive in the case of personality disorder. In the ICD-10, personality disorders are described as ”deeply ingrained and enduring behaviour patterns, manifesting themselves as inflexible responses to a broad range of personal and social situations”, and they represent ”either extreme or significant deviations from the way the average individual in a given culture perceives, thinks, feels, and particularly relates to others.” (The quotation is from Kendell 2002, pp. 110-111.) This suggests that personality disorders are simply abnormal varieties of sane psychic life, i.e. that “[t]he behaviours and attitudes that define personality disorders are probably graded traits present to a lesser degree in many other people” (ibid., p. 112), which, in turn, makes it tempting to adopt a dimensional approach. According to Pilgrim (2002), we should even abandon the concept (i.e. category) of personality disorder altogether.[25]

In short, if this there is any truth in the dimensional approach, this suggests that it is not always desirable to draw a sharp conceptual boundary between e.g. mental disorders and related non-pathological conditions. This is of course incompatible with the precision condition as formulated above, since this formulation presupposes the categorical model. The condition can be formulated in a way that is compatible with the dimensional view, however. On this view, there is a conceptual co-variation between the degree to which someone is disordered and his position in one or several other dimensions. To accept the precision condition for a dimensional definition is simply to require that it is made quite clear what these other dimensions are (cf. Brülde 2000a, 2000b).

5. The reliability condition. A definition should be practically applicable, it should be relatively easy *in practice* (and not just in principle) to determine whether a certain condition belongs to the category of mental disorder (as defined). If this is the case, it is likely that different observers can apply the concept in the same way, i.e. agree on what conditions that should be included in the category and what conditions that should be excluded. It is more likely that a definition will satisfy the reliability condition if the criteria for applying the concept are operational, i.e. if the concept is defined in descriptive terms, and if these terms are, moreover, observational.[26] The presence of a mental disorder can then be established on observational grounds.

Practical applicability is obviously important for communication purposes, i.e. it is likely that a definition that satisfies this condition can (if commonly accepted) improve or facilitate communication between different groups and individuals, both across different settings and cultures.

---

[25] Categorical and dimensional models need not be mutually exclusive, however, e.g. it is also possible to “combine qualitative categories with quantitative trait measurements” (Jablensky and Kendell 2002, p. 16).

[26] Or expressed in psychometric terms: A high degree of reliability seems to presuppose content validity, i.e. that the category has empirical referents (cf. Jablensky and Kendell 2002, p. 10).

6. The simplicity conditions. (a) The class of conditions categorized as mental disorders should be as homogenous as possible. This suggests that we should (other things being equal) prefer a theory which defines "mental disorder" in terms of one criterion only, or in terms of conjunction of different criteria, to a theory which defines "mental disorder" in terms of e.g. a disjunction of different criteria. For example, the idea that all harmful dysfunctions are disorders is more attractive than the idea that all dysfunctions that are either harmful to self or to others are disorders. (This can be regarded as a desire for monism or parsimony.) (b) A definition which does not contain a number of *ad hoc* exceptions or modifications is (other things being equal) preferable to a theory which contains such modifications. For example, the idea that all disabilities caused by mental factors are mental disorders is, according to this condition, more attractive than the idea that only health-related disabilities caused by certain kinds of mental factors are mental disorders. (This can perhaps be regarded as a "desire for unity".)

These six desiderata are all conditions of adequacy in the strict sense, i.e. conditions that are (in part) derived from certain purposes, but which do not by themselves constitute such purposes. We will now turn to some conditions which explicitly appeal to the practical purposes of a definition listed above, e.g. the idea that a definition should help us decide who is entitled to health care or compensation. It should be noted that these conditions imply that a definition can be criticized for not fulfilling these purposes.

7. The condition of normative adequacy. This is the idea that a definition of "mental disorder" should (ideally) help us make better decisions in a number of areas. For example, such a definition should help us determine what services and compensation arrangements there should be for people who suffer from a certain condition; it should help us determine what criminals that should be legally excused or sentenced to psychiatric care; it should help us decide coercive psychiatric treatment is appropriate; it should help us specify the goals of medicine, and to distinguish the special responsibilities of mental health professionals from those of other professionals; and it might perhaps also help us to relate to people with problems in a more appropriate way (on the interpersonal level). However, it is (as we have seen on p. 13 above) doubtful whether there is one single definition that fits all these practical purposes.

8. Other moral considerations. It is possible that we should somehow take it into consideration that it might have harmful consequences for people to be classified as mentally ill, e.g. that they are sometimes stigmatized as a result. For example, it might be argued that the concept should be defined in a way that makes abuse difficult, or that a definition should (ideally) not "open the way to regarding a wide range of purely social disabilities (such as aggressive, uncooperative behaviour or an inability to resist lighting fires or stealing) as mental disorders" (Kendell 2002, p. 112). This seems to suggest that we should exclude as many conditions as possible from the concept of mental disorder

(cf. p. 14 above). The question is whether there is any other way in which a definition can satisfy this condition. As far as I can see, there is only one other possibility, viz. the following one: It seems that this condition might put certain restraints on the evaluative content of the concept, e.g. that it tells us to prefer a less evaluative definition to a more evaluative definition, an explicitly evaluative definition to a definition that is merely implicitly evaluative, and a definition that relies on considerations of harm to a definition that permits to rely on judgements of normality when classifying something as a disorder.

This concludes our list of tentative desiderata for a good definition of "mental disorder". Let us now take a quick look at to what extent these different conditions are in harmony with each other, i.e. to what extent they pull in the same direction. We also have to ask ourselves what we should do if some of the conditions happen to pull in different directions, e.g. what desiderata that should be given most weight, and what kind of conceptual analysis that is most consistent with this choice.

What kind of conceptual analysis does best fit the criteria?

The theory condition, the precision condition, the simplicity conditions, and (to some extent) the reliability condition seem to pull in the same direction. Taken together, these conditions suggest that what we really want is a traditional conceptual analysis in terms of necessary conditions that are jointly sufficient. That is, the best way to satisfy these conditions is most probably to engage in traditional conceptual analysis.

It is not likely that our everyday notion (or "folk concept") of mental disorder can be analyzed in this way, however. Murphy and Woolfolk (2000b) are not the only theorists who reject the assumption that the folk concept of mental disorder is a unitary, coherent concept that can be traditionally defined, i.e. "that there is a consistent set of beliefs that provide necessary and sufficient conditions for analysis of a folk concept", and moreover, "that there currently exists a coherent set of scientific, clinical, or legal beliefs and practices that share a clear understanding of what mental disorders are." (p. 273)

This strongly suggests that there is a tension between the ordinary language condition and some of the other desiderata,[27] and that we have to make some kind of decision about how much weight we should give to the different conditions. One option is of course to allow for the possibility that a definition deviates (to some extent) from ordinary language, which would (in turn) allow for a traditional conceptual analysis of the concept. Another option is to give so much weight to the ordinary language condition that it becomes impossible to give a precise and coherent definition of "mental disorder" in terms of necessary and sufficient conditions. So what if we choose this option, what

[27] The conflict between the ordinary language condition and the simplicity conditions described on pp. 65-66 below is a good example of such a tension.

kind of conceptual analysis would this result in? Apart from the dimensional approach described on p. 20 above (which is not really an option here), what are the alternatives to the traditional categorical analysis in terms of necessary and sufficient conditions.

To simply point out that "mental disorder" is a "family concept" (in Wittgenstein's sense) that connects a number of conditions by "family resemblances" does not constitute much of an analysis. To point this out is, in my view, merely to observe that "[t]here need be no one bunch of things in common – necessary and sufficient conditions – for the same general word […] [e.g. "mental disorder"] to apply to a class of individuals" (Hacking 1995, p. 23), and that "[l]abels often work well without strict necessary and sufficient conditions" (ibid., p. 23).

A somewhat more interesting suggestion (based on this observation) is the idea that we should define "mental disorder" in the same way as some of the diagnostic categories in DSM are defined. We list a number of criteria, and then require that some of these criteria must be met, but not necessarily all, for something to count as a disorder. This does not sound very promising in this context, however.

According to Jablensky and Kendell (2002), this is a kind of "polythetic definition" (as opposed to the traditional strategy, which is "monothetic"), in the sense that members of a class share a large proportion of their properties but do not necessarily agree on the presence of any one property (p. 4) Another example of such a polythetic approach is the prototype-matching approach (ibid., p. 4). Many theorists seem to think that some kind of *prototype analysis* is the most appropriate if we want to capture our everyday concept of mental disorder. For example, "Lilienfeld and Marino (1995) maintain that mental disorder is an ostensive or Roschian concept, implying that the term can only be understood by considering the prototypes of mental disorder." (Kendell 2002, p. 113)

Jablensky and Kendell (2002) describes this "prototype-matching procedure" as follows: "In this approach, a category is represented by its *prototype*, i.e. a fuzzy set comprising the most common features or properties displayed by "typical" members of the category. The features describing the prototype need be neither necessary nor sufficient, but they must provide a theoretical ideal against which real individuals or objects can be evaluated. Statistical procedures can be used to compute for any individual or object how closely they match the ideal type." (p. 4) That is, something is a mental disorder, on this view, if it is "sufficiently similar to the prototypes of mental disorder (schizophrenia and major depression, perhaps)" (Kendell 2002, p. 113).[28]

---

[28] This rudimentary type of prototype analysis is probably sufficient for our purposes, i.e. to define "mental disorder". However, if the purpose is to explain in more detail how our ordinary concept actually works, this rudimentary analysis probably needs to be supplemented by some kind of dimensional (or multidimensional) approach. An example of such an approach is implicit in Hacking's (1995) notion of a *radial concept*. In Hacking's own words: "Theoretical linguists find more structure in classes than mere family resemblance. Each class has best examples […] [i.e. prototypes] and then other examples that radiate away from the best examples. […] Ostriches differ from robins [the prototypical bird] in some ways; pelicans differ from robins in others. We cannot arrange all birds in a single linear order

According to Hacking (1995),

> [t]he idea of a prototype is implicit in psychiatry. […] Prototypes [e.g. the examples given in the DSM Casebook], and radial classes, whether for birds or mental disorders, are not mere supplements to definitions. They are essential to comprehension. One can make a very strong argument, in the philosophy of language, that what people understand by a word is not a definition, but a prototype and the class of examples structurally arranged around the prototype. (p. 24)[29]

In short, what kind of conceptual analysis that is most appropriate seems to depend on what our everyday use of the term "mental disorder" is actually like, and on how central or important we take the ordinary language condition to be, i.e. how much weight we give this condition compared with e.g. the theory condition, the reliability condition, or the simplicity conditions.

Now that the stage is set, let us look at how the concept of mental disorder can and should be defined.

---

of birdiness, saying that pelicans are more birdy than ostriches but less birdy than robins. If we must draw a diagram, it should be a circle or a sphere, with ostriches and pelicans farther away from robins than hawks or and sparrows, but not in one straight line. The class of birds may be thought of as *radial*, with different birds related by different chains of family resemblances, the chains leading in to a central prototype. Likewise for mental illness, individual patients cannot be simply arranged as more 'close to' or 'distant from' standard cases. This is because the ways in which a patient differs from the standard may themselves be structured." (pp. 23-24) It may well be the case that "mental disorder" is a radial concept in this sense, e.g. that it makes little sense to say, of *any two people*, that one is more disordered than the other. These matters of "more" and "less" need not concern us here, however.

[29] Another possible way in which the folk concept of mental disorder might be captured is to "functionally define" the concept by showing how it figures in our folk theory of disorder. Such an analysis has to satisfy certain key "platitudes" about the concept that we all share, and that constrain any analysis of the concept (cf. e.g. Murphy and Woolfolk 2000b, p. 287). It is doubtful whether such a functional definition is an attractive option in this context, however.

Chapter Two. Conceptual theories of mental disorder

The general idea: Disorders as undesirable conditions caused by internal factors

It is often assumed that disorders (maladies like diseases and injuries) are physical or mental states or processes (e.g. underlying anatomical or physiological pathologies or abnormalities) that typically manifest themselves in different kinds of undesirable (e.g. harmful) symptoms. This idea can be spelled out as follows: A condition is a disorder if and only if (a) it is undesirable or bad (either in itself or because of its consequences), and (b) the condition is caused by some type of internal state or process (e.g. a lesion or a part dysfunction), i.e. the cause of the undesirable condition is inside the individual's body or mind.

If we apply this general idea to the mental case, we get the idea that *mental* disorders are undesirable (e.g. harmful) conditions caused by some kind of internal (presumably mental) state or process (e.g. a mental dysfunction). That a person has mental order means (roughly) that there is "something wrong" with his or her mind, and that this has undesirable consequences. This idea can be formulated as follows: A condition is a mental disorder if and only if (a) the condition is undesirable or bad (let us call this the *value component* of the concept), and (b) the condition is caused by some kind of "underlying" mental state or process (let us call this the *factual or explanatory component*).

This rudimentary conceptual theory of mental disorder offers us truth conditions for mental disorder statements, i.e. it tells us under what conditions a mental disorder is present. It also makes a claim about what kind of thing a mental disorder *is*, however, viz. a "condition" that is *caused by* e.g. an underlying dysfunction. This seems to suggest that mental disorders should be located on the level of the organism as a whole, e.g. that they are syndromes, i.e. "dynamic patterns of intercorrelated symptoms and signs that have a characteristic evolution over time." (Jablensky and Kendell 2002, p. 7) This is clearly the view of DSM-IV, where it is claimed that "each of the mental disorders is conceptualized as a clinically significant behavioural or psychological syndrome or pattern that occurs in an individual" (DSM-IV-TR, p. xxxi). It is also claimed that "[w]hatever its original cause, it [this syndrome or pattern] must currently be considered a manifestation of a behavioural, psychological, or biological dysfunction in the individual." (ibid., p. xxxi)

However, it might also be argued that it is the internal cause, rather than its symptomatic manifestations, that *is* the disorder. Suppose that some kind of harmful dysfunction analysis is correct, e.g. that a person has a mental disorder if and only if some mental mechanism fails to perform its function, and this causes him harm on the holistic level. In this case, it might be argued that the disorder should be identified with the part dysfunction itself rather than with the manifestation of this dysfunction, e.g. some clus-

ter of symptoms and signs. A third ontological possibility is to identify the disorder with the whole complex, i.e. to view both the underlying pathology and the symptoms as parts of the disorder. I don't think it is very important which alternative that is chosen, but personally, I tend to adopt the second view in the somatic case, e.g. if there is a harmful dysfunction, then the disorder is identical with the dysfunction (since this allows for the possibility that there are diseases without symptoms). However, in the mental case, it is more tempting to say that the disorder is identical with the syndrome that is caused by the dysfunction (since this implies that we can know that a disorder is present without having identified any underlying pathology).

So, the general view is that a person has a mental disorder if and only if there is an "underlying" (preferably mental) state or process that tends to give rise to undesirable symptoms. This mixed or hybrid view has been challenged in three different ways. First, there is the view that the concept of mental disorder is a purely evaluative concept, i.e. that there is really no need for a factual component. Second, it has been argued that the concept is a purely factual or scientific concept, that the presence of the right kind of internal cause is not just necessary for mental disorder, but also sufficient. And third, it has been suggested that more criteria are needed, i.e. that the value component and the factual component should be supplemented by a third component, or that at least one of these components need to replaced by another type of criterion. Let us look at these three objections, to see if there are any good reasons to reject the mixed view.

The pure value approach

On the radically evaluative view of disorder, the concept of mental disorder is a purely evaluative (or perhaps normative) concept. Wakefield (1992) writes: "The pure value account of disorder asserts that disorder is nothing (or almost nothing) but a value concept, so that social judgments of disorder are nothing but judgments of desirability according to social norms and ideals." (p. 376) This seems to be Szasz's view of how the concept of mental disorder is actually used in medical contexts (but not of how it *should* be used; as we have seen above, he thinks the concept should be abolished altogether). On this view, the term "mental disease" is typically used to refer to undesirable behavioural abnormalities (e.g. misconducts) or undesirable deviations from culturally identifiable social norms (unwanted by the self or others). In short, "what counts as psychopathology is based on a judgment of how the *person ought to function*"(Szasz 2000, p. 9). A possible example of a theorist who has actually adopted this view is Ausubel, who defined "disease" as "any marked deviation, physical, mental or behavioral, from normally desirable standards of structural or functional integrity" (quoted in Wakefield 1992, p. 376).

The most obvious objection to this view is that there are many undesirable mental afflictions and/or behavioural deviations that are *not* considered disorders, e.g. a disorder

is not just *any* undesirable behaviour.[30] It is therefore necessary to draw a line between the afflictions and behaviours that are disorders from the ones that are not, and it is hard to see how this can be done unless we add some factual requirement, e.g. that the cause of the undesirable condition is internal. This suggests that some mixed or hybrid theory is most reasonable, partly because such a theory can account for something that the pure value account fails to account for, viz. the fact that attributions of disorder are attempts to partially explain behaviour and/or symptoms (cf. ibid., p. 377).

The pure scientific approach

On the so-called traditional medical model (or "machine-fault model") of disorder,[31] disorders are "machine-faults", e.g. underlying structural or functional abnormalities. These underlying pathologies are typically accompanied by undesirable signs and symptoms, but on the pure medical model, these undesirable consequences are not defining characteristics of disorder. That is, the presence of a machine-fault is not just necessary for disorder, but also sufficient. If we add the assumption that the presence of a machine-fault can be established in an objective and scientific way, we get the idea that it is possible to state in scientific, objective and value-neutral terms that a certain phenomenon is a disorder-type phenomenon, or that a certain person has a disorder. That is, the idea is that the concept of disorder is a value-neutral concept that doesn't contain any value component (Boorse, who conceives of disease in this way, has a different idea about the concept of illness, however).

It is hard to offer any conclusive arguments against the scientific view at as this point, e.g. before we know more about how central notions like "part dysfunction" have been analyzed. The definitions that have been given have not been very successful, however. Consider the idea that disorders are part dysfunctions, and that a part dysfunction is a statistical functional abnormality that has a detrimental effect on survival or reproduction. This view seems to imply that people who are abnormally intelligent in the statistical sense have a disorder if they live somewhat shorter lives and have less children than the average person, even if they are e.g. much happier than this person. This is just one of many reasons why we should reject the scientific view. Another reason is that the view assumes that the category of mental disorder corresponds to a natural kind (see p. 17 above and pp. 42-43 below), a third reason is that it gives us a category of disorder

[30] That is, the pure value approach fails to satisfy the ordinary language condition. This approach is highly problematic in other respects as well, e.g. a purely evaluative definition is neither precise nor reliable. It might even be suspected that the pure value approach does not satisfy the value condition very well. The reason for this is that it tends to focus on the wrong kind of value, viz. evaluative normality rather than harm.

[31] Originally, this was a model of *disease* rather than of disorder. It is not very problematic to conceive of it as a model of disorder, however.

which is not practically useful, and a fourth reason is that it fails to explain that disorders are almost always undesirable conditions.[32] Many (but perhaps not all) of these objections can be met if we realize that the concept of disorder that best satisfies the desiderata for a good definition is, in part, an evaluative concept.

Definitions related to medical practice

Are there any other possible criteria for mental disorder apart from the evaluative and factual (explanatory) criteria mentioned above? The most common type of suggestion is probably normative criteria connected to medical practice, e.g. the idea that a disorder is, by definition, a condition that health professionals treat, or a condition that (we think) should be treated by health professionals or by medical means.[33] Reznek is an example of someone who claims that suitability for medical treatment is a necessary condition for mental disorder. On his view, something can only be a pathological condition if both (i) it requires medical intervention, and (ii) medical intervention is appropriate.[34]

This is not a good suggestion. First, there are disorders for which there is no existing treatment. If this objection is met by including possible (undiscovered) medical treatment in the criterion, it can be quite hard to interpret the term "possible" in a way that makes the criterion plausible. Second, disorders are not the only problems that we think should be handled by health professionals. There are also a number of other problems that are assigned to the medical professions because of their special skills, e.g. being short, child birth, marital conflicts and occupational problems. In short, there is no necessary connection between a condition's being a disorder and its requiring medical intervention (cf. Wilkinson 2000, p. 300). Third, if e.g. being entitled to treatment is a defining characteristic of disorder, then it is analytically true to say that someone is entitled to treatment *because* he is ill. And fourth, "this approach would paradoxically imply that lack of social concern can eliminate disorder." (Wakefield 1992, p. 377) All these problems could have been avoided if Reznek and others were clearer about the distinction between the content of a definition and the practical purpose of making it (e.g. to help us decide who is entitled to treatment).

---

[32] That is, scientific definitions fail to satisfy the ordinary language condition, the value condition, and the condition of normative adequacy. These definitions also have their strengths, however, e.g. they tend to be sufficiently precise, reliable, and simple.

[33] According to Kendell (2002), it has been suggested in the past (e.g. by Kräupl Taylor) that diseases are simply what doctors treat, but there are, as far as he can tell, "no contemporary advocates for such a simplistic view" (pp. 111-112).

[34] Cf. also Boorse's idea that illnesses are diseases which entitle to treatment and excuse normally criticazable behaviour. But does he really mean that this is so by definition?

To conclude, it seems reasonable to accept the general idea that mental disorders are undesirable conditions caused by internal (presumably mental) states and processes, i.e. that a plausible analysis of the concept should include both an evaluative component and a factual component. This assumption gives rise to two questions: (1) How should the evaluative content of the concept of mental disorder (its value component) be characterized? If mental disorders are by definition bad, then in what way are they bad, according to whom, and so on? (2) What kind of internal cause is essentially involved in mental disorder? How should the factual (or explanatory) component of the concept be characterized?

The next part will be devoted to the first question, whereas the last part of this essay will be devoted to the second question. (This question will also be discussed in more depth in Radovic (forthcoming).)

Chapter Three. The value component. Harm and other bad things

The idea that the concept of mental disorder is "evaluative" or "value laden" is basically a rejection of the idea that we can identify a mental disorder, or specify what a mental disorder is, in an objective way, i.e. without engaging in some form of value judgement. That is, the fundamental idea is that we *somehow* have to rely on value judgements to identify the class of mental disorder, that ascriptions of mental disorder "ineliminably involve values". But how should this idea be further specified, e.g. how should the "evaluative content" of the concept of mental disorder be characterized? There are at least three questions that are of relevance in this context, namely:

(i) What kind of values or evaluations do we have to rely on to identify the class of mental disorder? For example, if mental disorders are by definition bad, then in what way are they bad?

(ii) If attributions of mental disorder essentially involve values, is there any implicit reference to some specific evaluative standard? If so, whose standard? For example, if mental disorders are by definition bad, then according to whom are they bad?

(iii) If we assume that we *somehow* have to rely on value judgements to specify the class of mental disorder, how exactly do these indispensable value judgements enter the picture? In what way is the concept of mental disorder (and the judgements that contain this concept) value laden? Or more specifically, is the concept value laden "in the ontological or definitional sense" or "in the epistemic sense" (cf. Wakefield 2000b, p. 254)?

That the concept of mental disorder is value laden in the ontological or definitional sense means that it is an evaluative concept, i.e. that it has evaluative content in a literal sense, that the "correct definition" of the concept contains an explicit value-component. If the concept of mental disorder is evaluative in this sense, then judgements about mental disorder are a kind of value judgements, and the truth or falsity of these judgments are "dependent of the values that influenced them" (cf. ibid., p. 265). That the concept of mental disorder is value laden in the epistemic sense simply means that the *recognition* of mental disorders relies on value judgments, i.e. that we cannot pick out the class of mental disorders without recourse to values.

Once this distinction is made, it is rather obvious that a concept can be value laden in the epistemic sense without being value laden in the definitional sense. Suppose that we define "mental disorder" in terms of distress and disability, and that our reason for doing so is that we regard distress and disability as bad for the individual. In this case, the concept of mental disorder is not value laden in the definitional sense, i.e. it has no

evaluative content.[35] This would suggest that our disorder judgements are in principle factual, and that the truth or falsity of these judgements are "independent of the values that influenced them", i.e. the idea that distress and disability are bad for us. So the third question is really whether the concept of mental disorder is value laden in the ontological or definitional sense, or whether it is "merely" value laden in the epistemic sense.

In the following, the main focus will be on (i), i.e. the question of what kind of value judgements we have to rely on to identify the class of mental disorder. There are at least three kinds of evaluations (i.e. three kinds of "badness") that might be of relevance when we want to determine whether a certain condition should be regarded as a mental disorder: (1) Judgements about what is bad or harmful for the individual who has the condition. (2) Judgements about what is bad or harmful for others. (3) Judgements about abnormal functioning, e.g. the idea that the person's behaviour deviates from some standard of good or normal functioning (where this standard is not fully derived from or based on considerations of harm). This is probably the category to which attributions of irrationality belong. Since (1) is by far the most common suggestion, we will start our investigation here.

1. Harm

It can hardly be denied that mental disorders typically involve some kind of harm to the individual who has the disorder, e.g. distress or disability. As Wilkinson (2000) puts it, "we have a *prima facie* reason for believing that a condition is a disorder if it is a state of persons which causes them to be harmed (e.g., through death or pain)" (p. 289). This strongly suggests that the connection between disorder and harm is conceptual rather than contingent, i.e. that we rely rather heavily on considerations of harm when we want to determine whether a certain condition should be regarded as a mental disorder.

Many theorists go further than this, and claim that harm to the individual is a necessary condition for disorder, *and* that we don't have to rely on any other kinds of value judgements to delineate the class of mental disorder. For example, Wakefield (1992) claims that a condition cannot be a mental disorder unless this condition "causes some harm or deprivation of benefit to the person as judged by the standards of the person's culture" (p. 385).

The harms that are mentioned in this context (i.e. that are typically associated with disorder) are of three different kinds, viz. (a) displeasure (distress, pain, suffering, loss

[35] The same thing holds if we define the concept in terms of a certain evaluator or certain standards of evaluation, e.g. in terms of what is considered bad by the person who has the condition. In this case, the concept of mental disorder is clearly not value laden in the definitional sense, and it is probably not value laden in the epistemic sense either. This makes me suspect that these accounts are not really (at least not in all cases) definitions at all, but rather attempts to *explain* our disorder judgements in terms of actual evaluations. Cf. the section on constructionism on pp. 42-43 below.

of pleasure); (b) disability (incapacitation, impairment, loss of freedom, decreased ability to act, limitations of social abilities); and (c) significantly increased risk of suffering the harms mentioned in (a) or (b), or of suffering death.[36] All these three kinds of harm are mentioned in DSM-IV's definition of "mental disorder", where it is claimed that a disorder is a

> clinically significant behavioural or psychological syndrome or pattern […] that is associated with present *distress* (e.g., a painful symptom) or *disability* (i.e., 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. (DSM-IV-TR, p. xxxi, my italics)

It is worth noting that this definition differs from Wakefield's definition in that there is no explicit reference to the concept of harm. There are many other definitions that are just like DSM's definition in this regard. In these cases, the concept of disorder is not analyzed in terms of harm, i.e. in evaluative terms, but in terms of distress, disability, or the like, i.e. in descriptive terms. For example, Nordenfelt (1995) defines the concept of malady (an umbrella term for diseases, impairments, injuries and defects) as follows: "*M* is a type of malady in environment *E* if, and only if, *M* is an episode-type which, when instanced in a person *A* in *E*, causes with high probability illness in *A*" (p. 149), where a person is ill if and only if he is, in standard circumstances, disabled from realizing his vital goals. Whitbeck and Fulford also conceive of disorders as inner states or processes which tend to have a detrimental effect on our ability to act, whereas e.g. Spitzer and Reznek also includes a reference to suffering in their definitions.[37]

The claim that "mental disorder" should be analyzed in terms of e.g. distress and disability is really a rejection of the view that the concept is value laden in the definitional sense. That is, if the concept is value laden at all, it is value laden "merely" in the epis-

---

[36] It is worth noting that a condition (e.g. a symptom) can be harmful or bad for an individual in two different ways, viz. finally (bad as an end) or instrumentally (bad as a means). If a condition is harmful as an end (as in the case of suffering or distress), it has a "direct" (a priori, or non-causal) detrimental effect on the person's well-being, and if a condition is harmful as a means (as in the case of disability or incapacitation), the negative effect on the person's well-being is causal rather than "direct". This suggests that the harms in category (a) are "final harms", whereas the harms in categories (b) and (c) are (rather) "instrumental harms". We should also notice that if disorders are conceived of as syndromes (clusters of symptoms), then it is not really appropriate to regard them (as Wakefield does) as conditions that *cause* harm. Even the underlying pathology *constitute* rather than cause harm, since it is a form of instrumental harm.

[37] Notice that Nordenfelt's definition includes a reference to the circumstances under which a condition should be harmful to count as a disorder. This is something that every complete analysis of the concept must do, i.e. one must provide an answer to the question "harmful under what circumstances, in what environment?" For example, is "standard circumstances" the best reply (as Nordenfelt suggests)?

temic sense.[38] Does it matter whether mental disorders are regarded as, by definition, harmful (as in Wakefield's case) or whether they are, instead, regarded as, by definition, associated with e.g. distress and disability? Well, it seems that there are at least two good reasons for choosing the explicitly evaluative alternative, i.e. for not translating harm into distress and disability. First, it is important not to forget that the underlying reason why distress and disability are relevant to disorder is that they are bad. That is, if we believe that the concept is value laden in the epistemic sense, we might as well make it evaluative in the definitional sense, so that we don't loose track of the insight that judgements of harm are fundamental to our judgments about disorder.[39] And second, the list of possible harms due to the right kind of internal cause is (as Wakefield 1992 points out) potentially endless, which suggests that it makes good sense not to restrict ourselves to two or three central kinds of harm (cf. p. 381). However, it can also be argued that we should reject the explicitly evaluative alternative, e.g. because a descriptive definition (in terms distress and disability) is both more precise and more reliable.

Does this mean that we should endorse the view that mental disorders are harmful conditions caused by internal (presumably mental) states and processes, i.e. that harm is necessary for disorder? Before we turn to this question, let us first take a brief look at question (ii) on p. 31 above. Suppose that mental disorders are, by definition, harmful. Is this analysis complete, or do we also have to add something about from whose perspective it is harmful, or according to what evaluative standard?

Digression: Bad according to whom?

It is sometimes claimed that diseases or illnesses are, by definition, something that the afflicted individual himself regards as bad and undesirable (cf. Boorse's definition of "illness"). This suggests that the crucial thing is not so much whether the condition is bad *for* the individual as whether it is bad *according to* this individual, i.e. that he or she evaluates it as bad. Is this a plausible view? I think not. We first have to ask in what way the condition is bad according to the individual. Here, it is rather natural to assume that the relevant kind of badness is harm, i.e. badness-for-the-individual. So, it is reasonable

---

[38] On the assumption that terms like "distress" and "disability" are descriptive, that is. It is not quite certain that the concept of disability is entirely descriptive, however. Consider Culver and Gert's (1982) analysis of the notion of disability: To be disabled is (1) "to lack some ability that is characteristic [normal] of the species at the appropriate level of maturation, such as the ability to walk, talk, or see" (ibid., p. 76), or to have "an extraordinary low degree of that ability" (ibid., p. 77), and (2) this lack of ability is not "due to the lack of some specialized training not naturally provided to all or almost all members of the species" (ibid., pp. 75–76). It is also added that"[a]n overwhelming majority of nondiseased, noninjured members of the species must have had the ability for it to be characteristic of the species" (ibid., p. 77). This suggests that attributions of disability are, at least to some extent, based on evaluative judgements of what is normal and abnormal.

[39] Or in slightly different terms, an explicitly evaluative definition satisfies the value condition to a higher degree than a definition which is merely "implicitly" evaluative, viz. perfectly!

to assume that an internally caused condition can only be a disorder if the individual herself thinks it is bad for her? Suppose that an individual believes that a certain condition is good for her, e.g. when suffering from a manic episode, but that she is mistaken, i.e. that the condition is really harmful for her. In this case, it seems appropriate to say that she is suffering from a disorder. That is, the crucial thing is whether the condition is bad *for the individual*, not whether it is (in some way or other) bad *from her perspective*.

So, is there any reason to refer to any other standard in order to determine whether a certain condition is a disorder, e.g. "the standards of the person's culture" (as Wakefield seems to suggest)? I think not, viz. for the following reasons: Suppose we want to determine whether a certain newly discovered condition is a mental disorder. In this case, we might want to know whether it is really harmful (period), not whether it is harmful according to the standards of the person's culture, e.g. our own culture. This is only relevant when we want to *explain* why certain conditions are *actually* classified as disorders in a certain culture,[40] not when we want to know what conditions that *should* be classified as disorders. This is partly due to the fact that local cultural standards might be just as mistaken as the standards of the individual involved, and if we would come to realize that our own cultural standard is mistaken, it is highly unlikely that we would accept Wakefield's (1992) idea that disorders are conditions that cause "harm or deprivation of benefit to the person as judged by the standards of the person's culture" (p. 384). It is also worth noting that Wakefield's criterion is (in this formulation) a descriptive rather an evaluative criterion: To say that a certain condition is deemed negative by a certain "sociocultural standard" is a factual statement, and to analyze the concept of mental disorder in such terms doesn't just make the concept value-neutral in the definitional sense, but probably also in the epistemic sense. Hence, we should drop all references to perspectives and specific standards of evaluation. The crucial thing is whether a condition is bad (e.g. harmful), and not whether it is bad from a certain perspective or according to a certain standard.

Is harm really necessary for disorder?

Let us now try to find out whether the presence of harm is necessary for disorder. Or more specifically, whether there are any *types* of mental disorder that are either not associated with harm for the individual, or not considered disorders primarily in virtue of being associated with harm. That is, as long as we have manifest harms in mind, the question is not really whether one can find *instances* of disorder that are not associated with harm. (If we think of significant risk for harm as a kind of harm, however, it seems to matter less whether we have types or instances in mind.)[41]

---

[40] Cf. the section on constructionism on pp. 42-43 below.

[41] A question that can't really be discussed at this point is whether harm with the right kind of cause is *sufficient* for disorder. There are obviously many harmful conditions that are not disorders, e.g. "normal

An example of a condition that might not be considered a disorder by virtue of being harmful for the individual is pedophilia. In DSM-IV, this condition is characterized as follows:

> (A) Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger). (B) The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. (C) The person is at least 16 years and at least 5 years older than the child or children in criterion A. (DSM-IV, p. 528)

As it is characterized here, this condition is clearly harmful (due to criterion B). Now, assume that A and C are both satisfied, but not B. Is there really no disorder present in this case? Moreover, even if all three criteria are satisfied, it might be suspected that A and C are the decisive criteria, i.e. that it is not really B that makes the condition a disorder. In fact, it is hard to avoid the impression that criterion B has been added to fit DSM-IV's general definition of "mental disorder" (cf. p. 33 above).[42]

Now, consider the central diagnostic criteria for Antisocial Personality Disorder:

> There is a pervasive pattern of disregard for and violation of the rights of others occurring since age 15 years, as indicated by three (or more) of the following: (1) failure to conform to social norms with respect to lawful behaviours as indicated by repeatedly performing acts that are grounds for arrest; (2) deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure; (3) impulsivity or failure to plan ahead; (4) irritability and aggressiveness, as indicated by repeated physical fights or assaults; (5) reckless disregard for safety of self or others; (6) consistent irresponsibility, as indicated by repeated failure to sustain consistent work behavior or honor financial obligations; (7) lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another (DSM-IV, pp. 649-650, DSM-IV-TR, p. 706).

---

grief" or unrequited love (or infatuation), and the idea is that the difference between these conditions and disorders is that conditions of the latter type have "the right kind of cause", viz. some kind of internal immediate or proximate cause. But before we know how this cause should be characterized, we can't really tell whether this criterion draws the line where we intuitively want it to be, e.g. so that panic attacks count as pathological whereas normal grief does not.

[42] In this context, it is worth noting that criterion (B) is given a new formulation in DSM-IV-TR, namely: "The person has acted on these sexual urges, or the sexual urges or fantasies cause marked distress or interpersonal difficulty." (p. 572) That is, harm (distress or disability) is no longer a necessary condition for pedophilia. The general definition of "mental disorder" remains untouched in DSM-IV-TR, however.

Again, it might be asked if this is this really a disorder in virtue of the fact that it is harmful for the individual. It is true that a person who satisfies these criteria is highly likely to get into different kinds of trouble, but is this really the reason why we tend to think of his condition as a disorder? I think not. It seems more likely that we classify him as disordered because he tends to cause a lot of harm to others.

To conclude, if conditions like pedophilia and antisocial personality disorder are correctly classified as mental disorders, then they are not disorders by virtue of being harmful for the individual who has the condition, but rather because these conditions are abnormal and/or harmful to others.[43] This suggests that we should not draw the line between the pathological and the non-pathological on basis of harm-for-the-individual-evaluations alone, but that we sometimes need to make use of different kinds of evaluations as well. To see if this is a plausible suggestion, let us take a closer look at the two most intuitively plausible candidates, viz. harmful-for-others-judgements and judgements of abnormality (including attributions of irrationality). Do we ever have to make use of these types of evaluative considerations when we attribute a mental disorder to someone, or when we specify the class of mental disorder?

2. Harm for others

Is it ever plausible to regard a condition as a mental disorder because it has bad consequences for others, e.g. because the afflicted individual is dangerous, violent or threatening, or because he acts in a way that is weird or disturbing (due to the right kind of internal causes)? Or more specifically: (a) Do we ever need to resort to value judgements of the form "this is harmful to others" in order to come up with a plausible concept of mental disorder? Is this type of evaluative consideration ever relevant? And (b) is this

[43] Culver and Gert (1982) do not regard this as a good objection to the idea that harm is necessary for disorder. On their view, a pedophilia-type condition can only be a disorder if it is harmful for the individual who has the condition, e.g. if the relevant fantasy or behaviour is "ego-dystonic" (distressful and unpleasant for the person). So, what about those cases where condition (B) is not satisfied, i.e. where the relevant fantasy or behaviour is "ego-syntonic"? Here, they argue that ego-syntonic fantasies are never disorders, no matter how bizarre they are (p. 102). They do admit that some ego-syntonic behaviours are disorders, however, but *not* because they are abnormal and/or harmful to others. In their view, an ego-syntonic sexual behaviour can only qualify as a disorder (i) if it is "not under voluntary control", or (ii) "if it has a high likelihood of being generally known to or discovered by others, and the nearly universal reaction of others would be one of repugnance and revulsion" (p. 103). In my view, this is a rather far-fetched defence of the idea that harm is necessary for disorder. If having sex with very young children is a mental disorder, is it really a disorder by virtue of the fact that it is "universally considered to be repulsive", and that the person is likely to suffer harm as a result? Is it not more reasonable to assume that the reason why it is a disorder is identical with the reason why it is "the target of general repugnance" in the first place, viz. because it is abnormal and/or harmful to others?

type of evaluation ever sufficient as value component, i.e. is harmfulness to others *with the right kind of internal cause* ever sufficient for mental disorder?[44]

(a) This question has already been answered in the affirmative, but here is another argument. According to Szasz and others, the primary criterion for involuntary mental hospitalization is dangerousness, to self and/or others (cf. e.g. Szasz 2000, p. 10) Now, if psychiatric coercion is not just used to reduce the risk of suicide, but also to reduce the risk of e.g. homicide, it would be odd if this is not reflected at all in the concept of disorder. That is, if we can detain a mentally ill person because he is dangerous, it seems that this dangerousness can not be left out of the concept of mental illness altogether.

(b) Suppose that a certain individual is in a condition that makes him dangerous to others, and that this condition has the right kind of internal cause. Is this circumstance ever a sufficient condition for mental disorder? Well, we can't really tell until we know how this cause should be characterized, but my guess is that circumstances of this kind are never really sufficient for disorder. In my tentative view, the dangerousness in question must always be accompanied by some other kind of abnormality to be indicative of mental illness, e.g. the fact that the person behaves violently for no apparent reason (his behaviour is not intelligible) or the fact that his actions are compulsive.

3. Abnormal functioning on the holistic level

Is it ever plausible to regard a condition as a mental disorder because it is abnormal, e.g. highly irrational, or more specifically, because it is associated with abnormal functioning or deviant behaviour (assuming that the condition has the right kind of internal cause)? Or more specifically: (a) Do we ever need to resort to value judgements of the form "this behaviour or response is a marked and undesirable deviation from the normal" in order to come up with a plausible concept of mental disorder? Are evaluative considerations of this type ever relevant? And (b) do attributions of abnormality ever constitute the full evaluative content of the concept of disorder, i.e. is abnormality with the right kind of internal cause ever sufficient for mental disorder? We might also ask (c) whether some kind of abnormality is necessary for disorder.

Before we take a closer look at these questions, there are a number of things that need to be clarified. For example, what do people like Szasz or Ausubel mean when they say that the term "mental disorder" actually refers (or should refer) to mental or behavioural deviations from some norm or standard? For example, what kind of norm or judgement does Szasz (2000) have in mind when he claims that "what counts as psychopathology is based on a judgment of how the *person ought to function*" (p. 9)?[45]

---

[44] It might also be asked whether harmfulness to others is necessary for disorder, but this question is obviously too stupid to be taken seriously.

[45] This evaluative view must of course be distinguished from the idea that mental disorders are deviations from *the statistically normal* (on the holistic or behavioural level). The latter view is quite implau-

In Szasz's case, the relevant norms are, it seems, primarily moral or legal norms. On this view, the behavioural "abnormalities" that are regarded as mental disorders are often deviations towards the immoral or the illegal (i.e. misconducts). As I see it, other kinds of norms might also be relevant, however. For example, the idea that it is pathological to totally ignore one's own appearance might be based on an aesthetic norm, whereas the idea that it is pathological to grieve intensely for years and years after a loss has occurred can perhaps be seen as an attribution of irrationality. And as Wakefield (2000a) points out, irrationality is sometimes viewed as "the hallmark of mental disturbance", i.e. it is rather commonly believed that "mental disorder often consists of a breakdown in the capacity for rational thought or action" (p. 18).

It is important to note that for this view to make any sense at all, we have to assume that the relevant standards are not ideals (standards of good or perfect functioning), but "standards of normality". We also have to assume that the deviation from the norm is both marked and negative, since no one would regard a slight or positive deviation as pathological. Moreover, the idea that a condition can qualify as pathological by virtue of its abnormality can hardly constitute an alternative to the two harmfulness views unless we assume that the relevant standard of normal functioning is not based on considerations of harm alone.[46] Let us now return to the three questions:

(a) There is no doubt that our actual disorder judgements can often be *explained* in terms of what we find normal and abnormal.[47] But is this how it should be, i.e. is it sometimes necessary to make use of abnormality judgements to specify the class of mental disorder in a *plausible* way? I think so. To return to the example in (b) on p. 38 above,

---

sible, by the way, e.g. because it implies that positive statistical deviations like excellence in strength or intelligence are disorders. It is also important to distinguish the view that disorders are deviations on the holistic level from the view that disorders are (by definition) caused by deviations on the "part level", e.g. anatomical (structural) or physiological (functional) abnormalities. We will return to the latter idea on p. 50 and pp. 58-59.

[46] For example, it might be argued that irrational beliefs can only be indicative of mental disorder when they are associated with a lack in the "ability to believe", i.e. that irrationality is, in this context, best regarded as a disability (i.e. as a kind of harm) rather than as a mere abnormality. This view is adopted by Culver and Gert (1982), who suggest that "irrational beliefs are symptoms of mental maladies because they count as the lack of the ability to believe. What is wrong with having psychotic delusions, such as paranoid delusions, is not that they are false but rather that they show the person does not have the ability to believe; he does not respond to the overwhelming evidence in a way that is appropriate for someone with his knowledge and intelligence." (p. 113)

[47] For example, the fact that certain sexual conditions (e.g. "paraphilias" like fetishism, transvestism, or voyeurism) are classified as mental disorders is probably best explained in this way. According to Culver and Gert (1982), this is the primary reason why DSM-III lists some of the paraphilias as mental disorders (even though this is not explicitly stated). For example, in cases like transvestism, "the defining criteria for the disorder are whether some deviant fantasies and/or behaviours are present and whether heterosexual activity is weak or absent" (p. 106). On Culver and Gert's view, this suggests that DSM has adopted "the psychological theory that heterosexuality (fantasy and behavior) between consenting adults is the ideal and that except for ego-syntonic homosexuality the absence of heterosexuality is ipso facto a mental disorder" (ibid., p. 105).

a highly dangerous individual can (I think) only be classified as mentally ill if some other kind of abnormality is present, e.g. if his behaviour is, to a high degree, irrational or unintelligible.

(b) Is abnormality (with the right kind of internal cause) ever a sufficient condition for disorder? Let us first note that there are many behavioural deviations that are not regarded disorders, e.g. certain kinds of criminal or morally repugnant behaviour. It is hard to tell whether these behaviours are caused by the right kind of internal factors, however. So, let's reconsider the case of pedophilia, assuming that this condition has the right kind of cause, and that it is, in part, constituted by abnormal sexual preferences. Is the presence of this abnormality sufficient to make the condition pathological? I think not. In my view, the abnormality must be associated with some kind of harm to be indicative of mental disorder. That is, that a condition is abnormal is not sufficient to make it pathological, it is also necessary that it is harmful to the individual or to others.[48] The main reason why we should reject the idea that deviance (with the right kind of cause) can be a sufficient condition for mental disorder is practical.[49]

(c) Is the presence of an abnormality necessary for mental disorder? DSM-IV's definition seems to imply that it is, viz. when it states that a condition (in this case a syndrome or pattern) should not be classified as a disorder if it is "merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one" (DSM-IV-TR, p. xxxi). Assuming that this can be interpreted as a criterion of disorder (and not merely as an easily applicable rule of thumb that can help us determine if something is a disorder), is it plausible? Let us first note that many physical disorders are normal or natural reactions or responses to traumas or life experiences, e.g. influenza, measles, wounds, and burns. This suggests that it is both arbitrary and unjustified to exclude all expectable and culturally sanctioned (i.e. normal) responses from the category of mental disorder. Posttraumatic stress disorder is a possible counterexample, and if Wilkinson (2000) is right, "normal grief" (conceived of as a kind of mental injury) is another.[50]

---

[48] That a condition is both abnormal and harmful does not guarantee that it is a disorder, however, e.g. as in the case of extreme ignorance. But this may well be due to the fact that the condition has the wrong kind of (internal) cause.

[49] For example, "[t]he tendency to regard deviance as a sufficient feature of sexual disorders leaves psychiatry open to the justified criticism that deviance per se, sexual or political, has become closely linked in psychiatric thought with sickness" (Culver and Gert 1982, p. 107). The reason why it is undesirable to pathologize deviances in this way is clearly practical. Cf. also footnote 47 above.

[50] The idea that a disorder is never an "expectable and culturally sanctioned response" can also be regarded as an operationalization of the term "dysfunction" (rather than as a normality criterion). This is how Wakefield (1992) conceives of the idea. On his view, it is a bad operationalization, since it fails to capture the dysfunction requirement that inspired it (p. 381). This is how the dysfunction requirement is formulated in DSM-IV: "Whatever its original cause, it [a disorder] must currently be considered a manifestation of a behavioural, psychological, or biological dysfunction in the individual. Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual

In short, it is sometimes necessary to rely on abnormality judgements to specify the class of mental disorder in a plausible way. However, abnormality (with the right kind of internal cause) is neither necessary nor sufficient for disorder.

Conclusions

We can now summarize the chapter on mental disorder and value as follows:

(A) Mental disorders are necessarily undesirable conditions, but this does not imply that "mental disorder" is an evaluative term. The idea that disorders are bad might also function as a condition of adequacy (the value condition), which explains why we use (and should use) evaluations to identify the class of mental disorders, i.e. that the concept is value laden in the epistemic sense. However, it seems plausible to conceive of the concept as value laden in the definitional sense as well, i.e. to define "mental disorder" in explicitly evaluative terms.

(B) The evaluations we have to rely on to distinguish the pathological from the non-pathological are mainly of the form "this condition is harmful for the person", but the presence of harm is not necessary for disorder. If we want to specify the class of mental disorder in a way that is consistent with ordinary language, we need to make use of other kinds of evaluations as well, e.g. (i) judgements about what is harmful for others, and (ii) judgements about abnormal functioning, e.g. the idea that the person's behaviour deviates substantially from some standard of normal functioning (where this standard is not fully derived from or based on considerations of harm).

(C) The idea that several kinds of evaluations are relevant in this context is more in line with ordinary language than the idea that considerations of harm are the only relevant kind of evaluation, and a definition based on this idea might also be normatively adequate to a higher degree. However, the broader view gives us a less coherent concept of mental disorder, partly because harmfulness to others or deviating from some standard of functioning is neither necessary nor sufficient for disorder, and partly because these evaluations tend to be used in an *ad hoc* manner. This suggests that if we accept the broader view, we have to give up the idea that "mental disorder" can be defined in terms of necessary conditions that are jointly sufficient, e.g. that we have to engage in some sort of family resemblance analysis instead.

The more narrow idea that harm is necessary for mental disorder, and that we do not have to rely on any other evaluations to distinguish the pathological from the non-pathological, has other advantages, e.g. it gives us an evaluative content of "mental dis-

---

and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above." (DSM-IV-TR, p. xxxi) We will return to this criterion below.

order" that satisfies the simplicity conditions and the reliability condition to a higher degree. In short, as far as the evaluative content of a definition is concerned, there seems to be quite a tension between some of the desiderata listed on pp. 18-23 above, e.g. the ordinary language condition and the condition of normative adequacy, on the one hand, and e.g. the reliability and simplicity conditions, on the other.

(D) There is no need to refer to any particular perspectives or specific standards of evaluation – e.g. what the individual himself evaluates as bad and undesirable, or what is undesirable "as judged by the standards of the person's culture" – when defining the concept of mental disorder. The crucial thing is whether a condition is bad, not whether it is bad from a particular perspective.

This is all I have to say about the value component viewed in isolation. To further determine how well e.g. the harm criterion captures the concept of mental disorder, it is necessary to combine it with the idea that the harm in question must have a certain kind of internal cause. But in order to do this, we have to get a better grasp of what the relevant kind of internal cause is. But before we turn to this question, let us just take a brief look at the connection between the idea that "mental disorder" is an evaluative concept, on the one hand, and the idea that the category of mental disorder is socially constructed, on the other. How are these two ideas related to each other?

Value-ladenness and social constructionism

It is sometimes argued that the concept of mental disorder a social construction. Is this a plausible view? And how is it related to the idea that the concept is value-laden?

To be a social constructionist concerning mental disorder means (roughly) that one regards our present category of disorder as socially constructed, i.e. as a human invention or fabrication. This is an empirical claim that can be divided into the following three parts:

> (i) Our present category of mental disorder is created by us rather than "by the world". It is not a "natural category" that reflects the structure of the world itself, and we have not discovered it through careful observation of the world "as it is".
>
> (ii) Our present category of disorder (which might look "natural" or "unavoidable" to some) could well have been different, it is both contingent and arbitrary.
>
> (iii) The idea that the category is a human invention is often specified in terms of how the category has come into existence, and why it continues to exist once it has come into existence (why it is stable over time). Here it is often argued that the category is not a scientific discovery, but that it is (to a considerable extent) determined

by extra-theoretical or extra-scientific factors. For example, it might be argued that our present category of disorder is (in part) an outcome of "social and political struggles", or that we need to explain the category in terms of group interests or doctor-patient interactions. It seems highly likely that these extra-theoretical or extra-scientific factors involve evaluations. (Cf. p. 39, esp. footnote 47. Cf. also p. 36.)

However, some constructionists go further than this, and make the following metaphysical claim (called nominalism):

(iv) There is no correct or true definition of "mental disorder". There is no such thing as an objective distinction between the pathological and the non-pathological that can be discovered e.g. by scientific means (this is not how reality is structured). Either there is no sharp distinction between reality and our conceptualizations of it, or if there is an external reality that can be distinguished from our conceptualizations of it, there is no such thing as the single most valid conceptualization of it. And even if it happened to be the case that there is some overlap between our present category of disorder and certain naturally given categories, it is highly unlikely that any of these categories even remotely coincide with our category of mental disorder, which suggests that there is no such thing as a natural kind which can correctly be labelled "mental disorder".

The constructionist view is clearly a plausible view of our present category of mental disorder.[51] This category does not correspond to any natural kind, it has (to a considerable degree) been shaped by evaluative considerations, and so on. It also seems plausible to assume that nominalism is true in this area, and that there is no natural kind that even remotely coincides with our present category of mental disorder.

There is obviously an intimate connection between the idea that the category of mental disorder is socially constructed, on the one hand, and the idea that "mental disorder" is an evaluative concept, on the other. So how should this connection be characterized? Well, constructionism with regard to the concept of mental disorder (an empirical thesis) does not imply that this concept is value laden (a conceptual thesis). However, the idea that the concept of mental disorder is actually evaluative seems to provide the best explanation of why constructionism is true in this area. Moreover, the idea that this is the way it should be (given the purposes of a definition), i.e. that the concept *should* be evaluative, is probably part of the best explanation of why nominalism is true in this area.

---

[51] It is worth noting that this does not in any way imply that all or most of our present diagnostic categories are social constructions. However, the part of reality that can be subsumed under our concept of disorder is (in spite of some clustering) probably not very well structured at all, i.e. it might be hard or impossible to discover objectively valid diagnostic categories in this area. Cf. Brülde and Tengland 2003, chapter 6.

Chapter Four. The factual component. "Machine faults" and other internal causes

So, the general idea is not just that a mental disorder is an undesirable condition, but also that it is (by definition) caused by the right kind of internal factor. The concept of mental disorder is not a purely evaluative concept, it also contains a factual (or explanatory) component. But how should this factual component be characterized, what kind of internal cause is "essentially" involved in mental disorder? There are at least six possibilities here,[52] viz. the following ones:

(1) The idea that disorders are (or are by definition caused by) *lesions*, i.e. structural or anatomical abnormalities. On this view, a person can only have a disorder if some structural part of the organism (like a cell, a tissue, or an organ) is "damaged".

(2) The idea that disorders are (or are caused by) *part dysfunctions*, e.g. physiological or biochemical abnormalities or disturbances that may or may not be caused by anatomical abnormalities. There are several different versions of this view, depending on what conception of dysfunction that is taken for granted.

These two ideas are both versions of the so-called *traditional medical model*, or "machine-fault model", of disorder. On this view, disorders (e.g. diseases) are regarded as "machine-faults", i.e. underlying abnormalities (structural or functional) that tend to cause problems on the holistic level. As I see it, this view has two central features: (i) The "reductionist" idea that the symptoms are caused by some kind of "underlying" or "lower-level" pathology. The machine as a whole does not function because there is a fault somewhere in the apparatus, and this is how disorders must (by definition) be explained. (ii) The view that these machine-faults are abnormalities, i.e. deviations from normal structure or function, where normality is understood in terms of species design.

One reason why it might be tempting to define "mental disorder" in machine-fault terms is probably the phenomenon of *clustering*, i.e. the fact that symptoms often tend to cluster into syndromes.[53] In somatic medicine, many syndromes have been explained in terms of machine-faults, and this gives us a strong prima facie reason to suspect that

---

[52] These six possibilities are numbered 1, 2a, 2b, 3, 4, and 5.

[53] As Jablensky and Kendell (2002) point out: "Although the range and number of possible aetiological factors–genetic, toxic, metabolic, or experiential–that may give rise to psychiatric disorders is practically unlimited, the range of psychopathological syndromes is limited. The paranoid syndrome, the obsessive-compulsive syndrome, the depressive syndrome–to mention just a few major symptom clusters–occur with impressive regularity in different individuals and settings, although in each case their presentation is imprinted by personality and cultural differences." (p. 7)

there is such a thing as an underlying pathology or morphology (as opposed to aetiology) in the mental case as well. How else could we explain clustering?

Now, as Svensson (1990) points out, there are two fundamentally different versions of the "machine-fault model", viz. "the somaticist approach" and the genuine-mental-disease approach.

> (a) The somaticist view is "the view that the internal pathology in cases of 'mental disease' is *exactly* of the same type as in cases of physical disease, which is to say that 'mental disease' pathology, just like other pathology, is assumed to consist of structural or functional abnormalities of the human organism" (p. 86). On this view, the essential internal factor is some kind of neurophysiological or neuroanatomical disturbance, or some biochemical abnormality in the brain.
>
> (b) On the genuine-mental-disease view, "the abnormalities in cases of 'mental disease' are not, and need not be, of an organic nature for them to be correctly conceptualized in terms of disease." (ibid., p. 86). The relevant internal cause ("machine-fault") is located in the "mental apparatus" rather than in the "organic apparatus". It is the "psychic machinery" that has broken down, which implies that there is such a thing as mental disease *sui generis*, i.e. purely mental disease.

It is rather obvious that the lesion view is necessarily somaticist, i.e. that it is not really compatible with the genuine-mental-disease view. The reason for this is that there are no structural abnormalities in the mental realm. There is no mental structure or mental anatomy in any literal sense, the mind does not consist of parts which can be damaged.

Both the somaticist view and the genuine-mental-disease view can be combined with the dysfunction view, however, i.e. there are two possible versions of the dysfunction view: (2a) On the somaticist dysfunction view, mental disorders are by definition caused by physiological or biochemical abnormalities or disturbances. (2b) On the genuine-mental-disease version of the dysfunction view, mental disorders are by definition caused by underlying *mental dysfunctions*.

These three versions of the traditional medical model – i.e. (1), (2a), and (2b) – all purport to provide a necessary condition of mental disorder. It is normally assumed that this necessary condition is an objective criterion, i.e. that it can be objectively (e.g. scientifically) determined whether the condition is satisfied or not.[54] However, there are also

> (3) alternative dysfunction views which are not traditional medical views, and which do not make this assumption. These views assume instead that the relevant

[54] The objective criterion provided by the traditional medical model (e.g. part dysfunction) is always regarded as a necessary condition by the proponents of this model. Some of these proponents (e.g. Boorse) also regards the criterion as sufficient, however, whereas others (e.g. Wakefield) do not.

notion of dysfunction is value laden, at least in the epistemic sense (cf. pp. 31-32 above).

The last two suggestions both presuppose that some kind of harm is necessary for mental disorder:[55]

(4) the idea that the harmful condition (e.g. disability) has its immediate cause within the individual's mind, i.e. that it is caused by some mental state or process. Or more specifically, a mental disorder is a condition that belongs to a *type* that is normally harmful, and where conditions of this *type* are caused by internal mental episodes of another *type*. Or alternatively put, the relevant mental episode is of a *type* such that most or all instances of this *type* cause harm (e.g. disability). On this view, it is not possible to give any further specification of the kind of internal cause that is present when someone has a mental disorder.

(5) The idea that mental disorders are harmful conditions which are not due to any "distinct sustaining cause". This view is a version of the general idea that disorders are internally caused harmful conditions, combined with an analysis of the phrase "internally caused", where a condition is regarded as internally caused if there is an absence of a distinct sustaining cause.

Let us now investigate how plausible these suggestions are. The main question that will be asked in connection with each suggestion is whether it provides a plausible view of what kind of internal cause that is *necessary* for mental disorder. However, it is also important to find out, each suggestion, whether it provides a sufficient condition for disorder *when combined with the undesirability criterion* (e.g. the harm criterion). For example, the critical assessment of the dysfunction view will be based on the following two questions: (i) Is dysfunction necessary for disorder? (ii) Is undesirable (e.g. harmful) dysfunction sufficient for disorder?[56] In connection with this, it will also be asked whether the dysfunction view gives rise to a category of disorder that is too wide (too inclusive) or too narrow (too exclusive).

---

[55] That is, no one has ever suggested that the kind of internal causes that (4) and (5) refer to as essential are *sufficient* for disorder.

[56] In some of the cases – viz. (1), (2) and possibly (3) – it might also be of some interest to ask whether the proposed factual criterion is (in itself) sufficient for mental disorder, e.g. whether dysfunction (with or without harm) is sufficient for disorder. (Note that this question does not arise in connection with suggestions (4) and (5), where some kind of undesirability criterion has been incorporated from the very beginning.) I have already suggested (cf. the section on the pure scientific approach on pp. 28-29) that some kind of value criterion is necessary for disorder, but since there are theorists that have claimed that dysfunction alone is sufficient (e.g. Boorse), I will not avoid the question altogether.

1. The lesion view

On this view, disorders are (or are caused by) lesions, i.e. structural abnormalities or disturbances in organs, tissues, or cells. According to Szasz (2000), this is how the concept of disease is actually used in medicine: "Literally, the term *disease* denotes a demonstrable *lesion* of cells, tissues, or organs" (p. 4). However, Szasz does not deny that diseases involve functional abnormalities (disturbed bodily function) as well:

> [T]he *medical* concept of disease […] denotes a *bodily abnormality or somatic pathology*, that is, a physico-chemical–anatomical or physiological, structural or functional–*alteration of the body deemed to be undesirable*. (ibid., p. 4)

He seems to think that all functional disturbances are caused by lesions, though, as when he claims that "the basis of all disease is cellular pathology".

One reason why this idea is of no interest in this context is that it strongly suggests, by virtue of being a somaticist approach, that there is no such thing as mental disorder. Suppose that it is discovered that a condition like schizophrenia is really caused by some kind of lesion in the brain. In this case, schizophrenia would probably be regarded as a somatic disease rather than as a mental disease. This suggests that if the lesion view were true, there would be no mental disorders left, only somatic disorders. Or as Svensson (1990) puts it, it would, on this view, be hard to tell how, and why, mental disease should be distinguished from physical disease. If the disease that causes behavioural or experiential abnormalities is organic by nature, what is mental about it? (p. 87) That is, "the somaticist view seems to run the risk of making redundant the notion of mental disease. It seems difficult to make clear what is mental about 'mental disease' if the pathology of it is a physical pathology." (ibid., p. 117)

Is this likely to happen? It doesn't seem so. For example, Maj et al (2002) points out

> that [t]hirty years of biological research have not been able to identify a specific biological marker for any of the current diagnostic categories (and genetic research is now providing evidence for the possible existence of vulnerability loci which are common to schizophrenia and bipolar disorder). (p. ix)

And according to Jablensky and Kendell (2002), Alzheimer's disease is one of the few conditions appearing in psychiatric classifications where a specific brain morphology and a "tentative pathophysiology" have been identified. "Schizophrenia, however, is still better described as a syndrome" (p. 6). But regardless of what the future will bring, it is obvious that all somaticist views suffer from another weakness, namely that it implies that as long as we haven't found any underlying pathophysiology, we can't really know whether a mental disorder is present. This seems rather counterintuitive, espe-

cially if we consider our present level of ignorance (about underlying pathophysiologies).[57]

The lesion view is not plausible in the somatic case either. As Wakefield (1992) points out, the existence of a lesion is neither necessary nor sufficient for disorder. "There are physical disorders […] for which there are no known anatomical lesions", and "a lesion can be a harmless abnormality that is not a disorder" (p. 375).[58] For a lesion to constitute (or contribute to) a disorder, it needs to have a negative effect on function, which in turn has harmful effects on the organism as a whole. This strongly suggests that we should specify the factual component of the concept of disorder in functional rather than structural terms, i.e. in terms of dysfunction.[59]

2. Disorder as "part" dysfunction (or harmful dysfunction)

On the dysfunction view, disorders are, by definition, caused by "part dysfunctions". That is, the underlying abnormality ("machine-fault", pathology, or deviation from species design) is functional rather than structural, which is not to deny that the reduced function is, in most cases, due to some structural abnormality.

The part dysfunctions that are essentially involved in mental disorder are, it seems, either biological (e.g. biochemical or neurophysiological) or mental (as in the case of a genuine underlying psychopathology). This suggests that the relevant dysfunctions cannot really be behavioural, as is claimed in DSM-IV.[60] Such a claim is not quite in harmony with the central idea that a disorder can only be present if something has gone wrong with the organism's internal functioning, and that this internal dysfunction *ex-*

---

[57] On the assumption that at least some of the conditions that are currently regarded as pathological are reasonably regarded as such, that is. At this point, it is worth pointing out that the lesion view can, if combined with the fact that very few "pathophysiologies" have been identified, give some support to the idea that all or most of the conditions that are currently regarded as mental disorders are not really disorders at all. This is probably one reason why the lesion view is so attractive to an anti-psychiatrist like Szasz.

[58] A definition that combines the lesion view with some kind of undesirability criterion doesn't just fail to satisfy the ordinary language condition; it is also lacking in normative adequacy. However, a definition of this kind would satisfy most of the other conditions of adequacy quite well, including the value condition (a condition that would not be satisfied if the existence of a lesion were, instead, regarded as sufficient for disorder).

[59] Wakefield (1992) also criticizes Szasz's idea of how lesions can be identified, viz. by statistical means. The recognition of a lesion is not simply a matter of observing anatomical deviance. To recognize "a statistical deviation from a typical anatomical structure" as a lesion is also to recognize that this deviation impairs the ability of the particular structure to accomplish the functions that it was designed to perform ("part dysfunction"). Again, it is concluded that the concept of lesion is irrelevant unless it is conceived of as a functional concept.

[60] DSM-IV's definition of "mental disorder" contains the idea that "[w]hatever its original cause, it [a disorder] must currently be considered a manifestation of a *behavioral*, psychological, or biological dysfunction in the individual." (DSM-IV-TR, p. xxxi, my italics)

plains the symptoms that can be observed on the level of the organism as a whole (e.g. the behavioural level). We should not forget that the dysfunction view is a "reductionist" view, which assumes that the machine as a whole does not function because there is a functional fault somewhere in the apparatus. As Wakefield (2000a) points out,

> *[f]unction* and *dysfunction* in the medical sense refer, in the first instance, not to the quality of the person's performance in a given environment but to whether the mechanisms within the person are performing or failing to perform the functions they were designed to perform. (p. 20)

It is also important to note that the term "part" is "used here in a very broad sense to encompass hierarchically organized levels of physical and mental structure and function. For example, cells and organs are parts, but systems of interacting organs (e.g., the digestive system) are also parts that have functions corresponding to their level of description." (ibid., p. 25)

Another important feature of most dysfunction views is the idea that abnormal part functioning needs to be specified in terms of deviations from species design. Every species has a natural design, a species-typical design, which comprises the hierarchy of interlocking functional systems that supports the life of organisms of that type. As Svensson (1990) formulates this view,

> a [somatic] disease [dysfunction] is said to be present when some part of the body fails to perform its […] species-typical contribution to the life-processes of the organism. […] There is a 'natural order', a plan of organization developed through natural evolution, of the human organism, and diseases [dysfunctions] are deviations from this 'species-design' (p. 93).

Svensson also adds that ascriptions of dysfunctions are, on this view, not regarded as evaluative, but as objective statement of facts (ibid., p. 93).

In short, part dysfunctions are underlying functional abnormalities, internal functional deviations from species-design, which can explain e.g. why a person is not functioning well on the holistic level. This is not very informative, however, i.e. it seems that the notion of dysfunction is in need of further clarification. The central question is of course what (exactly) a dysfunction is, i.e. how the concept of dysfunction should be defined. However, we also need to know how dysfunctions should be identified, i.e. how we can and should determine whether a dysfunction is present (assuming that it might not be possible to apply the definition in a straight-forward way; cf. the reliability condition).

Two medical conceptions of dysfunction

There are at least two traditional medical views on what a dysfunction is, and on how we should determine whether a dysfunction is present, views which (if accepted) give rise to two different conceptions of disorder (somatic as well as mental disorder).

(i) On the first of these two views, the concept of dysfunction is analyzed in terms of *biological disadvantage*, where the notion of biological disadvantage is understood in evolutionary terms, viz. in terms of reduced survival or lowered reproductive fitness.[61] On this view, a biological function of a part or process is its ultimate contribution to survival and reproduction, e.g. that a part of an organism functions normally means that it contributes to survival and reproduction (reproductive fitness) with at least typical efficiency (cf. Wakefield 1992). A dysfunction is present if a part or process does not contribute survival or reproductive fitness in this way. If this notion of dysfunction is incorporated into the concept of disorder, we get the idea that a disorder is (roughly) a condition that reduces longevity or fertility.

This view is normally accompanied by the idea that dysfunctions are (as a rule) *statistical abnormalities*, e.g. that it can be established statistically whether a certain part performs its function or not. Now, it is not quite clear whether this should be understood as part of the definition of "dysfunction", or "merely" as a highly useful (and perhaps necessary) way to determine whether a dysfunction is present (cf. the distinction between the definitional and epistemic senses of "value laden" on pp. 31-32 above). That is, it is not clear whether dysfunctions are, by definition, statistical functional abnormalities that have detrimental effects on survival or reproduction, or whether dysfunctions are, by definition, abnormalities which have detrimental effects on survival or reproduction, where these abnormalities can be (or even have to be) established statistically. In any case, the idea is that we can gain knowledge about e.g. the species design of a certain organ by empirically establishing the statistically normal functioning of this organ (for a certain sex and age group), and that we can therefore (at least to some extent) determine what conditions are pathological through statistical methods. It might also be argued that there is no other way in which we can (in practice) determine this.

(ii) The second view of dysfunction (Wakefield's view) is also inspired by evolutionary theory. Here, a dysfunction is (roughly) regarded as a failure of some internal mecha-

---

[61] It seems that this view was first formulated by Scadding in 1967. In his view, to have a disease is (roughly) to "differ from the norm of the species in such a way as to […] [be placed] at a biological disadvantage" (Kendell 2002, p. 112). Scadding "never explained what he meant by biological disadvantage", however, "but Kendell (1975) and Bourse [sic] (1975) both argued that it must at least encompass reduced fertility and life expectancy" (ibid., p. 112). Kendell also argued that for a disorder to be present, the disadvantage must be due to innate or intrinsic factors. It is worth noting that this view was originally intended as an analysis of "disease" rather than "dysfunction". I don't think it is very problematic to regard it as a theory of dysfunction, however.

nism to perform its natural function, the function for which it was designed by nature, i.e. by evolution. According to Wakefield (1992), a

> natural function of an organ or other mechanism is an effect of the organ or mechanism that enters into an explanation of the existence, structure, or activity of the organ or mechanism. (p. 382).

The explanation that Wakefield has in mind is obviously explanation by natural selection. However, it is not quite clear whether the reference to evolutionary explanation and natural selection should be regarded as part of the definition of "dysfunction". Consider Wakefield's harmful dysfunction (HD) analysis of disorder as it was formulated in Wakefield (1992):

> A condition is a disorder if and only if (a) the condition causes some harm or deprivation of benefit to the person as judged by the standards of the person's culture (the value criterion), and (b) the condition results from the inability of some internal mechanism to perform its natural function, wherein a natural function is an effect that is part of the evolutionary explanation of the existence and structure of the mechanism (the explanatory criterion). (p. 384)[62]

Here, it sounds as if "natural function" (and hence, "dysfunction") is *defined* in evolutionary terms. However, it is quite clear that Wakefield (2000a) rejects this interpretation, e.g. when he claims that "[i]t is a scientific discovery, not a conceptual truth, that functions exist because of natural selection" (p. 39), something which allows him to say that his notion of function is no different from Harvey's (who didn't have a clue about natural selection).

In any case, Wakefield still believes that every disorder originates in a "mechanism with a natural selection history", and that "there must be at least one malfunctioning system responsible for every disorder" (Murphy and Woolfolk 2000a, p. 242). This suggests that the term "dysfunction" is "a scientific and factual term based in evolutionary biology" (Wakefield 1992, p. 374), and that dysfunctions (natural functions, and mechanisms) can, at least in principle, be objectively identified. However, Wakefield (2000a) does not deny that value judgements often play a critical role in identifying functions in practice. The reason why this is possible is that the prototypical functions are benefi-

[62] That is, dysfunction is, on Wakefield's view, necessary but not sufficient for disorder. A dysfunction (as delineated by evolutionary theory) can only be (or cause) a disorder if it the cause of significant harm to the person under present environmental circumstances and according to present cultural standards (cf. Wakefield 1992, pp. 383-384). That's why albinism, reversal of heart position, and fused toes are not considered disorders.

cial.[63] This does not mean that "function" is a value term, however. To attribute a function to a mechanism is not to make a positive evaluation of it. Some functions may even be harmful in our present environment, e.g. the taste for fat.[64]

These are the two most common medical conceptions of dysfunction. On both these views, "function" and "dysfunction" are factual concepts that can be analyzed entirely in non-evaluative and objective terms. According to Wakefield (2000a), both views are also "essentialist views", since they assume that "function" is a scientific concept, and scientific concepts have essentialist structures (p. 28). However, it is also possible to define the concept of dysfunction without any reference to either evolution or biological disadvantage, but it is not likely that such a conception of dysfunction would be value-free. (We will return to this third possibility on pp. 66-67 below.) But before this, we have to take a closer (critical) look at the two essentialist conceptions of dysfunction, and at the idea that dysfunction (defined in any of these ways) is necessary for mental disorder.

But first of all, we need to clarify what a *mental* dysfunction is, both in more general terms and more specifically, i.e. given the two "essentialist" definitions of "dysfunction" that have just been offered. The main reason for this emphasis on the mental is (as we have already seen in connection with the lesion view) that all somaticist versions of the "machine-fault" model makes it hard to see what is mental about "mental disorder". Or as Wakefield (1992) points out, it is not the kind of symptoms that makes a disorder mental. "[F]or a disorder to be mental, there must be a mental dysfunction, although the mental dysfunction might be secondary to a physiological dysfunction." (p. 384) This suggests that genuine mental disorder can only be present if the underlying pathology is of a mental kind, and that the only approach worth taking seriously in this context is the genuine-mental-disease approach.[65]

---

[63] This observation partly explains (together with e.g. the claim about Harvey above) why Wakefield has added condition (b) below to his 1992 definition of "natural function". Wakefield (2000a) defines the concept as follows: "An effect of a feature of an organism is a natural function of the feature if the effect (a) causally explains the existence, maintenance in the species, or detailed structure of the feature of which it is an effect, and (b) it explains the feature via the same essential processes by which the base set of non-accidental beneficial effects of eyes, hands, and so on [e.g. eyes seeing, hands grasping], explain the mechanisms that give rise to them." (p. 36)

[64] Murphy and Woolfolk (2000a) seems to believe that Wakefield resorts to local cultural standards about what constitutes normal behaviour to identify dysfunctions, at least in the mental case. On their view, Wakefield thinks we can identify a malfunctioning mental mechanism by relying on indirect evidence that indicates or correlates with the existence of internal dysfunction, e.g. by recognizing that certain responses are not normal, expectable, proportionate or appropriate. They also ascribe to him the view "that value-free judgments of appropriate or normal behavior can not only be made, but will turn out to be correlative with a sophisticated evolutionary ecologist's objective view of behavior produced by malfunction." (p. 250) We will return to this criticism later.

[65] It is also possible to argue that "[m]ental maladies are distinguished from physical maladies primarily *by the types of symptoms or evils that characterize them*" (Culver and Gert 1982, p. 88, my italics).

The genuine-mental-disease approach: What is a mental dysfunction?

On the genuine-mental-disease version of the dysfunction view, mental disorders are, by definition, caused by underlying mental "part dysfunctions". This view is just as "reductionist" as its somaticist counterpart, i.e. the relevant functional abnormalities are conceived of as mental "machine-faults" located somewhere in the "mental apparatus" or "psychic machinery". Moreover, these mental dysfunctions are regarded as (functional) deviations from the species-design of the mind.

This gives rise to a number of questions. The central question is of course what (exactly) a mental dysfunction is, and how we should determine whether a dysfunction is present. What is a mental function, and what mental functions are there? What is it that can function or malfunction, i.e. what are "the carriers" of mental functions? What are the parts of the psychic machinery, what does the mind (viewed as a "machine") consist of? Can minds malfunction in the same way as diseased bodies malfunction?[66] Is there a species-design of the mind that can be established in an objective and value-neutral way?

To view the mind functionally (as a "mental apparatus") is to view it as a functional (or multifunctional) *system* that consists of a number of lower-level "interlocking" (multi)functional systems (and so on).[67] The "ultimate parts" of the psychic machinery are (to use another mechanical metaphor) *mental mechanisms*. If the mind is viewed not just functionally but also in a reductionist manner, it is (ultimately) mechanisms or processes that have functions, and that can function or malfunction. The function of a mechanism is (very roughly) what it does for the organism, e.g. the contribution it makes to the survival and reproduction of this organism. A more precise account is (as we have seen) given by Wakefield (1992). On his view, a *natural function* of a mechanism is "an effect of the […] mechanism that enters into an explanation of the existence, structure, or activity of the […] mechanism" (p. 382), e.g. what it has been designed to do by evolution.

(This view is characterized more at length in footnote 94 on p. 71 below.) The main objection to this alternative view is that it implies, somewhat counterintuitively, that psychosomatic disorders are not mental disorders.

66 Or as Svensson (1990) puts it: Are there any mental functions that are relevantly similar or analogous to bodily functions, and if there are such mental functional entities, can they malfunction or break down in the same sense as that e.g. a diseased bodily organ malfunctions? (p. 117)

67 As Murphy and Woolfolk (2000b) point out, human beings admit of description at several psychological levels, e.g. there is a distinction between the computational and algorithmic levels of the visual system, and the mind can also be described on an intentional level, i.e. in terms of beliefs and desires (p. 277). Cf. also Dennett 1987, chapter 1.

The reductionism mentioned above[68] does not in any way imply that different mental mechanisms can be regarded in isolation. Instead, it is worth emphasizing that

> [m]ental mechanisms like those involved in perception, motivation, emotion, linguistic ability, and cognition play distinctive but coordinated roles in overall mental functioning, much as different organs play distinctive but coordinated roles in physical functioning. (Wakefield 1992, p. 378)

Moreover, the different mental mechanisms have (e.g. like organs)

> such strikingly beneficial effects and depend on such complex and harmonious interactions that the effects cannot be entirely accidental. Thus, functional explanations of mental mechanisms are sometimes justified by what we know about how people manage to survive and reproduce. (ibid., p. 383)

It is important to notice that "[o]ne mechanism can have many functions, and one function can be performed by an interaction of many mechanisms." (Wakefield 2000b, p. 267) That a dysfunction is present means that some function is not performed by the mechanism (or mechanisms) that is designed to perform this function, and not necessarily that some mechanism has broken down. That is, the notion of dysfunction should (strictly speaking) be understood in terms of "failed functions, not failed mechanisms per se" (ibid., p. 267).

At this point, it is worth noting out that all this talk about mental mechanisms and mental systems does not in any way suggest that there are free-floating mental mechanisms, or that these mental mechanisms occur in some kind of soul or mental substance. Instead, it seems reasonable to assume that mental mechanisms are nothing but processes and activities that take place in the brain. On this view (which is normally called token physicalism, or the token identity theory), every mental occurrence (or token) is identical with a physical occurrence (token) in the brain. But doesn't this view imply that we need to accept the somaticist view of mental disorder after all? Or is there a principled way in which we can distinguish mental from physical mechanisms, an account of mental mechanisms that would (so to speak) save the genuine-mental-disorder view?

Well, it seems so. The genuine-mental-disorder view is not really threatened by the idea that concrete mental events are identical with brain events. The important thing is that there is no type-type identity between the mental and the physical, i.e. that the relevant *types* of mental mechanisms do not correspond to *types* of brain mechanisms in any simple way. If there is no such correspondence, there might be mentalistically defined

---

[68] This "explanatory" type of reductionism should be carefully distinguished from the kind of ontological reductionism which claims that the psychological can be reduced to the physical, i.e. materialism.

disease-types that do not coincide with any physiologically defined disease-type (cf. Svensson 1990). So, is there any reason to believe that (type) physicalism (or the type identity theory) is true, i.e. that all mental occurrences of a certain (mental) type are also of the same physical or neurological type? Not really. It has been pointed out by many theorists that it is highly unlikely that every type of mental event is also a type of neurological event, e.g. that all occurrences of pain have some physical property in common (in virtue of which they belong to the category "pain". The fact that the type identity theory is most probably false allows for the possibility that the brains of patients who share the same "mental disease" do not show any distinctive neurological similarity, which suggests that the disease in question must be defined in mental rather than physiological terms (cf. Svensson 1990, pp. 91-92).

Is there any reason to believe that the genuine-mental-disorder view is threatened by the more plausible *functional identity thesis*, or *functionalism*? According to this view, mental states are functional states. What all mental states of a certain *type* (e.g. all pains) have in common is their causal role in the system as a whole, and the only reason why a certain particular occurrence belongs to a certain mental state type (e.g. what makes it a case of pain) is that it has the causal role that states of that type (e.g. pains) have. This is really a claim about how mental states should be individuated, specified, or defined, viz. that they should be defined by "a triplet of relations: what typically causes them, what effects they have on other mental states, and what effects they have on behaviour." (Blackburn 1994, p. 150; cf. also e.g. Block (ed.) 1980-81, Stich 1983)[69] Blackburn also observes that

> [f]unctionalism is often compared with descriptions of a computer, since according to it mental descriptions correspond to a description of a machine in terms of software, that remains silent about the underlying hardware or 'realization' of the program the machine is running. (ibid., p. 150).

The last sentence indicates that functionalism is neutral with regard to the nature of the mental occurrences (tokens), which implies that it is compatible with the token identity theory characterized above.[70]

If we assume that functionalism and the token identity theory are both true, this seems to suggest that it is primarily functions (types of functions) and dysfunctions that can be mental, and that it doesn't make much sense to attribute mentality to the mechanisms

---

[69] It is an open question to what extent our ordinary mental terms (e.g. "belief", "pain", or "shame") refer to such functional types, but in my view, it seems likely that most of our actual mental terms are *not* functional in this sense. If they were, our so-called folk theory of the mind would be close to perfect, but it isn't.

[70] The idea that there are functional systems that have non-physical realizations, e.g. that the underlying hardware is purely mental, is a rather strange view in this day and age. It is not incompatible with functionalism, however.

which have (or carry) these functions. As far as I can see, this view implies that there is really no difference between (types of) mental functions and dysfunctions, on the one hand, and (types of) brain functions and dysfunctions, on the other, at least as long as we have the higher levels of organization in mind. The question is whether or not this constitutes a threat to the genuine-mental-disorder view. Is it necessary to assume that there is no type-type correlation between mental dysfunctions and higher brain dysfunctions in order to avoid the somaticist view (physical reductionism)? My tentative answer is "no", i.e. it seems to me that functionalism is compatible with the idea that there is such a thing as genuine mental disorders.[71]

The characterization of mental mechanisms and their functions that has been given so far is pretty abstract. To clarify the relevant notions further, we need some concrete examples of mental mechanisms and mental functions. What mental functions are there, and how can we establish what functions there are?

At this point, it might be tempting to believe that our ordinary psychological terms refer to such mechanisms or functions. For example, it might be assumed that cognition, emotion, perception, belief, desire, motivation, and memory are types of mental mechanisms or functions. On such a common sense view, appropriate fear might be viewed as a function of some affective fear-generating mechanism, and false or irrational beliefs might be viewed as failures of the belief-generating mechanism to perform the function that it has been designed to perform, i.e. as mental dysfunctions. Wakefield (1992) sometimes seems to endorse this view, e.g. when he tells us that cognitive, motivational, affective, personological, hedonic, linguistic, and behavioral dispositions and structures are examples of mental mechanisms that have functions (p. 375).

We should resist this common sense temptation, mainly because it tacitly accepts the so-called "folk theory" of the mind, which is a bad theory. First, it seems to assume that the mind is a collection of discrete mental faculties, and that our mental life can be explained in terms of these faculties. Such "faculty psychology" is not plausible, however.[72] An important reason for this is that such a theory can't really explain very much. To claim that someone hallucinates because the perceptual mechanism fails to perform its function, or that someone suffers from a total lack of motivation because the desire-function has broken down, is just as uninformative as the claim that sleeping pills causes sleep because of their dormative powers.

---

[71] We may also ask whether functionalism is true, e.g. whether it is plausible to "see mental similarity only when there is causal similarity" (Blackburn 1994, p. 151). I will not deal with this problem here, though.

[72] According to Murphy and Woolfolk (2000a), "[f]aculty psychology is the view that the heterogeneity of our mental life must be explained by the existence of distinct psychological mechanisms" (p. 251, footnote 5). This suggests that there are many possible faculty psychologies besides our folk psychology, e.g. many other versions of the view "that the mind is a collection of discrete, multifarious mechanisms that make distinctive contributions to overall functioning." (ibid., p. 242) In fact, a rather large number of such theories can be found all through the history of philosophy.

A related reason why we should the reject the common sense faculty view is that the distinction between "part functions" and holistic functioning collapses if this view is accepted. The dysfunction view is based on the assumption that there is a difference between the two levels, and that manifest disabilities (on the holistic level) can be *explained* in terms of underlying part dysfunctions. But if we think of failures in perception, cognition or emotion as part dysfunctions, this idea is no longer valid. Moreover, if we think of underlying pathologies in this common sense manner, it seems that they can no longer be identified in a value-free way. Consider the case of abnormal motivation. It is not just that such motivations are manifest rather than underlying, the abnormality in question is also the wrong kind of abnormality, i.e. it is more like abnormal behaviour (where the abnormality is evaluative) than abnormal biological functioning (where it is not).

This suggests that what we need in order to identify the mental mechanisms and their functions is a well-elaborated functional theory or model of the mental "apparatus", an explanatory model of the human mind which refers to mental structures, mechanisms and functions in a way that does not involve any reference to the manifest behaviours or emotions that we want such a theory to explain. On this view, mental mechanisms and their functions are theoretical constructs that figure in plausible explanatory theories (or maybe just the best theory?) of the human mind. Examples of the kind of theoretical constructs that are needed can be found in Freud's psychoanalytic theory, which includes mental structures (with functions) like the Superego, the Ego, and the Id, as well as defence mechanisms, like e.g. repression.[73] Other examples are the learning mechanisms to which the behavioural theory refers and constructs like "the impulse control function" (that tends to fail in e.g. psychopaths).

The problem with this more sophisticated theoretical account of mental mechanisms, functions and dysfunctions is that there is no such thing as a generally accepted theory in this field. And even if a theory like psychoanalysis would win eventually (it is the most promising attempt according to e.g. Boorse and Wakefield), it might be hard to apply in practice in order to determine whether a mental dysfunction is present, e.g. because there are too few observational criteria (cf. the reliability condition).

This concludes the section on what a mental dysfunction is, and how such dysfunction might be identified. Let us now return to two essentialist conceptions of dysfunction that were characterized above. Do any of these two views offer a satisfactory account of what a mental dysfunction is, and is it reasonable to assume that mental dysfunction (defined in any of these ways) is necessary for mental disorder? And do any of the views succeed in showing us how mental dysfunctions can be identified (in a value-free

[73] Wakefield (1992) uses Freud's repression account of neurotic disorder as an example of what this type of mental dysfunction account might work. Repression is a mechanism that sometimes fails to perform its function (what it was designed to do) in a satisfactory way, and when this happens, the person might become neurotic

way)? I will not restrict myself entirely to the mental case, however. The reason for this is that the dysfunction views of disorder offered by e.g. Boorse, Kendell, and Wakefield purport to define the generic concept of disorder, i.e. not just "mental disorder", but also "somatic disorder".

Dysfunctions as statistical abnormalities which give rise to biological disadvantage

On the first "essentialist" view of dysfunction, dysfunctions are *either* (by definition) statistical functional abnormalities that have detrimental effects on survival or reproductive fitness, *or* they are (by definition) functional abnormalities that have detrimental effects on survival or reproduction, where it is assumed that these abnormalities can (or must) be identified by statistical methods. If this notion of dysfunction is incorporated into the concept of disorder, we get the idea that a disorder is (at least in part) a condition that reduces longevity or fertility, and that is caused by a statistically establishable functional abnormality. Is this a plausible view of disorder in general, and of mental disorder in particular?

Let us first look at the statistical component (assuming that it is part of the definition of "dysfunction" and not merely a means to identify dysfunctions). It is easy to see that the presence of a dysfunction or abnormality cannot be established on purely statistical grounds, e.g. we cannot distinguish between positive and negative abnormalities (above-par and sub-par functioning) in this way. That is, statistic abnormality is not sufficient. But it doesn't seem to be necessary either, since a purely statistical criterion would make it conceptually impossible that the majority of a reference group suffers from a disease, i.e. the majority would be "debarred from being regarded as ill" (cf. Wakefield 1992). It has also been suggested, e.g. by Wakefield (1992) that a purely statistical criterion can "not distinguish between disadvantage due to dysfunction of internal mechanisms and disadvantage due to harmful environments" (p. 379).

Now, this criticism can clearly be avoided if there is no reference to statistical abnormality in the definition of "disorder", and if the relevant statistics are, instead, regarded as one of several means to identify dysfunctions. So, is it more plausible to define "disorder" solely in terms of (internally caused) biological disadvantage? We have already seen that reduced survival or reproduction is not sufficient for disorder (cf. p. 28 above). So, is it more plausible to assume that reduced survival or reproduction is necessary for disorder? It doesn't seem so. There are clearly a large number of e.g. harmful conditions that we intuitively classify as mental disorders that might not reduce fertility or longevity at all, like e.g. different phobias.[74] What these examples suggest is that it is more cru-

[74] To conceive of reduced survival or reproduction as sufficient for disorder doesn't just fail to satisfy the ordinary language condition. Such a conception of disorder also fails to satisfy the value condition (there are many things besides reduced survival or reproduction that are bad for an individual), the con-

cial whether a condition causes harm than whether it has detrimental effects on survival or reproduction in our present environment (which is a rather specific type of harm). As Wakefield (1992) points out,

> the biological disadvantage approach mistakenly uses decreased longevity and fertility in the present environment as the criterion for mechanism dysfunction. The fact that the organism's mechanisms were originally selected because they increased longevity and fertility in a past environment does not imply that some mechanism is malfunctioning when longevity and fertility decrease in the present environment. (p. 379)

He also adds that "[i]t is the failure of specific mechanisms to perform their assigned tasks, rather than lowered fitness in itself, that shows that something has gone wrong with the organism." (ibid., p. 379) So, let us now try to find out if Wakefield's own version of the dysfunction view is more plausible.

Dysfunctions as failures to perform natural functions

On Wakefield's conception of dysfunction, a dysfunction is a failure of some internal mechanism to perform its natural function, where a natural function of a mechanism is (roughly) an effect of this mechanism that explains why it was naturally selected. If this notion of dysfunction is incorporated into the concept of disorder as a necessary condition, we get the idea that a condition cannot be a disorder unless "the condition results from the inability of some internal mechanism to perform its natural function" (Wakefield 1992, p. 384).

Now, on Wakefield's view, a disorder cannot be mental unless it is caused by a *mental* dysfunction, i.e. unless it results from the failure of a *mental* mechanism to perform a natural function for which it was designed by evolution (ibid., p. 373).[75] That is, it is as-

---

dition of normative adequacy (it doesn't give us any plausible normative guidance at all), and the reliability condition (it is sometimes hard to determine whether a certain condition is associated with biological disadvantage in the relevant sense). To conceive of reduced survival or reproduction as necessary for disorder, e.g. to define "mental disorder" as harmful dysfunction (where dysfunction is defined in terms of reduced survival or reproduction) is more plausible, due to the value component. Such a conception also suffers from certain defects, however. Apart from the fact that it fails to satisfy the ordinary language condition, it gives us bad normative guidance (e.g. it would be implausible to require that a condition is associated with reduced survival or reproduction in order to treat it), and it is as unreliable as the idea that reduced survival or reproduction is sufficient for disorder.

[75] Wakefield's (1992) full analysis of mental disorder is a straight-forward application of his harmful dysfunction analysis of disorder in general (see p. 51 above) to the mental case: "A condition is a mental disorder if and only if (a) the condition causes some harm or deprivation of benefit to the person as judged by the standards of the person's culture (the value criterion), and (b) the condition results from the inability of some mental mechanism to perform its natural function, wherein a natural function is an

sumed that mental mechanisms and processes have natural functions, i.e. that they can be naturally selected in the same way as e.g. organs have been selected (cf. ibid., p. 375). As Murphy and Woolfolk (2000a) put it, Wakefield is "an empirical adaptationist about the mind", he believes "that human psychology, like human physiology, is a product of natural selection", and "that the mind consists of mechanisms designed by natural selection" (p. 242).[76]

Wakefield's evolutionary view has been criticized by many writers, e.g. by Lilienfeld and Marino 1995, Murphy and Woolfolk 2000a, 2000b, and Nordenfelt 2003. Most of the objections to Wakefield's theory purport to show that dysfunction (as defined by Wakefield) is not a necessary condition for mental disorder, i.e. that someone may well suffer from a mental disorder even when there is no "evolutionary malfunction". Some of these arguments try to establish (1) that "many mental functions are not direct evolutionary adaptations, but rather adaptively neutral by-products of adaptations" (Lilienfeld and Marino 1995), and that some disorders are failures with regard to these functions rather than with regard to natural functions. That is, the idea is that disorders can be caused by mechanisms that have no adaptive function. Other arguments of this type (2) purport to establish that mental disorders can even be caused by mechanisms that are working exactly as designed by evolution, e.g. that "many consensual disorders represent evolutionary adaptive reactions to danger or loss" (ibid.).

It is also possible, however, to (3) criticize the idea that dysfunctions (natural functions, and mechanisms) can be objectively identified. That is, it can be argued that in the absence of an established evolutionary science of the mind, Wakefield's concept of mental disorder is almost impossible to *apply*, and that we therefore have to rely on value judgements (e.g. intuitions about what responses are normal, expectable, proportionate or appropriate) to identify the relevant mental mechanisms and to recognize when they malfunction.[77] This is a striking argument, but it may not be as strong as it seems. First,

---

effect that is part of the evolutionary explanation of the existence and structure of the mental mechanism (the explanatory criterion)." (p. 385)

[76] According to Murphy and Woolfolk (2000a), Wakefield "goes further; he assumes that the mind is a collection of discrete, multifarious mechanisms that make distinctive contributions to overall functioning." (p. 242) It is somewhat doubtful whether he actually believes this, however. See p. 54 above.

[77] This is how Murphy and Woolfolk (2000a) argue, e.g. when they claim that "in the absence of some capacity to discern the natural kinds or psychological essences that Wakefield presumes, there is no way to identify a class of dysfunctional mental phenomena without engaging in some form of value judgment. Inevitably, we must start out with what we already think of as pathological […]." (pp. 250-51). Now, on their view, we can only draw the line between the pathological and the non-pathological if we are able to discriminate normal, proportionate and appropriate responses from abnormal, disproportionate and inappropriate ones. But how do we decide e.g. whether a certain situation warrants a depressive reaction? According to Murphy and Woolfolk, we cannot do this without recourse to some evaluative standard of normality: "We believe that the methods Wakefield thinks are required to apply the concept of dysfunction will inevitably incorporate judgments of moral and aesthetic value, because they can be applied only against a background of shared values about normal behavior." (ibid., p. 247)

it does not, unlike (1) and (2), concern Wakefield's *definition* of "mental disorder". It does not focus "on the ontological or definitional question of the nature of function and dysfunction but on the epistemological question of how we come to judge that a condition involves a dysfunction" (Wakefield 2000b, p. 264). Second, it seems reasonable to assume that we, *in practice*, often have to rely on judgements of *disorder* to determine whether or not a dysfunction is present, and I don't think anyone would deny that value judgements often play a critical role when we decide that a certain condition is a *disorder*, i.e. that "values and local norms may [actually] influence such judgments in a variety of ways" (ibid., pp. 264-265). The important thing here is not whether judgments of dysfunction are actually objective, however, but whether they *can* be objective, i.e. "whether an appeal to values is intrinsic to the logic of the evidence for such judgments." (ibid., p. 264) And to settle *this* question, it is not sufficient to point out that our *actual* judgements of dysfunction can often be explained in terms of "local values and norms" or "socially constructed criteria of normality". In fact, and this is the third point, it seems plausible to assume that *if* Wakefield's analysis of dysfunction is correct, then dysfunctions can (in principle) be identified objectively.[78] (This is not to say that *disorders* can be identified objectively, however, at least not if the *harmful* dysfunction analysis of disorder is correct.) As Wakefield himself puts it, even when our judgements of design are in fact influenced by local values, this does not in any way imply that these judgments themselves are "constituted, even in part, by such values. The truth or falsity of the judgments about design […] are in principle factual and independent of the values that influenced them" (ibid., p. 265).[79]

Let us now take a closer look at the more serious objections, e.g. arguments of type (1) and (2). However, there will be a more thorough discussion of these arguments in Radovic (forthcoming), and I will therefore be somewhat brief here.

(1) The arguments in the first group purport to show that mental disorders can be caused by failed mechanisms that have no adaptive function, e.g. by spandrels, exaptations, or vestigal parts (or free riders). Consider the case of *spandrels*, i.e. "adventitious by-products of the development of other traits" that "themselves have never possessed any adaptive function" (Murphy and Woolfolk 2000a, p. 243).

> If *mental* spandrels exist, then there are mental mechanisms that are the by-products of evolution but have themselves never possessed adaptive functions (in Wakefield's evolutionary sense) and, therefore, could not malfunction [in the relevant

[78] Even Murphy and Woolfolk (2000b) admit that malfunction, in the evolutionary sense, is an objective notion, and that it is possible to determine an evolutionary malfunction objectively. As we will soon see, they don't think this is the correct notion of dysfunction, however.

[79] He also believes "that we often correct our views of disorder when we learn about the cross-cultural evidence, even if this goes against our own values", and that this alleged fact constitutes "[o]ne powerful piece of evidence" for this view (ibid., p. 265).

sense]. Such mechanisms, however, could produce pathological behavior. (ibid., p. 243)[80]

This suggests that Wakefield's notions of function and dysfunction are not the most plausible notions in this context, and that the evolutionary notion of (natural) function needs to be replaced by some other notion of function, e.g. the notion of a Cummins-function (see p. 66 below).[81]

Wakefield (2000b) defends himself by maintaining that "the failure of a spandrel implies a disorder when and only when it implies the failure of a naturally selected function" (p. 254). For example, if someone is incapable of learning to read even under optimal learning conditions (a possible case of mental disorder), "we infer that there is something wrong with some internal neurological mechanism that, when functioning as designed, supports the capacity to read (although it supports reading accidentally, not by design)" (ibid., p. 256). This argument does not seem convincing to me, but it is hard to prove him wrong (or to see how one might convince him that he is wrong). But cf. Radovic (forthcoming).

(2) The arguments in the second group purport to show that mental disorders can be caused by mechanisms that perform their natural functions, i.e. that are working exactly as designed by evolution. An example of such an argument is given by Murphy and Woolfolk (2000a), when they claim that disorders can be caused by mechanisms that are processing "pathogenic input" but are doing so according to their design (cf., pp. 244-245). On their view, it is

> quite possible that underlying psychological mechanisms, functioning as designed, may give rise to deviant behavior because they are being fed, as input, pathogenic information that the subject acquired through idiosyncratic learning. Such behavior should be treated by reeducating the subject, not by assuming that some internal mechanism is broken. […] There is a clear and consequential distinction between a

---

[80] Nordenfelt (2003) refers to this argument as the argument from exaptation. This argument "says that there are several human functions (in the intuitive sense of function) which have never been selected in the evolutionary sense. This holds, in particular, for some of our advanced mental functions, such as reading, writing, or calculating. A disability with regard to such functions is normally regarded as a disorder or disease. However, since these are failures with regard to exaptations and not with regard to natural functions they do not fulfill Wakefield's criterion of diseasehood." The other argument offered by Nordenfelt is "the argument from free riders", which says that "there are several organs present in existing organisms today which have never been selected in the evolutionary sense. According to the definition of natural function these organs cannot have natural functions. Thus, they cannot be disordered. This is a counterintuitive conclusion." (These two quotations are both from Nordenfelt's own summary of the 2003 article, formulated in private communication.)

[81] It is worth noting that the first dysfunction view (see pp. 58-59 above) cannot be criticized in this manner.

> broken mechanism and a functioning mechanism operating over deviant input. Wakefield's analysis obscures this distinction. (p. 245) [82]

As it stands, this is not a very convincing argument. First, it seems to assume that dysfunctions are broken mechanisms. This is not true, however. For a dysfunction to be present, it is sufficient if some mechanism fails to do what it is designed to do. Second, the fact that a mental disorder is "a result of problematic social learning", that it is "acquired through normally operating mechanisms of learning and conditioning" (ibid., p. 245), does not mean that there is no dysfunction present. Injuries due to external trauma are obviously disorders that involve dysfunctions, and so are inflammatory reactions and infectious diseases due to invasions of noxious agents. There are mental injuries of this type as well, e.g. posttraumatic stress disorder.

However, the fact that some mental disorders are not just acquired by "learning", but that they can also be treated by psychotherapy, suggests that there is really no mental dysfunction involved in these cases. The difference between the phobic and the non-phobic is most probably not that there is some specific mechanism that is malfunctioning in the one case but that is functioning as designed in the other. Instead, it seems more reasonable to assume that the difference between the phobic and the non-phobic is simply that they *function in different ways*, where "the phobic way of functioning" is far more harmful than the non-phobic way. Or consider the personality disorders. As Kendell (2002) points out, personality disorders are often "distinguished from mental illness by their enduring, potentially life-long nature and by the assumption that they represent extremes of normal variation[83] rather than a morbid process of some kind." (p. 111) Kendell also points out that this idea is, in part, due to the fact that "as yet little is known of the underlying mechanisms of which they [the personality disorders] are a manifestation" (ibid., p. 113). All this suggests that personality disorders do not involve

---

[82] Murphy and Woolfolk (2000a) also present another argument of type (2), viz. the idea that disorders can be caused by mechanisms working as designed by evolution in circumstances they were not designed to be in. On their view, "disordered behavior in our modern environment is a poor guide to underlying dysfunction. […] The environment in which selection pressures acted so as to leave us with our current mental endowment is not the one we inhabit now. For any mental mechanism producing harmful behavior in the contemporary world, it may be that the mechanism is fulfilling its design specifications to the letter, but the environment is not one in which the design promotes well-being." (p. 244) They also add that mental disorders characterized by fear and avoidance are possible examples of such "environment-design mismatches". Against this, Wakefield (2000b) argues that "it is just not true that when we believe that the way we are designed has become problematic due to new environmental conditions, we consider the problem a disorder. For example, […] [w]e are designed with a 'fight or flight' reaction to threat, and this reaction is apparently harmful in the modern world of constant human competition and stress […], but this reaction is not considered a disorder." (p. 258) According to Wakefield, Murphy and Woolfolk "take it as obvious that such undesirable mismatches are disorders, which begs the question." (ibid., p. 259) In my view, this is a convincing piece of defense.

[83] Or alternatively put, "[t]he behaviours and attitudes that define personality disorders are probably graded traits present to a lesser degree in many other people" (ibid., p. 112).

dysfunctions (e.g. in Wakefield's evolutionary sense). That is, it seems that *if* we conceive of the personality disorders as mental disorders, then we have to reject this type of dysfunction view.[84] In short, it seems that we can't assume that all mental disorders involve "failing functions", or "sub-par functioning". In some cases, it makes more sense to speak of different *ways* of functioning rather than different *degrees* of functioning.

It is of course possible to defend oneself against this criticism by simply *insisting* that e.g. conditions acquired through normal learning processes can only be regarded as disorders when they involve dysfunctions. However, it seems that one has to pay a considerable price for doing so. Consider the following line of reasoning in Wakefield 2000b:

> [I]f a repressed sense of loss from early in life leads to rigidly fixed unconscious beliefs that consistently cause one to greatly overinterpret routine losses, then there is a failure of function of the loss-assessment mechanism that provides input to the sadness response, even if the beliefs were originally attained through normal learning processes. (p. 263)

In order to save the idea that "whenever there is a disorder, there simply has to be a dysfunction", Wakefield has to postulate "loss-assessment mechanisms" and other strange mental mechanisms. It is not just that this sounds like a dubious form of faculty psychology, it is also somewhat doubtful that all the mechanisms (faculties) that need to be introduced to save the idea are naturally selected.

Is *any* essentialist dysfunction view plausible?

In short, it seems that Wakefield's evolutionary notion of dysfunction is somewhat problematic, and so is the concept of mental disorder that is (so to speak) based on it.[85] Is

---

[84] However, it can be doubted whether personality disorders should be regarded as mental disorders, and not just because they might not involve dysfunctions. For example, we might want to exclude them from the category "mental disorder" because they are too unlike somatic disorders, because they can or should not be managed by the health care system, or because people with personality disorders do not deserve to be accorded the priviliges of the 'sick role'. Another way in which the dysfunction view might be defended is by questioning the idea that personality disorder can be sharply distinguished from "mental illness". According to Kendell (2002), the idea that personality disorders are enduring conditions that represent extremes of normal variation, whereas mental illnesses are morbid processes, "appear increasingly questionable, and as a result the distinction between illness and personality disorder is starting to break down" (p. 113). For example, "[s]ome schizophrenic illnesses have the same time course as personality disorder", and "it is becoming increasingly clear that the genetic bases of affective personality disorders and mood disorders, and of schizotypal personality disorder and schizophrenia, have much in common" (ibid., p. 113).

[85] There are two types of problems with Wakefield's definition of "mental disorder": those related to the idea that harm is necessary for disorder, and those associated with the idea that dysfunction (in the evolutionary sense) is necessary for disorder. The idea that dysfunction is necessary for disorder suffers from several weaknesses. It doesn't just fail to satisfy the ordinary language condition (malfunctioning

there any reason to believe that we can *ever* come up with an essentialist and value-free notion of dysfunction that is also necessary for disorder? There are several reasons for doubting that all disorders (by definition) involve objectively establishable dysfunctions, at least if we do not deviate totally from our present psychiatric classification systems, where e.g. phobias and personality disorders are classified as mental disorders.

First, the dysfunction theory implies that if no underlying pathology has been (objectively) identified, we can't really know whether a certain condition is a disorder. This is rather counterintuitive, however.

Second, the idea that some mental disorders are best conceived of as examples of harmful internal *functioning* rather than as harmful *dysfunctions* is not just an objection to Wakefield's evolutionary theory of disorder, but to all essentialist or scientific dysfunction views of disorder. These theories assume that all conditions properly classified as mental disorders have the *same type of internal cause*, and that this type is (roughly) a natural kind, a class that exists "out there" in nature. This seems to imply that the class of mental disorder is quite homogeneous, that all mental disorders are "the same kind of thing", i.e. that they belong to the same ontological category. It is likely, however, that the different mental disorders (as they are commonly defined) do *not* belong to the same ontological category. As May et al (2002) point out, it is possible "that the nature of psychopathology is intrinsically heterogeneous [i.e. that the simplicity condition cannot be satisfied] consisting in part of true disease entities and in part of reaction types or maladaptive response patterns." (p. x)[86] In short, it seems that some of the disorders listed in DSM are diseases, whereas other diagnostic rubrics refer to ontological categories like syndromes (in the medical sense), isolated symptoms, habitual patterns of behaviour, subjective experiences, and personality traits.[87] If we accept this heterogeneous picture as valid (e.g. because it satisfies the ordinary language condition and the condition of normative adequacy to a high degree), it seems hard to maintain the view that all disorders have the same type of internal cause, viz. that they are all caused by some mental dysfunction.

---

spandrels do not qualify as disorders) and the reliability condition (it is very hard to determine whether or not a dysfunction is present), it also fails to satisfy e.g. the condition of normative adequacy (if we think that a condition should not be treated unless it is a disorder, it suggests that malfunctioning spandrels should not be treated).

[86] For example, it is possible that bipolar disorder is a "disease arising from a defect in the brain machinery", whereas some of the anxiety disorders rather "arise from a dysregulation of defenses" (ibid., p. x).

[87] According to Jablensky and Kendell (2002), "[t]he term 'disorder', first introduced as a generic name for the unit of classification in DSM-I in 1952, has no clear correspondence with either the concept of disease or the concept of syndrome in medical classifications. It conveniently circumvents the problem that the material from which most of the diagnostic rubrics are constructed consists primarily of reported subjective experiences and patterns of behavior. Some of those rubrics correspond to syndromes in the medical sense, but many appear to be sub-syndromal and reflect isolated symptoms, habitual behaviours, or personality traits." (p. 6)

To sum up, the traditional medical (essentialist) dysfunction analysis of disorder is problematic for several reasons – mainly because it is inconsistent with ordinary language and because it fails to satisfy the condition of normative adequacy – and it should therefore be rejected. Look us now look at some alternatives, i.e. some other attempts to capture the relevant kind of internal cause.

3. Other dysfunction alternatives

Before we reject all kinds of dysfunction views, we have to ask ourselves whether there is any alternative notion of dysfunction that is more plausible in this context. What we want to know is whether it is possible to define the concept of dysfunction without any reference to evolution or biological disadvantage, and in a way that permits us to say that e.g. spandrels have functions.

An example of such a definition is the one offered by Cummins in 1975. On this view, a function is (roughly) a causal role in a system: The function of a certain structure or mechanism is the causal contribution it makes to the overall operation of the system that contains it, i.e. a certain structure has a "Cummins-function" if it contributes to the operation of the overall system. So, can the notion of Cummins-function be of any use in this context, i.e. can it give us a notion of mental dysfunction that can help us distinguish between disorder and non-disorder? Well, it certainly seems preferable to Wakefield's evolutionary notion of dysfunction, mainly because it has the advantage of not presupposing that we can discriminate mental adaptations from e.g. mental spandrels and exaptations in order to apply the concept of dysfunction correctly: they all have Cummins-functions (cf. Murphy and Woolfolk 2000b).

It is not likely that this concept of function can be applied without value judgments, however. To determine whether a certain mechanism has a function in this sense, and what this function is, we have to determine how it contributes to the operation of the overall system. Now, this is only possible if we have some idea of what the system and its goals are, and this can hardly be determined in an objective way.[88] As Murphy and Woolfolk (2000b) put it, "[t]he background context of inquiry, what makes the concept of proper functioning or dysfunction within a Cummins functional system intelligible at all, is laden with human interests from the very beginning" (p. 283). In short, it seems that we have to rely on value judgements in order to identify a dysfunction in a Cummins function, e.g. to determine the failure of a spandrel.

[88] In physiological contexts, there are exceptions to this rule, e.g. the goal of the temperature regulation system can clearly be determined objectively. At this point, it might be argued that survival and reproduction are the obvious goals of biological systems, but as far as I can see, this is a somewhat arbitrary choice. Moreover, we have already seen on pp. 58-59 above that this suggestion does not help us to come up with a reasonable definition of "disorder".

Megone's (2000) "Aristotelean" notion of dysfunction is an example of such a Cummins-type of view. According to Megone, "the ultimate goal for a good human being" ("the human function") is "to live the life of a fully rational animal" (p. 49). After having identified "the function of the whole" in this way, he can then maintain that "[t]he function of the bodily and mental parts of a human is to operate in ways which contribute instrumentally or constitutively to the realization of" this goal (ibid., p. 49). Illness, "whether bodily or mental", can then be analyzed as "an incapacitating failure of bodily or mental capacities, respectively, to realize their functions" (ibid., p. 49), or alternatively, as an "incapacitating failure to realize (actualize)" a fully rational life (ibid., p. 56).

Is this a plausible notion of mental dysfunction, i.e. can it help us to draw the line between mental disorders and other conditions in a reasonable way? Well, this seems to depend on how the goal state is specified, and this is not just because some possible specifications of goal states give rise to definitions that fail to satisfy the value condition. It is also quite hard to identify the relevant "part functions" and "dysfunctions", regardless of what the goal state is. For example, a definition which is based on the assumption that the goal state is "a fully rational life" is probably lacking in precision as well as in reliability, but the same thing seems to hold for a definition based on the assumption that the goal state is "to be able to realize one's vital goals" (see p. 68 below). The plausibility of this analysis also depends on how well it deals with the group (2) arguments on pp. 62-64 above. As far as I can see, these arguments are not much of a threat to the Cummins-type of dysfunction view. On the other hand, this suggests that there is little or no difference between this view and the idea that disorders are simply harmful conditions caused by internal states or processes. That is, to keep things simple, it seems that we might just drop the notion of dysfunction altogether. To see if this is so, let us now take a closer look at the idea that any internal cause will do.

4. The modern medical model: Any internal cause

We have seen that the traditional medical dysfunction view is most probably too narrow. For example, if dysfunction (defined in evolutionary terms) is regarded as necessary for disorder, this seems to exclude too many consensual disorders from the category of disorder. On the other hand, it seems quite clear that if someone has a mental disorder, this means that something has gone wrong with the person. So the question arises: How should we specify the central notion of "going wrong", now that we can no longer appeal to dysfunctions? Well, we have seen that the appeal to failed Cummins-functions is one possibility. A very similar suggestion is the idea that any internal cause will do, or more specifically, that any *type* of internal cause will do.

On this type of view (adopted by e.g. Nordenfelt, Pörn, Reznek, Whitbeck, and Sedgweck), diseases and other disorders are, by definition, internal conditions that tend to

compromise people's health, where "health" is, as a rule, defined in terms of functional ability. Or alternatively put, a disorder is an internal state or process that tends to give rise "illness", defined in terms of incapacitation or disability. Nordenfelt's (1995) definition of malady (an umbrella term for diseases, impairments, injuries and defects, i.e. disorder) is a good example of this type of view. According to Nordenfelt, "*M* is a type of malady in environment *E* if, and only if, *M* is an episode-type [physical or mental] which, when instanced in a person *A* in *E*, causes with high probability illness in *A*" (p. 149), where a person *A* is ill if and only if *A* is, in standard circumstances, disabled from realizing his vital goals.[89]

The main reason why Nordenfelt and others put so much emphasis on *types* is that it allows for the possibility that there are disorder-instances (e.g. undetected cancer) that do not affect the health of a person. To think in terms of types rather than in terms of instances is also crucial if we want to come up with a plausible distinction between mental disorder and somatic disorder. What makes a disorder mental is supposedly that it is caused by some mental factor (cf. p. 52 above).[90] And since it may well be the case that all mental episodes are identical with episodes in the brain, the only plausible way of distinguishing bodily from mental illness is, as Svensson (1990) suggests, "by way of dividing the *types* of internal factors that cause the illness into physical and mental" (p. 108, my italics).

If we replace the idea that disorders, by definition, tend to give rise to incapacitation with the more "generous" idea that a disorder is, by definition, harmful (cf. pp. 32-34 above), we get the following definition of "mental disorder": A mental disorder (token) is a type of condition that tends to give rise to harm, and that is caused by some internal episode of a mental type. That is, the relevant internal cause is of a mental type, and it is of a type such that most or all instances of this type cause harm (e.g. disability). On this view, it is not really possible to give any further specification of the kind of internal cause that is present when someone has a mental disorder.

Now, this view is obviously far more "inclusive" than the dysfunction view, at least on the assumption that there are many types of internal factors besides e.g. failures of naturally selected mechanisms that cause harm. In fact, this "inclusive" view may well be over-inclusive in the sense that it seems to include a number of consensual non-disorders in the category of disorder, e.g. conditions like ignorance, fanatical beliefs, normal grief, or being madly and unhappily in love. If Nordenfelt and others think that these conditions should *not* be regarded as mental disorders, they have to explain why,

---

[89] The concept of disease is defined as follows: "*D* is a disease-type in environment *E* if, and only if, *D* is a type of physical or mental process which, when instanced in a person *P* in *E*, would with high probability cause illness in *P*." (p. 108) That is, what distinguishes diseases from other maladies is, on this view, that they are *processes*.

[90] But again, cf. footnote 94 on p. 71 below.

i.e. they have to show us how genuine mental disorders should be distinguished from other mental incapacity-producing conditions. As Svensson (1990) puts it,

> [b]y making consequences on the whole-person-capacity level the defining characteristics of disease, this approach seems to run into difficulties in separating 'mental diseases' from other mental, illness-producing, circumstances. […] [T]he characterization of disease in terms of illness-producing circumstances is not sufficient to delineate a mental-disease category, unless we would want to expand this category to include a variety of mental incapacity-producing circumstances that we do not ordinarily view as in any way pathological. (p. 117)

It is hard to see on what grounds a theorist of this type can argue that e.g. fanatical beliefs or unhappy love are not disorders. It is of course possible to refer to what is considered normal in the culture in which the condition appears, but this would really turn it into another theory, which includes a new evaluative element. Another possibility is to argue that e.g. ignorance or fanatical beliefs are only harmful in certain environments, and that they should therefore be conceived of as suboptimal behaviours (or the like) rather than as disorders. However, this strategy cannot explain why conditions like unhappy love or "normal grief" should not be regarded as pathological.

Another difficulty that the "any internal cause" theorist has to deal with is how to distinguish the internal from external, or more specifically, how to determine whether a certain incapacitating condition is caused by internal factors or whether it is "merely" a reaction to environmental factors. For example, it seems pretty clear to me that almost every painful emotion is caused by a combination of internal and external factors, and that the immediate (proximate) cause is normally internal, viz. that the subject interprets and evaluates the external situation the way he does. This is as true in the case of grief (which we tend to regard as non-pathological) as it is in the case of phobic reactions (which we tend to regard as pathological). Now, if the "internal cause" theorist agrees with the idea that phobic reactions are pathological whereas "normal grief" is not, he has to explain why we should only think of the phobic reaction as internally caused. In short, he has to distinguish internally caused conditions from externally caused conditions in a way that works in this context, i.e. helps us to delineate the concept of mental disorder.[91] (It is worth noting that dysfunction theories never encounter this type of problem.) One possibility here is to appeal to some evaluative standard of normality, and to claim that the grieving of a loved one is externally caused because it is a "nor-

[91] Here is another difficult case. Suppose that someone repeatedly finds himself in a certain (external) situation because of his personality traits. Suppose also that the situation he is in (e.g. a situation characterized by loneliness and unemployment) causes him to suffer, and that this situation would be painful to almost everyone. Should we regard his distress as internally or externally caused (in the relevant sense)?

mal" and "intelligible" reaction, whereas intense anxiety in the presence of a dog is internally caused because it is "abnormal" or "unintelligible". This strategy is not really open to the "internal cause" theorist *qua* "internal cause" theorist, however, since it would (again) turn his theory into a different theory, a theory with an additional evaluative element. (There is a promising strategy that we will soon look at, however, viz. the idea that a condition is internally caused if it has no distinct sustaining cause.)

In short, the idea that a harmful condition is a mental disorder if it is caused by some kind (any kind) of internal mental factor is quite problematic. It fails to satisfy the ordinary language condition by being overly inclusive, and it probably fails to satisfy the condition of normative adequacy for the same reason. For example, if conditions like normal grief and unhappy love are regarded as pathological, we get little guidance from the idea that people who suffer from mental disorders are entitled to health care or compensation. Moreover, it is not quite clear how we should distinguish those conditions which are internally caused from those that are externally caused, and the definition is therefore lacking in precision as well as in reliability. It can also be argued that there is something resigned about this type of analysis, e.g. compared to Wakefield and others, who are at least trying to say something substantial about the factual component of the concept of disorder. Is it really impossible to say more about what makes a condition a disorder than that it is harmful and internally caused (in some vague sense)? Well, it is at least possible to define "internal cause" in an interesting way, viz. in terms of the absence of a distinct sustaining cause. Let us now take a closer look at this idea.

5. Culver and Gert: No distinct sustaining cause

Culver and Gert (1982) share the general view that disorders are (roughly) internally caused harmful conditions. To clarify this view, they specify the phrase "internally caused" in terms of the absence of a distinct sustaining cause:

> We could say that to be a malady, the evil-producing condition must be part of the person. However, for reasons of conceptual rigor, a more formal negative statement is preferable: the person has a malady if and only if the evil he is suffering does *not* have a sustaining cause which is clearly distinct from the person. (p. 72)[92]

The notion of a distinct sustaining cause is defined as follows: X is a distinct sustaining cause of a condition C if and only if (a) X is a cause of C; (b) X is not part of the person with C, i.e. it is distinct from this person; and (c) if X were removed, C would cease to

[92] As I see it, this idea is closely related to Reznek's idea that diseases are, by definition, involuntary conditions which cannot be eliminated by a decision, and that we are (therefore) not immediately responsible for our disorders. On the assumption that distinct sustaining causes are normally possible to eliminate, that is.

exist almost immediately, i.e. X is necessary for sustaining C [93](cf. Wilkinson 2000, p. 301). That is, the idea is that if a condition has a distinct sustaining cause, then it is externally caused (in the relevant sense), and it cannot be a malady (disorder). This is not to deny that a malady may have been caused *originally* by factors distinct from the person. The important thing is that to be a malady, a condition is *not now* "in a state of continuing dependence on those [external] factors; rather, it is present even in their absence" (Culver and Gert 1982., p. 72).

Is this a plausible view, i.e. is it reasonable to regard maladies (disorders) as harmful conditions that are "not in continuing dependence upon causes clearly distinct from oneself"? Well, on the face of it, this view certainly seems preferable to the vaguer "internal cause" view that we have just looked at. However, it is worth noting that Culver and Gert's analysis does not help us to distinguish those internal causes that are mental from those that are physical, and that we therefore (to some extent) have to rely on our intuitions to separate the mental disorders from the physical ones.[94]

So, is this analysis reasonable? Suppose that someone has a phobic reaction (a panic attack) in the presence of a dog. Suppose also that the anxiety would disappear almost immediately if the dog were removed. This suggests that the phobic reaction is not really a disorder, which is counterintuitive. At this point, it might be objected that there is more to phobia than panic attacks in the presence of the feared object, viz. that it essentially involves a kind of involuntary "avoidance behaviour" which tends to have disruptive effects on life. This behaviour is not sustained by any distinct cause, however, and the same thing holds for the volitional disability associated with it. Thus, phobia cannot really be regarded as a counter-example to Culver and Gert's analysis after all. However, as Culver and Gert themselves observe, even though phobias are "frequently associated with intentional unvoluntary [sic] behavior" (ibid., p. 119), these conditions "do not always involve a volitional disability; they sometimes involve only the suffering of inappropriate anxiety" (ibid., p. 111). Or more specifically, "only the irrational fear is

---

[93] In Culver and Gert's (1982) own words, a sustaining cause is "a cause whose effects come and go simultaneously (or nearly so) with its respective presence and absence" (p. 72).

[94] On the assumption that it is not the kind of symptoms that makes a disorder mental, that is, but the type of internal factor (underlying pathology) that gives rise to the condition (cf. p. 52 above). Culver and Gert (1982) reject this idea, however, and they therefore feel no need to distinguish between different types of internal causes. On their view, "[m]ental maladies are distinguished from physical maladies primarily by the types of symptoms or evils that characterize them. Though etiology plays some role, as does the type of treatment that is effective in relieving the symptoms or curing the malady, it is the dominant symptoms that play the largest role in determining whether we regard a person as having a physical or mental malady. […] Physical maladies have as their predominant symptoms physical pain and physical disability. […] Mental maladies are maladies in which the evil(s) being suffered are primarily mental pain and mental disability. Mental disability includes both cognitive and volitional disability." (p. 88)

necessary, not any actual avoidance behavior" (ibid., p. 120).[95] This suggests that phobia constitutes a counter-example after all, i.e. that there are disorders that are sustained by distinct causes. (It is rather obvious that the phobic *disposition* does not have a distinct sustaining cause, however.)[96]

Now, consider the case where someone is distressed or incapacitated because of his fanatical religious beliefs, e.g. because these beliefs don't really allow him to live "a normal life". If we conceive of a person's beliefs as an integrated part of this person,[97] this suggests that the religious fanatic suffers from a mental disorder, which might also be regarded as somewhat counterintuitive. On the other hand, it might be argued that fanatical beliefs are best regarded as "viruses of the mind", and that they should therefore be classified as mental disorders.[98]

A third type of case that might be problematic for Culver and Gert is an incapacitating condition like severe (but "normal") grief. Suppose someone is grieving because he just lost a loved one. Does this condition have a distinct sustaining cause? Well, it seems that the sustaining cause of the grief is really the person's belief that he has lost someone he loves. So, is this belief distinct from the person? Well, if it is not, this would suggest that "normal grief" is a disorder (contrary to what Culver and Gert themselves think). However, it might be argued that the belief in question is sustained by the state of the world, viz. the fact that the loved one is gone. That is, if the world would change and the loved one would return, the grief would disappear, which suggests that grief is not a disorder after all. As Wilkinson (2000) points out, it is somewhat strange way to think of eternal facts as sustaining causes, however. If we want to regard the grief (and the belief) as externally caused, it seems far more plausible to regard the loss (i.e. the concrete event) as the cause, and although this cause is clearly distinct, it can hardly be regarded as sustaining. In short, it seems that Culver and Gert's analysis imply that normal grief is a mental disorder after all, which might be linguistically counterintuitive to many of us. A concept of mental disorder that implies that grief is a disorder might also be considered less normatively adequate than a concept that does not have this implication.

---

[95] For example, "[i]t is possible for someone with claustrophobia to feel great anxiety when the situation requires him to enter an elevator, but to enter it in spite of the anxiety. In this case, the phobia is a mental malady [disorder] because of the intense, inappropriate anxiety, but it does not involve a volitional disability" (ibid., p. 120).

[96] Another possible example of a condition that is best regarded as disorder even though it is sustained by a distinct cause, is where someone is in a deeply confused condition due to sensory deprivation. It is not implausible to conceive of this condition as pathological. The example is Helge Malmgren's.

[97] In the somatic case, Culver and Gert suggest that e.g. poison, viruses, and germs should not be regarded as distinct parts of the person if they are biologically integrated (ibid., p. 73). In a similar way, it can be argued that a person's beliefs are parts of this person if they are "mentally integrated", e.g. if it is hard for the person to give them up.

[98] The suggestion is Frank Lorentzon's.

To conclude, it seems reasonable to assume that as a rule, those conditions that are dependent on some continuing environmental state are not disorders. For example, it is generally the case that the course of a disease is relatively independent of external influences. However, Culver and Gert do not draw the line between the pathological and the non-pathological where most people intuitively want to draw it. Their theory might be considered too exclusive and too inclusive at the same time, e.g. because it seems to imply that "normal grief" and fanatical beliefs should be classified as disorders, whereas panic attacks should be excluded from the category of mental disorder. That is, the theory does not fully satisfy the ordinary language condition. To some extent, it also fails to help us decide e.g. who is entitled to health care or compensation, at least if we assume that e.g. phobics are entitled to health care whereas e.g. religious fanatics are not. It is likely that Culver and Gert's analysis satisfies the other desiderata to a high degree, however.

Conclusions

We can now summarize the chapter on the factual component of "mental disorder" (i.e. on what kind of internal cause that is "essentially" involved in mental disorder) as follows:

(A) Wakefield's idea that disorders are, by definition, caused by mental dysfunctions (defined in evolutionary terms) is a well-founded theory, but it tends to exclude some consensual disorders from the category of mental disorder.

(B) The most promising alternatives to the scientific type of dysfunction analysis are (i) the idea that a mental disorder is an undesirable (e.g. harmful) condition caused by a dysfunction in a mental Cummins-function, and (ii) the idea that a mental disorder is a harmful condition that are caused by some kind of mental internal episode, where this is taken to imply that the condition does *not* have a distinct sustaining cause.

Both these ideas suffer from certain weaknesses, however. For example, (i) presupposes that we can conceive of the mental apparatus as a "Cummins functional system" and that we can attribute some goal or goals to this system. The main problem with (ii) is rather that it tends to include some more or less consensual non-disorders (e.g. "normal grief", fanatical beliefs, and unrequited love or infatuation) in the category of mental disorder.

(C) In my view, this strongly suggests that the different desiderata for a good definition (cf. pp. 18-23 above) are pulling in different directions. For example, it seems that if we rely on our linguistic intuitions (i.e. if we give considerable weight to the ordinary language condition), then we probably have to give up the idea that "mental disorder" can

be defined in terms of necessary conditions that are jointly sufficient (e.g. we have to sacrifice the theory condition and the simplicity condition). In fact, it seems likely that if we want to capture the concept of mental disorder as it is actually used in everyday speech, we should probably settle for some kind of family resemblance analysis or prototype analysis.

This observation is highly consistent with one of the main conclusions of chapter three, which concerned the evaluative component of "mental disorder". One of the basic assumptions in this essay is the idea that mental disorders as undesirable conditions that are caused by internal mental factors. On this view, the concept of mental disorder does not just have a factual (or explanatory) component – a condition can, by definition, not be a disorder unless it has the right kind of internal immediate or proximate – but also an evaluative component – a disorder is, by definition, undesirable. In chapter three, we saw that the evaluations we have to rely on to distinguish the pathological from the non-pathological are mainly of the form "this condition is harmful for the person". We also saw that if we want to specify the class of mental disorder in a way that is consistent with ordinary language, we also need to make use of other kinds of evaluations, viz. judgements about what is harmful for others and judgements about abnormal functioning. That is, the idea that several kinds of evaluations are relevant in this context is more in line with ordinary language than e.g. the idea that considerations of harm are the only relevant kind of evaluation. However, the latter idea has other advantages, e.g. it gives us an evaluative content of "mental disorder" that satisfies the simplicity conditions and the reliability condition to a higher degree. In short, as far as the evaluative content of a definition is concerned, there seems to be quite a tension between e.g. the ordinary language, on the one hand, and e.g. the reliability and simplicity conditions, on the other (cf. conclusion (C) on pp. 41-42 above).

In short, regardless of whether we have the value component or the factual component in mind, we can observe that if we rely heavily on our linguistic intuitions (i.e. if we give considerable weight to the ordinary language condition), then we most probably have to give up the idea that "mental disorder" can be defined in terms of necessary conditions that are jointly sufficient, and settle for e.g. some sort of family resemblance analysis instead. However, we might also give less weight to the ordinary language condition, e.g. because we want a coherent and precise concept or because we want a concept that is normatively relevant to a high degree.

As far as the evaluative component is concerned, there is (in my view) only one way to go if we decide to put less emphasis on our linguistic intuitions, viz. "the way of harm". That is, we simply have to accept the idea that considerations of harm are the only kind of evaluation we have to rely on to distinguish the pathological from the non-pathological. Concerning the factual component of the concept, there are (as far as I can see) two possible ways to go if we decide to give less weight to the ordinary language

condition, viz. “the scientific way” (e.g. to develop the dysfunction view further) or “the holistic way” (e.g. to improve on the idea that a mental disorder is a harmful condition that does not have a distinct sustaining cause). The first approach would probably give us a more narrow concept of disorder of high normative relevance, while the second approach would probably give us a broader and more generous concept that is of less normative relevance. As far as I can see, these alternative approaches are just as respectable as “the ordinary language approach”, probably even more so.

References

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV). Washington, DC: APA.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders,* 4th ed., text revision (DSM-IV-TR). Washington, DC: APA.

Blackburn S. (1994). *The Oxford Dictionary of Philosophy*. Oxford: Oxford University Press.

Block N. (ed.). *Readings in Philosophy of Psychology, Vol. 1-2*. Cambridge, Mass.: Harvard University Press, 1980-81.

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

Boorse C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44: 542–573.

Brülde B. (2000a). On How to Define the Concept of Health: A Loose Comparative Approach. *Medicine, Health Care and Philosophy*, 3: 305-308.

Brülde B. (2000b). More on the looser comparative approach to defining 'health': A reply to Nordenfelt's reply. *Medicine, Health Care and Philosophy*, 3: 313-315.

Brülde B. and Tengland P-A. (2003). *Hälsa och sjukdom – en begreppslig utredning*. Lund: Studentlitteratur.

Culver C.M. and Gert B. (1982). *Philosophy in Medicine. Conceptual and Ethical Issues in Medicine and Psychiatry.* Oxford: Oxford University Press.

Dennett D.C. (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press.

Hacking I. (1995). *Rewriting the Soul. Multiple Personality and the Sciences of Memory.* Princeton: Princeton University Press.

Hacking I. (1999). *The Social Construction of What?* Cambridge, Mass.: Harvard University Press.

Jablensky A. and Kendell R.E. (2002). Criteria for Assessing a Classification in Psychiatry. In: Maj et al (eds.). *Psychiatric Diagnosis and Classification*. Chichester: John Wiley, 1-24.

Kendell R.E. (1975). The concept of disease and its implications for psychiatry. *British Journal of Psychiatry*, 127: 305-315.

Kendell R.E. (2002). The Distinction between Personality Disorder and Mental Illness. *British Journal of Psychiatry*, 180: 110-115.

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

Maj M., Gaebel W., López-Ibor J.J., and Sartorius N. (eds.) (2002). *Psychiatric Diagnosis and Classification*. Chichester: John Wiley.

Malmgren H. (1984). *Några reflexioner kring sjukdomsbegreppet*. Göteborg: Filosofiska institutionen.

Megone C. (2000). Mental Illness, Human Function, and Values. *Philosophy, Psychiatry, and Psychology*, 7: 45-65.

Murphy D. and Woolfolk R.L. (2000a). The Harmful Dysfunction Analysis of Mental Disorder. *Philosophy, Psychiatry, and Psychology*, Vol. 7, No 4, December 2000, pp. 241-252.

Murphy D. and Woolfolk R.L. (2000b). Conceptual Analysis versus Scientific Understanding: An Assessment of Wakefield's Folk Psychiatry. *Philosophy, Psychiatry, and Psychology*, 7: 271-292.

Nettleton S. (1995). *The Sociology of Health and Illness*. Cambridge: Polity Press.

Nordenfelt L. (1995). *On the Nature of Health. An Action-Theoretic Approach*, 2nd revised edition. Dordrecht: Kluwer.

Nordenfelt L. (2003). On the Evolutionary Concept of Health. Health as Natural Function. In: Nordenfelt L. and Liss P-E. (eds.). *Dimensions of Health and Health Promotion*. Amsterdam: Rodopi Press.

Pilgrim D. (2002). Personality Disorder (Correspondence). *British Journal of Psychiatry* (2002), 181: 77-78.

Radovic F. (forthcoming). Natural functions. (working title)

Reznek L. (1987) *The Nature of Disease*. London: Routledge.

Scadding J.G. 1967. Diagnostics: the clinician and the computer. *Lancet*, ii: 877–882.

Sedgewick P. (1981). Illness – Mental and Otherwise. In: Caplan S.L., Engelhardt H.T., and McCartney J. J. (eds.). *Concepts of Health and Disease: Interdisciplinary Perspectives*, Reading, Mass.: Addison-Wesley, 119-129.

SOU 2002:3. *Psykisk störning, brott och ansvar*. Betänkande av Psykansvarskommittén.

Stich S.P. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, Mass.: MIT Press.

Svensson T. (1990). *On the notion of mental illness: Problematizing the medical-model conception of certain abnormal behaviour and mental afflictions*. Linköping: Linköping Studies in Arts and Science, 54.

Szasz T. (2000): Second Commentary on "Aristotle's Function Argument". *Philosophy, Psychiatry, and Psychology*, 7: 3-16.

Tännsjö T. (1999). *Coercive Care. The ethics of choice in health and medicine*. London: Routledge.

Üstün T. B., Chatterji S. and Andrews, G. (2002). International Classifications and the Diagnosis of Mental Disorders: Strengths, Limitations and Future Perspectives. In: Maj et al (eds.). *Psychiatric Diagnosis and Classification*. Chichester: John Wiley.

Wakefield J.C. (1992). The Concept of Mental Disorder. On the Boundary Between Biological Facts and Social Values. *American Psychologist*, 47: 373-388.

Wakefield J.C. (2000a). Aristotle as Sociobiologist: The "Function of a Human Being" Argument, Black Box Essentialism, and the Concept of Mental Disorder. *Philosophy, Psychiatry, and Psychology*, 7: 17-44.

Wakefield J.C. (2000b). Spandrels, Vestigal Organs, and Such: Reply to Murphy and Woolfolk's "The Harmful Dysfunction Analysis of Mental Disorder". *Philosophy, Psychiatry, and Psychology*, 7: 253-269.

Whitbeck C. (1981). A Theory of Health. In: Caplan S.L., Engelhardt H.T., and McCartney J.J. (eds.). *Concepts of Health and Disease: Interdisciplinary Perspectives*, Reading, Massachusetts: Addison-Wesley, 611–626.

Wilkinson S. (2000). Is 'Normal Grief' a Mental Disorder? *The Philosophical Quarterly*, 50: 289-304.