Rethinking the dysfunction criterion for medicine and psychiatry

Kelly Alexandra Roe

Philosophy Program
Research School of Social Sciences

Australian National University

Contents

Introduction

Mental disorders are often considered to be a major health problem. It has been estimated that one in four adults will experience mental illness at some point in their lives. The economic and social cost is thought to be greater and an even greater number of individuals are indirectly affected.

There are many controversial issues around the accurate diagnosis, nature, and treatment of mental disorder. For instance, some theorists maintain that there isn't any such thing as mental disorder and advocate doing away with the whole institution of psychiatry (e.g., Szasz). Others maintain that mental disorders are far more prevalent than is commonly supposed and they argue for its increasing prioritization in health budgets and health insurance plans. This thesis will attempt to come to grips with aspects of this controversy, hopefully in an enlightening way, so that we are better positioned to see our way forwards.

One of the issues that philosophy has traditionally been concerned with is the relationship between words, concepts, and stuff in the world. For example, we have the English word 'water', we have the idea, or concept of WATER, and we have the watery stuff out there in the world that the word or concept picks out or denotes. There is a fairly standard story as to how these three things relate - though there are many controversies over the finer details of an adequate account.

More in particular, this thesis will focus on words such as 'bio-medical disorder', 'mental disorder', 'schizophrenia', the concepts or meanings of those words, and the phenomena in the world that the terms and concepts refer

to (when all goes well). The basic idea is that work that has been done in philosophy on the word-concept-world relation can be usefully applied to the philosophy of psychiatry. I also believe that the philosophy of psychiatry presents an interesting case for philosophy as it shows certain complications to arise for the standard view that don't seem to have been extensively considered thus far.

For example, while scientific discovery is rightly thought to play an important role, I will argue that the role of concepts and conceptual analysis has largely been undervalued. Concepts are cheap, however, and which concepts we *should* adopt is best viewed as being crucially dependent on what it is that we aim to do with those concepts. Medicine in general and psychiatry in particular is an interesting case for so obviously being an applied science. Whether or not an individual is believed to be disordered is often thought to have important implications for that individual. I think that we need to face up to the fact that a number of decisions need to be made about what it is that we value and where it is that we wish to head as a society. An appropriately sensitive conceptual analysis needs to take due consideration of these applied issues.

Much of the literature in the philosophy of psychiatry has focused on answering the fol- lowing three questions:

- 1. What is the difference between the presence and absence of biomedical disorder?
- 2. What is the difference between having a mental and non-mental disorder?
- 3. What (if any) kinds of mental disorders are there?

The standard way of conceptualizing these questions is the medical model whereby they progress from general to specific. According to the medical model the disordered individuals are a subset of all individuals. The mentally disordered individuals are a subset of the disordered individuals. The individuals with a particular kind of mental disorder are a subset of the men-

tally disordered individuals. Another feature of the medical model is that the nature of mental disorder is thought to be something that exists independently from and prior to its application. Theoretical science is thought to inform us about the nature of stuff and then the applied sciences put that information to special use for various projects (e.g., medicine, conservation, engineering).

One way of understanding the first question is to view it as being focused on delineating the difference between health and sickness. Putting the issue this way is controversial, however. There are a variety of intuitively related or synonymous notions including but not limited to: 'disease', 'dysfunction', 'malady', 'abnormality' or 'illness'. One theorist might attempt to define or offer a conceptual analysis of one or more of these terms and perhaps offer an account of subtle differences between them. Problems arise when different theorists define their terms differently or offer different conceptual analyses of the same term. It can be hard to figure out whether theorists are genuinely disagreeing on substantive issues or whether theorists are simply talking past one another because they have defined their terms differently. One of the things that I am very concerned to do here is to avoid purely verbal disagreement and try and make some progress on the substantive issues. As such I shall follow the current convention (that isn't without its critics) of using the term 'disorder' throughout. I also shall not attempt to further define or analyze any of the terms of the debate. Rather, I will try and restrict myself to a meta level of analysis in an attempt to remain abstract enough to get to the heart of the debate. The reader can judge whether this strategy is or is not successful in avoiding verbal dispute.

The second question focuses on the relationship between mental disorder and bio-medical disorder. The medical model conceptualizes mental disorders as a subset of bio-medical disorders just as immunological, cardio-vascular, and respiratory disorders are sub-sets of bio-medical disorder. So this question can take the form of focusing on the relationship between psychiatry and the rest of medicine. On this understanding the question may be restated as 'in virtue of what does an individual have a *mental* (as opposed to non-

mental e.g., cardiovascular) disorder?' This question can also take the form of focusing on the relationship between psychiatry and non-medical fields such as social work, education, and psychology that seem intuitively to be concerned with the same subject matter (of mentally disordered individuals) even though these other disciplines are not themselves regarded to be specialist fields within medicine. On this understanding the question may be restated as 'in virtue of what does an individual have a mental *disorder* as opposed to a mental condition (e.g., grief or sadness). An alternative to the medical model is the view that mental disorders are not kinds of bio-medical disorders at all, which is to deny that mental disorders are a sub-set of bio-medical disorders more generally. On this view there isn't a difference (that we might find intuitive) between conditions like grieving and depression, where neither of them are correctly thought to be bio-medical disorders.

The third question addresses the issue of natural kinds. There is much current debate over whether mental disorders are best thought of as categorical where there are objective facts about which type or kind of mental disorder a person has, or whether there aren't any kinds of mental disorder at all. This denial of kinds of mental disorder need not take the form of eliminativism about mental disorder or about psychiatry's status as a branch of medicine. Rather, it may be focused on mental disorders as being extreme variants on a continuum rather than being categorical in nature. Something like the continuum view has become popular with respect to what are currently regarded 'personality disorders' for instance. The idea is that people with personality disorders are simply display extreme versions on certain traits rather than exhibiting categorical difference from non-disordered individuals. While we might be inclined to think that there is a simple objective fact of the matter about whether an individual has a certain condition like diabetes, cancer, or hepatitis, it might be that mental disorders are not so amenable to categorical analyses.

Answers to this third question relate back and may even be considered prior to the other questions insofar as we think that what makes it the case that an individual is disordered or mentally disordered is that they have a type or

kind of 'condition' which is a (mental) disorder. Throughout I shall use the term 'condition' to refer to such things as cancer, depression, attention deficit, being homosexual, being pregnant, and voting democrat. The thought is that only some of these 'conditions' (in this neutral sense) are disorders and the aim is to get clearer on which (if any) are and why.

As is traditional in philosophy theorists often attempt to answer these questions by attempting to analyze certain concepts. In this case, concepts such as DISORDER, MENTAL DISORDER, and DEPRESSION. Also other concepts mentioned previously such as ILLNESS, MALADY, SICKNESS, HEALTH etc. The aim has traditionally been to provide individually necessary and sufficient conditions for the correct application of the concept to phenomena in the world. While it is only too easy to get caught up in the process of offering definitions of our terms or concepts or attempting to provide counter-examples to motivate revision of proposed definitions I think it will ultimately be more profitable to take a step back from this process.

A number of philosophers came to be skeptical of the utility of conceptual analysis and they thought that what had come to be known as 'Gettierology' was giving epistemology a bad name. We had the Wittgensteinean notion of a game that cast doubt on the utility of traditional conceptual analysis. The idea of family resemblance concepts that didn't have necessary and jointly sufficient conditions. We started to turn to issues of what concepts were, and then issues of caring less about our concepts and more about the nature of the world. People become interested in scientific notions and natural kinds. One of the virtues of adopting a more meta perspective is that it becomes apparent that while there is a great deal of controversy over terminology and particular details of proposed definitions the main source of tension for psychiatry can be traced back to differences of opinion as to the relative contribution of the following two lines of inquiry:

- 1. Scientific discovery
- 2. Normative considerations

This tension involves whether what makes a person disordered is more a

matter of some- thing being objectively, scientifically discoverably 'wrong' with them, or whether it is more a matter of society judging the person to be violating certain kinds of (yet to be specified) norms in that they and / or we would be better off if we changed the individual. While some theorists maintain it is solely a matter of one or the other the standard view is that both of these aspects play a role. The main problem has been to offer an account of the relative contribution of each and get clearer on what that entails for the relationship between mental and non-mental disorder, for psychiatry's status as authoritative when it comes to the diagnosis and treatment of mental disorder, and for psychiatry's status as a specialist field within medicine.

I will offer an account of this tension here and attempt to motivate the idea that there is much to be gained by considering conceptual analysis and applied science as distinctive aspects with important roles to play. So ultimately I will offer a model of the inter-relationship between the following four areas of investigation:

- 1. Pure science
- 2. Applied science
- 3. Normative philosophy
- 4. Conceptual analytic philosophy

I will not attempt to define mental disorder or to provide individually necessary and jointly sufficient conditions for the correct application of the concept. Such attempts, I maintain, risk defining mental disorder in relatively arbitrary ways and also begging the question by assuming rather than motivating a view on the relative contribution of scientific discovery and normative judgement. Such attempts also give insufficient attention to the conceptual issues and also to the applied aspects of science. The idea is that we will come to a better understanding of where we should look for criterion that determine whether a person has a disorder, whether they have a mental disorder, and what kind of disorder they have. We will have a better idea of

the division of labour and potential contributions of theorists involved in the above four lines of inquiry.

Again, I think that a number of decisions need to be made around which concepts we adopt for the field of psychiatry. We are in the position of being able to choose our own adventure but in order to choose wisely we need to understand something of the potential consequences of our decision. Which concepts we have in fact is to a significant degree an empirical issue. Which concepts we have in fact is largely uninteresting, however, compared to the issue of which concepts we are best to adopt. I will attempt to motivate this issue here.

The thesis I wanted to write was one that grounded psychiatry firmly in bio-medicine which in turn was grounded firmly in the biological sciences more generally by way of evolutionary theory. Much work has been done on this already (e.g., Neander) and I hoped to contribute to this cause. Unfortunately, the more I looked into it the more trouble I ran into with the whole grounding project. I have come to doubt that psychiatry can be successfully grounded in the natural sciences - and especially evolutionary biology. Far from concluding 'so much the worse for psychiatry' as many other theorists have been led to conclude I really can't see how the rest of medicine fares any better, however. In other words, what started as an attempt to show psychiatry to be just like medicine (successfully grounded in biology) has resulted in the realization that medicine is just like psychiatry. Perhaps the thing to do is not so much to conclude 'so much the worse for psychiatry' as to conclude 'so much the worse for medicine'.

While one conclusion that we could take from this is the radical revisionist project of doing away with the institutions of medicine and psychiatry more moderate responses are possible. Indeed our work to help individuals with HIV and cancer is important and I'm not at all arguing otherwise. Our values play a much greater role than is commonly acknowledged, however. Epistemic norms or values. But also, and especially, (I'll argue) with respect to the metaphysics. We need to face up to this and work on sorting it out (hold it up for dialogue and debate). Rather than thinking that science

will sort it out. We need to appreciate that this isn't business as usual for science and current and future answers given by scientists are likely to assume answers to the questions that really interest us.

Is attention deficit really a disorder or are these individuals some combination of energetic, excited, bored? is individual x really mentally disordered or is there nothing wrong with them? Answers to these questions matter. They can matter a great deal. Access to treatment. Access to a label for self / other understanding. Reduction of guilt / responsibility. Stigma. Depersonalization.

The main thing I want to say here is that it is a very popular conception that we just need to sit back and wait for the verdict that science will deliver us. I think that this is mistaken, however. I don't see what scientific finding could settle this issue - though I do of course think that a number of considerations come to bear and that scientific findings are important considerations. They are not the only relevant considerations, however. The normative aspect. We need to engage in debate and critique. We need to understand the social consequences of this. We need to face up to decisions.

Post-script, 2023. In more recent years we have seen medicine become more like psychiatry as the anti-psychiatrists conceived of it. The 'new phrenology' seems to be less about neuro-psychology and neurology and brain science and more about imaging in general. Once we were concerned with the social construction of mental illness. Now we have come to be concerned with the social construction of viral illness, such as Covid. Instead of locking people up when they haven't done anything wrong in the name of psychiatry we lock people up when they haven't done anything wrong in the name of public health. It does not seem to be science that is making decisions. It is hard to see how science could tell us that these are decisions that are best to be made. People have been consented for injections where manufacturers liability has been waived. Things seem to be more about politics and political control than about science, or about values, or about caring for persons. Indeed, the entire issue of chapter one of this thesis pre-supposed a basic morality that is mostly lacking. Apparently the world is hostile and

competitive and there is no such thing as co-operation merely self-interested co-ordination because co-operation never could have evolved. And of course there aren't any triangles or squares. It's just about targeting individuals because you are getting away with it and why be moral? Who is going to make you? Who and whose army? Sterelny would say that we've been taken over by psychopaths. Over-taken by psychopaths. Well, then, who am I to disagree? Selfish selfish selfish some people are just that selfish. And there we go. They would have things be no other way. Everything for them and their selfish selfish selfish genes.

Chapter 1

Defining 'mental disorder': Why should we care?

Chapter introduction

A plausible thought often expressed in the seminar room is that it is rather pointless to ask and attempt to answer questions of the form 'what is x?' without firstly saying something about the use to which the notion is to be put in both theory and practice. Indeed, without some idea of the role that the notion is supposed to play it is hard to see what considerations would come to bear on assessing the adequacy or inadequacy of an analysis. The majority of theorists who are involved in attempting to analyse the notion of 'disorder' or of 'mental disorder' seem to appreciate this when they start out attempting to motivate why we should care about whether a condition is a disorder or whether an individual counts as being disordered. They then go on to offer an analysis that is supposed to determine whether a condition really counts as a disorder or whether an individual really counts as being disordered. Despite this, it is often considerably less clear how the account that they offer measures up with respect to the criterion of adequacy that they initially embraced in order to motivate their account in the first place.

If theorists disagree too radically on the role that the notion is supposed to

play then it starts to look increasingly like theorists are simply talking past each other rather than radically disagreeing. This concern is particularly pressing when considering notions such as 'bio-medical disorder' and 'mental disorder' that are employed by a diversity of theorists in a diversity of ethical, legal, social, medical, scientific, and philosophical contexts. While one option is simply to say that there are different notions in play in these different contexts this seems unsatisfactory insofar as we have the intuition that there is genuine disagreement between these theorists, however.

In this chapter I wish to follow suit insofar as I'll begin by introducing why it is that we should care about the notion of disorder and what is supposed to be at stake in whether an individual is regarded as disordered or a condition is regarded to be a disorder. I will show how scientific, normative, and conceptual issues have arisen around the notion of 'mental disorder'. While much work focuses on one of these aspects above the others I will use this chapter as groundwork for making a case for each of these considerations having a role to play in a criterion of adequacy on an account of mental disorder in chapter two. While I won't attempt to offer a definitive criterion of adequacy or an account I will attempt to offer something diagnostic with respect to locating the areas of genuine compared to purely verbal disagreement. These considerations are best appreciated when viewed in the historical context in which they emerged and thus this chapter will largely be a historical explication.

1.1 The birth of psychiatry

1.1.1 Disorder vs non-disorder

People have been interested in helping those who are suffering well before the rise of medicine as the institution that we know of today. Delusions, hallucinations, seizures, depression (especially catatonic), and mania have been described since antiquity. Davidson and Neale (2001) discuss instances of what would now be regarded as psychopathology in the early Chinese, Egyptians, Babylonians, Greeks, and Hebrews, though these people attributed

the cause of mental disorder to Gods or possessing spirits. In the fifth century B.C Hippocrates classified mental disorders into mania, melancholia (or depression) and phrenitis, or brain fever. He thought these disorders were caused by imbalance in the humours. We have Moses seeing burning bushes in the old testament and Jesus casting out demons in the new. Cases like this have been attributed to demons in the dark ages and witches in the middle ages. The Cartesian 'mental' 'non-mental' distinction that is so familiar to us today is a relatively recent one. People have been described as suffering from afflictions we would now regard to be 'mental' for as long as they have been described as suffering from afflictions we would now regard to be 'non-mental'. For instance, Ellenberger (1970, p.5) presents an early taxonomy of disease on the basis of aetiology (cause): Disease-object intrusion, Loss of the soul, Spirit intrusion, Breach of taboo, Sorcery. He also presents an early taxonomy of cure for the diseases: Extraction of disease object, to find bring back and restore the lost soul, exorcism, mechanical extraction of foreign spirit, transference of foreign spirit into another living being, confession and propitiation, and counter-magic.

Medicine gradually came to distinguish itself from alternatives as the 'winning team'. Relative success of medical treatments compared to treatments by priests and the like. Ellenberger on mechanisms and treatments that could be provided by any trained practitioner vs. the idea that there is something special about the particular practitioner (e.g., a priest gifted with the power to heal or exorcize). Medicine came to develop as an institution. To develop standards for induction to become a doctor. To develop standards of treatment and care.

1.1.2 Mental vs non-mental disorder

Edward Shorter (1997, p.8) places the birth of psychiatry as a specialist field within medicine in the late 1700s in the context of a more general trend towards specialism in medicine. He also places the birth of psychiatry as being due to the notion that confinement could be therapeutic and the the rise of the institution. While institutionalization or hospitalization might be

thought to be a fairly extreme form of management or treatment for mental illness to us today it is important to note that we still do utilize it in severe cases. While advances in psychiatric treatment might well have resulted in mentally ill individuals functioning at a higher level, current advances in psychiatry might well have resulted in individuals who would not have been regarded to be mentally ill currently being regarded to be so[1].

Psychiatry was originally concerned with what came to be known as the severe 'functional psychoses'. While this terminology is no longer part of official nomenclature a fairly rough guide is that the functional psychoses were thought to include disorders that would now be regarded as schizophrenia, severe mania and bi-polar, and catatonic depression.

In response to Focault's characterization of mentally ill individuals frolicking on the commons prior to the rise of the institution Shorter (1997, p.3) describes:

> In the 1870's just prior to introducing an asylum, officials in the French-speaking Swiss canton of Fribourg conducted a census of the mentally ill... One-fifth of the 164 mental patients they identified had been under restraint at home, mostly in unheated rooms and stables, "narrow, dark, damp, stinking lockups."... As Louis Caradec, a retired marine surgeon practising in Brittany, commented in 1860 of the surrounding countryside, "In our rural areas, where people are still imbued with absurd prejudices, public opinion sees having madness in the family as shameful and will not send the person to an asylum. This is the principal reason that motivates our peasants to keep such poor afflicted individuals at home. If the insane person is peaceful, people generally let him run loose. But if he becomes raging or troublesome, he's

[1]Perhaps mental illness is more prevalent today than it once was. Or perhaps we have relaxed the criteria such that more individuals get to count as being mentally ill. This makes a difference as to whether it is or is not likely that a person with a diagnosis of mental illness will or will not recover, does or does not need to be institutionalized, and is or isn't likely to pass the affliction onto future generations. I will have much more to say about this especially in the final chapter. The choices we make seem to matter.

chained down in a corner of the stable or in an isolated room...

Mental disorders were thought to be genetic defects that were passed on with increasing severity to future generations. Family members who did not suffer observable symptoms were still thought to suffer polluted blood lines. Many families went to considerable lengths to conceal that a relative was suffering mental disorder and thus external assistance was often sought only for the most severe cases when concealment became problematic. Individuals who were afflicted were often treated concealed at home by relatives or handed over to a doctor to care for them in an institution, private clinic, or even the doctors own personal residence.

With the stigma around diagnosis of psychiatric disorder on the one hand and the increasing numbers of particularly middle class individuals seeking treatment from medical practitioners on the other, neurology arose as a branch of medicine that was concerned with what came to be known as the 'functional neuroses'[2]. Prescriptions included bed rest and visits to spas that were thought to have curative mineral properties. Also prescriptions substances such as tobacco, snuff, cocaine, barbiturates, and opiates. While we now typically think of Freud as being concerned with the psychiatric conditions of post-traumatic stress, hysteria, and neurosis Freud trained as a neurologist and developed his talking cure psychoanalysis and hypnotism in the context of treating the traumatic stresses associated with war veterans suffering what used to be known as shell-shock (now post-traumatic stress) then hysterical blindness and paralysis (now somatoform disorders). Psychoanalysis filled a niche for intellectual understanding of the upper classes. Neurology and neurological disorder did not suffer from the stigma that psychiatry and mental disorder did. Differentiating the presence from absence of disorder arises un-

[2]'neurosis' and 'psychosis' are no longer part of official nomenclature. Such terms are thought to be excessively theory laden and aren't accepted by the majority of clinicians. It makes it hard to talk about early psychiatry where such terms are rife. It is also not straightforward to attempt to translate old diagnoses e.g., 'neurosis' into current diagnostic categories e.g., 'generalized anxiety' even when some theorists make a case that we have merely re-labelled the same phenomenon and / or that the presentation of the same disorder has evolved over time and / or that our (theoretical) understanding of the same phenomenon has evolved over time. More on this in the last chapter.

der such conditions. Unless we think that everyone who believes themselves to be suffering from a disorder is in fact we need to distinguish between those with a disorder and those without - otherwise known as the 'worried well'[3].

At a first pass mental disorders might be thought of as disorders of cognitive processes, such as thinking, emotion, or desire. Current classification regards cortical blindness as neurological rather than a psychiatric, however. This seems to be in line with our common-sense intuitions though it creates tension with the intuition that mental disorders are disorders of cognitive processes as vision would be a paradigmatically mental process. Indeed, other visual disturbances such as hysterical blindness and hallucinations are typically regarded as psychiatric rather than neurological and thus the concept of mental that is employed in common sense and in current nosology seems to be under-inclusive. Current nosology might also be thought of as over-inclusive. The essential feature of Tourettes is tics and there wouldn't seem to be anything particularly mental or cognitive about a motor disturbance even if it were caused by neurological malfunction. Perhaps Tourettes really has an essentially cognitive component that is neglected by current nosology, or perhaps Tourettes is not appropriately classified as a mental disorder and the current nosology is over-inclusive with respect to this case. The current distinction that is drawn between psychiatric or mental disorders and neurological or non-mental disorders is thus problematic. There are several things that we can do in the face of an inadequate concept, but first I want to turn to the main area of controversy, that of the nature of disorder, illness, or disease more generally as it is employed in both folk-psychology and in medicine.

Murphy (2006, find page ref) maintains that whether a condition is currently considered neurological or psychiatric is a matter of contingencies of history rather than due to any principled theoretical difference. He claims that none of the distinctions between the mind and brain currently on offer in the philosophy of mind, psychology, or cognitive neurosciences can provide a

[3]More on this later. By way of preview - everyone attends better when given amphetamines (aka: ADD / ADHD meds) but presumably not all of us have attention deficit disorder.

distinction that works to justify our currently regarding some conditions to be psychiatric rather than neurological and vice versa. PPP discussion on this. I think this is fairly persuasive. Perhaps we need to alter our current classification so that it does fall in line with some principled distinction or other. Perhaps we need to work harder to find the right level of abstraction at which we can describe a principled distinction that captures all and only the cases listed in the current version of the DSM. Or perhaps we need to ask ourselves: What difference does it make?

One difference that it might make is whether a person is seen by a neurologist or a psychiatrist. They utilize different screening tests and are familiar with different differentials. Neurologists focus typically on testing reflexes and order MRI's for the majority of their patients. Psychiatrists typically focus on mental state exams (questionnaires). The issue may be one of what kind of practitioner / what focus is most useful to the patient. The question then shifts as to why psychiatrists should be trained one way rather than another, however. Perhaps it is largely about how much curriculum you can get through in your induction to be one kind of specialist and this is (even if due to contingencies of history) the way things have gone in training and hence in subject matter.

There is a tension in that if psychiatry is too much 'just like medicine' then it may perhaps undermine itself. Will psychiatry survive if mental disorders are neurological disorders really? On the other hand there is tension in that if psychiatry isn't 'just like medicine' then it may undermine itself the other way. Will psychiatry survive if mental disorders are psychological disorders really, or problems of poor person / environment fit or problems of a messed up society? This leads into the turf wars: Who is authoritative when it comes to mental disorder? (Psychiatry / medicine currently) Who should be? Before we turn to this issue we will consider the present state of classification.

1.2 Classification of mental disorder

1.2.1 In the beginning: Chaos

In the early days for psychiatry nearly every psychiatrist had their own system of classification. Classification was very theory laden and involved assumptions about cause, course, and treatment. We have seen something of this already with many terms that are no longer part of official nomenclature because of too much theoretical dispute (e.g., psychosis, neurosis, hysteria, multiple personality disorder, shell shock, dementia praecox, manic depression).

Bentall (2003, p.46) relates that the World Health Organization attempted to bring unity to the field by expanding their manual the *International Classification of Diseases Index* to include non-fatal diseases for the sixth edition in 1951, but that it did not come to be widely accepted.

Bentall (2003, p.47) describes how a committee chaired by Erwin Stengal was mandated to investigate how consensus might be achieved:

> Stengels advice that diagnoses should make no reference to aetiology was fol- lowed for the eighth edition of the *International Classification of Disease*, which was published in 1965 and officially adopted by WHO in 1969. ICD-8 was the product of an unusual degree of co-operation between psychiatrists in different countries. Scandinavian and German psychiatric societies supported the new taxonomy, and the American Psychiatric Association agreed to base a revision of their *Diagnostic and Statistical Manual* on the ICD-8 system. Accordingly, in 1965 the APA appointed a small committee of eight experts and two consultants, and DSM-II was published three years later.

The DSM and ICD were thus both intentionally a-theoretic as to aetiology / cause / inner mechanism in order to be less controversial and more widely embraced by clinicians of different theoretical orientations.

1.2.2 Theory neutrality - behavioural symptom clusters

The ICD and the DSM are similar in the way that they distinguish between different kinds of disorders. Even though there are minor differences between the diagnostic categories they are designed such that translation between them is possible. They both focus on providing clusters of behavioural symptoms, or cognitive symptoms identifiable by verbal behaviour. When an individual has significant impairment in their functioning and they meet enough of the behavioural symptoms then the person may be regarded as having that particular kind of mental disorder. While some of the kinds of disorder have essential symptoms the majority do not, rather the person only need exhibit a certain number of symptoms. There are also exclusion criteria such that when an individual meets diagnostic criteria for more than one kind of disorder one diagnosis may take priority and exclude the other. There are other exclusion criteria as well, such as that the behaviour isn't caused by a general medical condition or the effects of a substance or toxin, or that the behaviour isn't performed solely as a matter of political protest or religious conviction. By sticking to observable behaviour and refraining from requiring certain aetiology or theoretical convictions (e.g., that ones social withdrawal be caused by overbearing mothers) the hope was that all clinicians could embrace the classification systems and unity could be achieved. Inter-rater reliability. Consensus over diagnosis.

Ian Hacking (1995, check page no.) maintains that even more important than the DSM definition of mental disorder and kinds of mental disorder the accompanying casebook. The case book provides numerous case studies of people who are prototypical instances of someone both being mentally disordered and meeting a certain diagnostic category (including commentary with differential etc). Clinical judgement may thus be thought to consist largely of experience with a variety of more or less prototypical cases so that a clinicians judgement falls in line with the judgement of other health professionals. Case studies form an important part of abnormal psychology and clinical psychiatry texts. Part of the process of initiation into medicine

in learning to diagnose similarly to other diagnosticians. Despite this inter-rater reliability (between trained clinicians) is still very poor (find reference - abnormal psychology textbook find primary study).

The current system is thought to be neo-Kraeplinean insofar as it favours empirical observation over theoretical constructs. Kraeplin had high hopes for empirical observation of behavioural symptom clusters:

> Judging from our experience in internal medicine it is a fair assumption that similar disease processes will produce identical symptom pictures, identical pathological anatomy and an identical aetiology. If, therefore, we possessed a comprehensive knowledge of any of these three fields - pathological anatomy, symptomatology, or aetiology - we would at once have a uniform and standard classification of mental diseases. A similar comprehensive knowledge of either of the other two fields would give us not just as uniform and standard classifications, but all of these classifications would exactly coincide.

Kraeplin, (1907) quoted in O. Reider (1974) 'The origin of our confusion about schizophrenia', *Psychiatry*, 37: 197-208 from Bentall.

> This was Kraepelin's big idea, announced tentatively in the second edition of the Compendium (renamed the Textbook of Psychiatry), which appeared in 1887, and which he elaborated until just before his death in 1926, soon after which the ninth and last edition was published. Mental illness fell into a small number of discoverable types, and these could be independently identified by studying symptoms, by direct observation of brain diseases, or by discovering the aetiologies of the illnesses (for example, by finding out whether they ran in families and were therefore determined by heredity). Of course, the only practical method of classification available at the time was by symptoms, as very little was known about the neuropathology or aetiology of psychiatric disorders. However, precisely because individuals with the same

> illness, defined by symptoms, were assumed to have the same brain disease, it could confidently be assumed that the identification of the illness would lead directly to an understanding of aetiology. On Kraepelin's analysis, therefore, the correct classification of mental illnesses according to symptoms would provide a kind of Rosetta stone, which would point directly to the biological origins of madness. (Bentall, 2003, p.12-13) .

So the thought is that scientific progress will be made by following the recommendation of Hempel. In the beginning there is chaos and not even a common language. Different people use terms differently and terms are highly theory laden (e.g., psychosis, neurosis). The first stage of science consists in (relatively though of course not perfectly) description of observations of empirical phenomena. In this case that consisted in Kraeplin's efforts to collect data on cases and to describe the course of symptoms for those cases. The next stage of science is thought to be the theory stage - where we can get at the underlying mechanisms and / or the causal aetiology of the phenomenon. This stage needs to occur after systematic observation, however, and cannot get up off the ground it if occurs to prematurely (as was previously the case in psychiatry). It is also thought that if we get the symptom observations right then this really will be the key because we can read off different kinds (where each kind has its distinct essence, distinct cause, and distinct course).

This is a very important picture. It is one that I will challenge significantly. It is one of the key notions that I will challenge, actually. It is also an important idea because it seems to be (though I guess it doesn't have to be) tied up with the idea that we can read off three independent stages - causes, constituents, and effects (just like how we can read off essential vs non-essential or constituents) from nature. This is a view (separating the theory from the application or the science from the application or the values from the science) that I will be concerned to undermine.

The current classification systems in psychiatry are commonly regarded as Neo-Kraeplinean. Bentall (2003) states that

> According to this paradigm psychiatric disorders fall into a finite number of types or categories (dementia praecox, manic depression, paranoia, etc.), each with a different pathophysiology and aetiology and that the way that they [psychiatrists and clinical psychologists] assign diagnoses and decide treatments for their patients, the way that they conduct their research into the causes of madness reveals that the Kraeplinian paradigm remains almost unchallenged within the mental health professions as a whole.

In support of this claim he cites four observations. Firstly, that modern textbooks of psychopathology are typically organised according to some variant of the Kraeplinean system with chapter headings on the different diagnoses. Secondly, the official diagnostic systems that are endorsed by such influential bodies as the WHO and APA are similarly organised. Third, most research is based on the paradigm in that the basic unit of research is typically the diagnoses on the assumption that individuals with the same diagnosis share something in common. Finally, clinicians typically employ Kraeplinean diagnostic concepts when explaining what is wrong (e.g., bi-polar) and when deciding on a treatment (therefore, lithium).

Sadock and Sadock (2003, p.288) also relate that:

> Advances in scientific psychiatry are to a great extent shaped by its system of classification. Systems of classification are fundamental to all sciences, containing the concepts upon which theory is based and influencing what can and cannot be seen. The classification of illnesses (nosology) has always been an integral part of the theory and practice of medicine.

On the commensurability of the DSM and ICD Index Sadock and Sadock (2003, p.288) relate:

> There was a strong consensus that diagnostic systems used in the United States must be compatible with the ICD to ensure uniform reporting of national and international health statistics. In addition, Medicare requires that billing codes for reimbursement

follow ICD ICD-10 is the official classification system used in Europe and many other parts of the world. All categories used in DSM-IV-TR are found in ICD-10, but not all ICD-10 categories are in DSM-IV-TR. The code numbers for disorders in DSM are fully compatible with ICD.

1.2.3 Defining (mental) disorder

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

Before offering a definition of mental disorder the American Psychiatric Association (2000, pp.xxx-xxxi) begins with some caveats.

> ...although this manual provides a classification of mental disorders, it must be admitted that no definition adequately specifies precise boundaries for the concept of "mental disorder". The concept of mental disorder, like many other concepts in medicine and science, lacks a consistent operational definition that covers all situations. All medical conditions are defined on various levels of abstraction - for example, structural pathology (e.g., ulcerative colitis), symptom presentation (e.g., migraine), deviance from a physiological norm (e.g., hypertension), and etiology (e.g., pheumococcal pheumonia). Mental disorders have also been de-

> fined by a variety of concepts (e.g., disease, dysfunction, dyscontrol, disadvantage, disability, inflexibility, irrationality, syndromal pattern, etiology, and statistical deviation). Each is a useful indicator for a mental disorder, but none is equivalent to the concept, and different situations call for different definitions.

There are several issues that are raised by this section of the DSM. Firstly, the APA is explicit about attempting to offer an operational definition that enables clinicians to identify which individuals are mentally disordered. The APA is also explicit about attempting to offer an operational definition that justifies which conditions are included in the DSM as mental disorders. Now, it might be the case that the features that we use to identify individuals and / or conditions are inessential to mental disorder similarly to how we might fairly reliably identify samples of water on the basis of colour and taste, for example, even though the essential feature is that it is composed of H2O. Despite a possible divergence between essential features of the phenomenon and characteristics which may be useful to enable people to identify, the DSM is explicit about being a handbook for clinicians. I shall return to the issue of the purposes of a classification system in a later section, but firstly I wish to consider the issue of definition in more depth.

While the APA states that the definition is provided because it was actually used to determine which conditions should appear in the DSM as mental disorders Rachel Cooper (2005, 2007) maintains that the definition was instead provided in the attempt to justify why certain conditions were included. Cooper has noted that the attempt to define mental disorder occurs at the time in which the APA was under considerable pressure from gay rights activists and anti-psychiatry lobby groups for the APA to justify how they decided that certain individuals / conditions were mentally disordered. In particular, the attempt was to ground psychiatry (or to justify psychiatry's status) as a speciality within medicine, and to directly counter the concern that psychiatry was in the business of confining and treating people who were merely in violation of social and moral norms as some anti-psychiatrists maintained. The DSM definition addresses this latter issue quite specifically

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

One thing to note about the DSM definition is that it maintains that dysfunction (or malfunction) is necessary for mental disorder. While there has been considerable controversy over the notion of disorder and related notions like disease, illness, sickness, disability, dysfunction etc the majority of the debate has been over which (if any) of these notions are thin (which is to say non-evaluative) and which are thick (which is to say they have an evaluative component). It is typically granted that mental disorder is a thick concept but a common line is to attempt to ground psychiatry as a specialist field within medicine by attempting to show that there is a thin aspect to mental disorder and to provide a non-normative account of that thin component. The DSM assumes that the relevant notion of disorder is one that is shared with general medicine and the APA point to a plurality in the notion of disorder that is employed in general medicine maintaining that the psychiatric notion is comparably pluralist (will return to pluralist views at some point).

The DSM can be viewed as being an example of a general approach that Murphy dubs the two-stage view. On the two stage view there are facts about whether an individual is malfunctioning that are thin concepts in the sense of there independence from our social, moral, or political norms. It is by appealing to the necessity of malfunction that the APA means to counter the claim made by some anti-psychiatrists that mental illness is solely a matter of norm violation. Murphy maintains that the majority view within psychiatry is a two-stage view where there is a non-normative thin notion that is relevant together with an evaluative aspect.

The American Psychiatric Association continues on:

> Despite these caveats, the definition of *mental disorder* that was included in DSM-III and DSM-III-R is presented here because it is as useful as any other avail- able definition and has helped to guide

> decisions regarding which conditions on the boundary between normality and pathology should be included in DSM-IV. In DSM-IV, each of the mental disorders is conceptualized as a clinically significant behavioural or psychological syndrome or pattern that occurs in an individual and that is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a behavioural, psychological, or biological dysfunction in the individual. Neither deviant behaviour (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.

The DSM definition is a version of a two-stage view according to which there is a non- normative element - dysfunction that is necessary. They are fairly liberal with the other criterion - could be normative could be non-normative. They say that it has been used to guide decisions as to what to include or exclude. Are they talking about homosexuality here? It is unclear how this definition justifies its decision to exclude homosexuality from the DSM.

It is important to note, though, that it isn't such a problem that there isn't an adequate definition. Biology textbooks need not start out by defining 'life' in a way that satisfactorily distinguishes it from non-living things. Can offer a classification of living organisms without having a satisfactory account of that distinction. When there is controversy over whether something should be included as living or as a disorder then it seems to matter, however. Insofar as this is a more salient issue for psychiatry and general medicine than it is for the biological sciences it seems more pressing to get the definition right.

The main systems of classification are thus provided by two health organiza-

tions - the WHO and the APA. While alternative classification systems have been offered their use is more limited to clinicians with a particular theoretical orientation. The *Psychodynamics Diagnostic Manual* will probably only be used by theorists and clinicians of a psychodynamic orientation, for example, and for billing and statistical purposes DSM or ICD codes must still be provided. The most prevalent view of mental disorders are thus that they are a certain kind of medical condition. The authoritative bodies who decide which conditions will be included and excluded from a classification of mental disorders are typically psychiatrists. The profession that is taken to be authoritative with respect to the diagnosis and treatment of mental disorders is psychiatry.

Diagnosis of mental disorder seems to consist of (as least) two interrelated components. Firstly there is the issue of how we identify whether or not an individual is mentally disordered, and secondly there is the issue of how we identify what particular kind of mental disorder they have. I shall address both of these in turn. With respect to the first issue of identifying mental disorder in general we can distinguish two further related problems. The first is how to distinguish a disorder from a problem in living, The second is the issue of how to distinguish mental or psychiatric disorders from non-mental, neurological disorders, or general medical conditions.

With respect to the first issue the DSM provides a global assessment of functioning (or GAF) scale that is meant to capture the extent of the disability, disorder, dysfunction, or distress. Without significant impairment in functioning a clinician should not diagnose an individual as having a mental disorder even if they meet diagnostic criteria for a particular kind of mental disorder. The GAF scale reflects the notion that the DSM is primarily concerned with providing a tool to enable clinicians to make diagnostic decisions. The DSM also lists the following features that clinicians are supposed to use to assess whether an individual has a mental disorder: statistical infrequency, violation of norms, personal distress, disability or dysfunction, and unexpectedness. With respect to unexpectedness Davison and Neale (p.6) maintain that, for example, an anxiety disorder is diagnosed when anxiety

is unexpected and out of proportion to the situation, as when a person who is well off worries constantly about his financial situation. The DSM takes this list to not only be a way of identifying individuals on whom to intervene, however, it takes it as an attempted definition of the nature of mental disorder, though it is acknowledged that current definitions are inadequate to capture the phenomenon that is of interest.

1.2.4 The aims of the DSM

The American Psychiatric Association (2000, p.xxiii) states that:

> The utility and credibility of DSM-IV require that it focus on its clinical, re- search, and educational purposes and be supported by an extensive empirical foundation. Our highest priority has been to provide a helpful guide to clinical practice. An additional goal was to facilitate research and improve communication among clinicians and researchers. We were also mindful of the use of DSM-IV for improving the collection of clinical information and as an educational tool for teaching psychopathology. An official nomenclature must be applicable in a wide diversity of contexts. DSM-IV is used by clinicians and researchers of many different orientations (e.g., biological, psychodynamic, cognitive, behavioural, interpersonal, family / systems). It is used by psychiatrists, other physicians, psychologists, social workers, nurses, occupational and rehabilitation therapists, counsellors, and other health and mental health professionals. DSM-IV must be usable across settings inpatient, out- patient, partial hospital, consultation-liaison, clinic, private practice, and primary care, and with community populations. It is also a necessary tool for collecting and communicating accurate public health statistics. Fortunately, all these many uses are compatible with one another.

They also carve off the above scientific projects from other extra-scientific projects when they state:

> It is to be understood that inclusion here, for clinical and research purposes, of a diagnostic category such as Pathological Gambling or Pedophilia does not imply that the condition meets legal or other non-medical criteria for what constitutes mental disease, mental disorder, or mental disability (xxxvii).

This carving off of the scientific concerns from the extra-scientific concerns is similar to Murphys take when he focuses his book on the scientific notion of mental disorder rather than the extra-scientific notion that comes up in issues to do with moral and legal responsibility. Murphy also places treatment as an extra-scientific concern. While I'm with the DSM in placing treatment as a scientific concern issues around when involuntary treatment are justified seems to have more to do with the notion of responsibility / rationality etc.

I don't think that we can separate out these two stages of theory that occurs prior to application. One of the things that I want to argue for is that the application is inexorably tied up such that it is partly constitutive of the phenomena (or something like this).

The APA has this to say:

> The three main aims that are provided are to firstly, be of use to clinicians so that they can identify and treat people with psychiatric disorders. Secondly, to be of use to researchers so that they can identify people with psychiatric disorders and investigate the causes of them and treatments for them. Thirdly (we may assume that this is derivative) to provide a classification system that can be used to compile health statistics.

In what follows I want to grant the DSM its stated aims. While one project would be to critique how much the DSM really is guided by these aims as opposed to other influences (e.g., political), and another project would be to critique the above mentioned aims as worthy of pursuing, neither of those projects is my project here. What I wish to do here is to take the stated aims as primitive and engage in a critical investigation of how the DSM can progress as a science given its stated aims. For my project I'll thus take the

stated aims of the DSM as primitive. One can thus read my thesis as an investigation of the hypothetical question "if the aims of the classification system are as the DSM states, then how can the DSM better move towards them?"

One of the issues I wish to consider is whether the three stated aims come into line as much as the APA regards them to. While the APA maintains that it is fortunate that clinicians and researchers can share a common classification system I think it is far from obvious that the system that it most useful for clinicians would be most useful for researchers. While clinicians need a map from identifiable features to treatments it might turn out to be the case that researchers need a classification system that diverges from this. Murphy and others have critiqued the DSM from focusing fairly exclusively on observable behavioural symptoms rather than on internal generative or causal mechanisms. It is unclear that a classification system that was based on internal generative or causal mechanisms would be optimally useful for clinicians, however, even though such a classification system could be optimally useful for researchers.

The DSM states that its third aim is to provide a classification system that facilitates communication between clinicians. Prior to the development of the DSM and ICD index there were a proliferation of nosologies that were very theory dependent on which variety of psychodynamic theory the theorist subscribed to. Part of the motivation from moving from a classification of inner causes to a classification of behavioural symptoms was that regardless of theoretical orientation clinicians could agree as to whether an individual exhibited this or that symptom. As the diagnostic categories are built out of behavioural symptoms this also allowed clinicians to agree as to what diagnosis a patient should have, regardless of the clinicians theoretical orientation. The issue here is thus one of inter-rater reliability. When a behavioural symptom or a diagnostic category has good inter-rater reliability then different clinicians would attribute the same symptoms and diagnostic category to the same individual. Both construct validity and inter-rater reliability would seem to be required in order for compilation of statistics on prevalence rates

to be meaningful.

These two aims of facilitating research and promoting communication between clinicians might be thought to map onto two different aims of providing a nosology that is scientifically fruitful with respect to generalisation and prediction and providing a nosology that is useful for clinicians with respect to identifying which individuals are requiring intervention. The DSM takes these aims to be complimentary and indeed they do seem to be related. One would hope that nosology is useful with respect to identifying what kind of disorder an individual actually has, for example, and one would also hope that a scientific nosology would provide information as to what kinds of interventions are likely to be effective. It might turn out to be the case that these aims diverge, however. While purely behavioural symptoms might be most useful with respect to identifying the individuals who require intervention purely behavioural symptoms might be less than optimal with respect to enabling us to identify the underlying causal mechanisms that provide information as to the optimal points of intervention.

The majority of research takes the diagnostic categories provided by the DSM as the basic unit of research analysis. When people search for a genetic basis, the structural or functional neurological abnormalities, the efficacy of medication or therapy, the cross-cultural variation, or the course of illness, the DSM criteria is used to identify the individuals with the disorder that is the subject of research. While it is important to distinguish clearly between the nature of disorder on the one hand and how we go about identifying individuals with the disorder on the other, the two are clearly related in the sense that we need to identify individuals in order to commence investigation into the generalisations and predictions that we can make about them as a group and our findings about individuals in the group could lead to subsequent revisions of the diagnostic categories.

The relevant notion here is the notion of construct validity. The DSM provides a list of constructs, or kinds of disorder. A construct is thought to be valid when there are scientific generalisations and predictions that can be made about an individual on the basis of identifying the individual as an

instance of the category picked out by the construct. As such, constructs can be more or less valid depending on whether they support more or less generalisations and predictions. The notion of a category that is in play here seems to be in line with Boyd's homeostatic property cluster theory of kinds where we note that there are observable properties (in this case behavioural symptoms) that are found to be clustered together in nature. Because these properties are found to be clustered together we can form a construct of the category and we can make fairly accurate generalisations from the presence of some properties, or symptoms, to the likely presence of some other properties, or symptoms. When we observe some of those properties, or symptoms we can also make fairly accurate predictions such as response to treatment or the future course of illness, for example. The homeostatic property cluster view might only be one way in which we could get projectability, however.

1.2.5 The turf wars. AKA: Introduction to politics

It is also important to note that while psychiatrists study and treat mental disorders and have special authority when it comes to the American Psychiatric Association having special authority with respect to diagnoses other fields are also interested in investigating and treating people with mental disorders. While the following is rough to be sure it is worth considering some of those other fields and the sorts of things that they take themselves to be doing. Different fields today are involved in researching and treating people with mental disorder. The 'turf wars'. Concern that some of these areas will be phased out. Concern that some of these areas are prioritized above others.

Psychiatry

Psychiatry is a specialist branch within medicine. Psychiatrists train as medical doctors and then go on to specialize in psychiatry in the same way that other medical doctors go on to specialize in neurology, oncology, or pediatrics. This being said, there is some debate about what is distinctive of a medical field - especially when people who treat individuals with disorders can

come from fields such as clinical psychology, social work, and education which are paradigmatically not medical fields and require no medical background to train in them. The typical view is that the medical model is committed to medical disorders being a certain kind of physical disorder (arising from dysfunction in the brain, for example). I shall call this the weak view of the medical model as it is one that is often endorsed by theorists outside the medical paradigm - though it is not without its critics as we shall see. The stronger view of the medical model would be that it is also committed to their best being treated with paradigmatically medical treatments such as hospitalization (or institutionalization) and fairly direct interventions to the physical dysfunction such as pharmaceutical or surgical interventions. This is a view that is less widely accepted - indeed it is most commonly held by psychiatrists and perhaps only held by other professionals with respect to the most severe 'functional psychoses' that we considered in the previous section.

If mental disorders are conceptualized as disorders of the brain then one might well wonder what the distinction is between neurology and psychiatry. The brain is composed of nerves and if mental disorders are neurological disorders in the sense of being disturbances of the brain then it would be hard to see what the distinction between neurology and psychiatry might be. While we have already seen that some disorders that used to be considered neurological came to be regarded as psychiatric and some disorders that used to be considered psychiatric came to be regarded as neurological it is much harder to come up with a principled reason why some conditions have been allocated (or re-allocated) one way rather than the other.

There have been a variety of attempts to come up with a principled way of distinguishing psychiatric disorders from neurological disorders. Some are attempts at apologetics in the sense that they aim to describe what cases that are currently regarded as one one the other have in common such that there is a principled reason to justify our classification practices. Other attempts are more revisionist in that they prescribe that some are inappropriately classified as one or the other. Some theorists maintain that there is no principled distinction in the subject matter of each field.

One way in which people have attempted to make the distinction is to say that mental disorders are understandable on the intentional level whereas non-mental disorders are not. This disagrees with Jaspers, however, who maintains that delusions proper are not understandable from the intentional level. Another way in which people have attempted to make the distinction is to say that neurology deals with comparatively peripheral neurological disturbances whereas psychiatry deals with comparatively central systems neurological / cognitive processes.

One way of getting some kind of grasp on the current distinction (which of course doesn't tell us anything about whether the current distinction is either defensible or well founded) is to consider what we might be told if we were unsure whether we wanted to specialize in neurology or psychiatry. There are paradigmatic cases that the fields deal with and the paradigmatic cases are different. This doesn't tell us what (if anything) the paradigmatic cases have in common, of course. But it is a start. We also have differences in the day to day practice of clinicians of both fields. Neurological assessments involve a lot of testing fairly peripheral reflexes to look for problems with fairly peripheral nervous functions. Psychiatric assessments involve a lot of asking questions to look for problems with thought and mood and daily activities. It has been estimated that around 80 percent of patients seen by neurologists have an MRI taken of their brain. The majority of neurologists see people with epilepsy or tumours to check for tumours or other gross abnormalities.(around 80 percent of neurology patients have one of the 'big five' conditions). Neurologists refer on to neurosurgeons for surgeries and prescribe medications that overlap with those of psychiatrists (e.g., anti-epileptics can be useful for mood stabilization and / or sedation). Psychiatrists refer on to neurologists for MRI's and do not do them as a matter of course. They do have more say over involuntary commitment or hospitalization. Neurologists tend to deal with sleep conditions even though they are in the DSM. Not many psychiatrists deal with intellectual handicap even though they are included. Also personality disorders, or addictions.

We might be tempted to maintain that mental disorders have mental symp-

toms either as the cause or as the manifestation or both. There seem to be clear examples of each, psychosomatic disorders seem to have a mental cause, mood disorders seem to have a mental manifestation (though may be caused by neurological disturbance). There seems to be a problem with developing a unified view, however.

Murphy (e.g., 2006) maintains that there is no tenable distinction between neurology and psychiatry at the end of the day and that psychiatry is a variety of cognitive neuroscience and as such should be integrated into it. Just because there doesn't seem to be a hard dividing line between them doesn't mean that there isn't a usefulness to the distinction, however. I'm not at all sure that Murphy is advocating that neurology and psychiatry should merge rather than remaining distinct medical specialities. Similar problems of justifying field divisions can occur with other medical specialities and this is not simply a problem for psychiatry. When should one see an ear nose and throat surgeon etc. It may be that there is a difference in practice rather than a difference in subject matter (as Murphy advocates) - though I think it is important to attempt to spell out the difference in practice given that we are indeed dealing with an applied field and whether an individual is mentally disordered or physically disordered is a controversial issue and what we are trying to do here is to get clearer on what is supposed to be at stake. The controversy does not seem to be over whether a person should be best treated by a neurologist or a psychiatrist?

While neurology used to traffic in these disorders which are now typically treated by psychologists (psychiatrists) it seems that many of the success cases in psychiatry are gifted to neurology. If you look at a textbook on how to conduct an assessment in psychiatry compared with neurology neurology seems much more focused on testing neurological reflexes involving the peripheral nervous system for the most part (e.g., Plantinger reflex). If the disturbance is to peripheral nerves and / or localized tumours to peripheral regions then things are well understood. Hence an attempted distinction between psychiatry dealing in 'higher' functions / dysfunctions. Central system processes. Concern that if psychiatry is neurology at base then might well

be subsumed. MRI scans.

Clinical psychology

Clinical psychology is a non-medical field that does not deal with differential diagnostics between psychiatric vs other medical conditions insofar as psychologists are not trained in the diagnostic of medical conditions. This is important because psychiatric diagnosis often have medical exclusionary criterion e.g., that the symptoms not be due to a medical condition. Psychology, as a field, focuses on psychometrics and empirically validated forms of talk therapy as may be distinct from other, though related, professions.

Scientific psychology arose as a distinct field in the early 1900's - much later than the time in which medical specialization was occurring. While early work focused on mental phenomena such as memory and individual differences in response times it wasn't too long before behaviourism became the ruling paradigm for psychology. The behaviourist paradigm conceived of psychology as the science of behaviour which was largely a reaction to psychoanalytic theories that had proliferated and where it was unclear what sort of evidence could be taken to support one of these theories over the other. The behaviourists were particularly interested in interventions that could alter behaviour - both normal and pathological - and a variety of effective interventions were found to deal with phobia and anxiety in particular. It is interesting that some of the early behaviourists started out being interested in nervous reflexes such as salivation. The cognitive revolution in psychology brought interventions of its own in the form of restructuring or re-framing unhelpful cognitions and beliefs. Psychology has also contributed much to the field of psychopathology especially with regards to the development of psychometric tests for intellectual handicap and the development of psychometric tests for both normal and abnormal personality traits.

Modern clinical psychology basically accepts the weak view of the medical model even though strictly speaking behaviourists can be neutral about it. Indeed Skinners version of utopia involved employing behaviourist techniques to shape and reinforce behaviour to what was regarded socially accepted

with little recourse to whether the behaviour was criminal undesirable or symptomatic of dysfunction. The behaviourists unwillingness to traffic in mental causes enabled them to remain agnostic with respect to the inner dysfunction view - though they may be regarded as holding that is was circular. Behaviourists were caricatured in novels such as A Brave New World precisely because there was a concern about whose values would be reinforced and ethical issues were brought to the fore with respect to altering criminal conduct in The Clockwork Orange. Clinicians who specialize in therapy and behavioural intervention often dismiss the strong version of the medical model, though not uniformly.

Another aspect to psychology is that of neuropsychology and cognitive neuropsychology. There is a move at present for some clinical psychologists with a background in neuropsychology to have additional training and to obtain limited prescription rights. This is something that is extremely controversial within medicine and psychology alike. Psychiatrists are protective about their prescription rights as this is a social practice that aligns them closely with medicine. It has been argued that the general medical background is necessary in order for a clinician to be able to correctly assess differentials, interactions between psychiatric and non-psychiatric medication, and to be adequately knowledgeable in issues of the effects of age and differences in metabolism of psychiatric medications. There is a concern within psychology that the strong medical model will reign supreme and that psychologists will be conceived of as second rate medical doctors rather than as first rate psychologists. There is a considerable politics tied up in this issue with respect to differences in salary as well where psychiatrists tend to make a great deal more than psychologists and the concern is that giving psychologists prescription rights would be a financial move more than anything else. While psychiatrists do sometimes provide therapy this is typically not state funded as it costs more to get a psychiatrist to deliver therapy than to do a medication review - and skeptics claim that the move to allow clinical psychologists to prescribe is similarly motivated by cost cutting reasons as it will cost less to employ a psychologist to prescribe than to employ a psychiatrist.

Another concern is that the efficacy of therapy intervention (of empirically validated forms) isn't being taken as seriously as it should be. Psychologists should thus work to promote these alternative forms of intervention rather than desiring to be part of what is considered the 'winning team' of medicine. Clinical psychology is in something of a state of identity crisis where there seems to be the threat of encroachment form psychiatry on the one hand (with the development of medications) and social work and education (in the form of counselling) on the other hand.

The Boulder Model of training in clinical psychology is a scientist-practitioner model one that is represented in PhD clinical psychology programs. See (Compas & Gotlib, 2002, p.18). The view is that clinical psychologists should be trained in research and in clinical practice as well. Some theorists maintained that a person who made a good research academic might make a poor clinician and vice versa. The upshot of this was the development of a PsyD program (doctor of psychology rather than philosophy) which would be a practice based program that could grant the qualification of clinical psychology. This issue has become political too. The most selective programs are clinical psychology programs and some doctor of psychology programs don't grant funding to their students. The PsyD programs tend to have broader scope for different varieties of therapy, however, and tend to be more eclectic in their focus.

There has been much controversy over how to characterise mental disorder. Textbooks in Abnormal Psychology often refer to features such as statistical infrequency (e.g., mental retardation), violation of norms (e.g., sociopathy), personal distress (anxiety, depression), disability or dysfunction, and unexpectedness. While clinical psychology isn't a specialist branch within medicine and as such need not commit itself to the disease model of mental disorder clinical psychology is often characterised as the study of abnormal psychology or psychopathology or mental disorder where the relevant notion of mental disorder is the one employed in psychiatry.

Clinical psychology regards itself as having the same subject matter as that of psychiatry as is reflected in both fields adopting the classification sys-

tems provided by the American Psychiatric Association (The Diagnostic and Statistical Manual of Mental Disorders) and / or the diagnostic system and / or coding provided by the World Health Organisation (The International Classification of Diseases Index). Textbooks in abnormal psychology and psychiatry are typically organised according to the different diagnostic categories endorsed by those systems, and both fields study the nature, course, and interventions of the categories endorsed by those systems. While clinical psychologists often characterise the subject matter as abnormal, deviant, or pathological rather than diseased, and while it is common for clinical psychologists to maintain that they reject the disease model of psychopathology, this denial seems to have more to do with resistance to biological reductionism than denial of there being a shared subject matter. It is controversial whether biological reductionism is entailed or implied by the disease model and yet concerns that such reductionism is entailed or implied seem to have been the driving force behind the American Psychological Associations threatening to sue the American Psychiatric Association if they characterised the subject matter of the DSM as mental disease rather than mental disorder.

Counselling, social work, education

Under this rubric we have humanistic / psychodynamically oriented (as opposed to psychoanalytic), and social change theories. We also have a consumer movement, though that is probably better kept distinct rather than tied to professions.

The empirically validated approaches that clinical psychologists are trained in (and the psychometric tests) have similarly become the defining feature of psychology that is thought to carve it off from related fields such as social work and counselling. The focus on running clinical trials to test efficacy and training therapists in the techniques that have empirically been shown to be validated is what is meant to distinguish clinical psychology from these related fields. Another thought is that while counsellors and life coaches deal in the 'worried well' of people having relationship difficulties and so on clinical psychology deals with severe psychiatric disturbance. Despite this there is

still a prevalent perception that some mental disorders are best treated pharmacologically (e.g., schizophrenia, psychotic disturbance, severe depression, etc) while others may be best treated with therapy and / or a combined approach (e.g., depression that isn't as severe / while the patient is in remission and so on). Life coaches and counsellors are thought to assist basically normal or well people to assist them when those individuals are not considered disordered. Despite this there is a blurring as social workers in particular attempt to trademark the term 'counsellor' as psychologists have trademarked 'clinical psychologist' and medical professionals have trademarked the term 'psychiatrist'. Many counsellors advertise as interested in dealing with issues that overlap with those that have traditionally been the subject matter of psychiatry and clinical psychology. Trauma, addiction, anxiety, depression, relationship counselling etc. Partly it is a difference in training and entry requirements.

Psychiatry is in a state of identity crisis and clinicians suffer from being considered 'second rate doctors' on the one hand, and 'second rate therapists' on the other.

There is controversy over whether psychiatry (with its emphasis on medication) should be the relevant authority as opposed to clinical psychology (for example) with its emphasis on behavioural intervention and therapy. When psychologists protest the APA and WHO setting the authoritative diagnostics manuals this might be seen to be a critique of the medical model insofar as the medical model grants priority to health professionals having the majority of control over the authoritative diagnostics manuals. Similarly when psychologists maintain that (some or all) individuals with certain kinds of mental disorders are best treated by therapy and behavioural interventions rather than by medication this might be seen to be a critique of the medical model. Indeed, one might wonder what makes a condition medical as opposed to something else? One answer might be that medical doctors are the authority on identifying and diagnosing which individuals are disordered. Another might be that medical doctors are the authority on treating individuals who are disordered. Still another might be that medical disorders are

biological disorders of the individual. While these three answers are often run together (and an acceptance of one is often taken to entail the acceptance of the others) I think that it is worth separating these issues out as distinct as it isn't fully obvious why one answer to one entails a particular answer to the others - even though this might ultimately turn out to be the case. Insofar as psychology agrees that mental disorders are biological disorders of the individual at base (as it tends to do) psychology may be characterized as basically accepting the medical model. The issues of which professional organization should have control over classification and which discipline is best placed to treat mental disorders might be better kept distinct.

While psychologists may be thought of as offering a critique of the medical model insofar as they question who the relevant authority should be and which variety of intervention is most effective, there are critiques of the medical model that run deeper than this. Some theorists have maintained that mental disorders are not a kind of medical (or biological) disorder at all and that psychiatry has more in common with law and other systems of social control than with medicine and biological science. While one might be tempted to disregard these views as too radical in the face of fairly obvious acceptance of the medical model insofar as it is biological, it is important to note that advocates of these views have had an important role to play in shaping current classification systems.

The title 'counsellor' is trademarked in much of the world the same way that the term 'psychologist' and 'clinical psychologist' and 'medical doctor' and 'psychiatrist' are. Master of counselling programs typically require an undergraduate degree in a related field such as psychology, education, or social work. These programs tend to be even more psychodynamically and humanistically based than the typical clinical psychology program.

The medical model is probably less accepted here than elsewhere. Partly due to the humanist focus, partly due to the broader focus in helping people - compared to fields who are more focused on mental disorders and a diagnosis of disorder.

Rape crisis counselling, domestic violence assistance, grief counselling, marriage counselling, tough love (teenage problems assistance), lifeline etc. Alcoholics anonymous (AA) and narcotics anonymous (NA) etc. Therapeutic communities (hippy commune treatment centres). Weekend workshops. Crystal healing. etc. Notion that one needs to have been through whatever oneself in order to help others through the same thing (an idea that is especially prevalent in addiction studies and to a lesser extent with respect to sexual abuse). More scope here to focus on the environment. Social work in particular has more scope to look into and assist with such things as employment, housing etc than traditional psychology (that doesn't focus on such things).

Life coaches, motivational therapists, 'therapist', self help, consumer support.

Which individuals are disordered? How do we tell (who gets to say) who gets to say what is most effective. What kinds (if any) of disorder are there? Moral / legal responsibility.

The consumer rights movement

I should probably say something about this. And the continental project. Experiential. It looks like this is the way that things are going at the moment. E.g., with what recently happened with the Oxford conference that went from an academic (largely male professor) guest speaker line up to being free where graduate student and consumer speaking was prioritized etc. I should say something about this... I think that perhaps things are moving too near the other extreme (that we are broadening the notion of 'mental illness' such that most everyone gets to have one which means that of course stigma gets less but that the end result is that those most in need of assistance might well not get it and treatment parity isn't likely to happen when it is (for example) cancer vs an individuals inability to focus for more than 20 minutes on a task (which is something that most people 'suffer' from). Of course we need a 'comparable severity' clause - but that is part of the problem, I think.

We have relaxed the criteria of mental disorder one hell of a lot. So much

that there isn't anything much wrong with people who are disordered. We shouldn't stigmatize people with mental disorder, we shouldn't discriminate against them (e.g., worrying about their occupational or social functioning) precisely because we have relaxed the criteria so much such that people who are functioning very highly indeed get to count as having a mental disorder. If we tighten up the criteria such that (once again) only the most severely functionally disturbed get to count as having a mental disorder then we would seem to rightly have concerns about the social and / or occupational functioning of individuals with mental disorder. The decisions we make (how broadly or narrowly we choose to define the concepts / how widely or narrowly we construe the phenomenon that gets to count) are crucially important with respect to determining the nature of the phenomenon. How SHOULD we choose to define our concepts / how broadly or narrowly we cast our net? That is something that we need to think about and discuss openly and honestly. But first: We need to face up to this being the reality of the situation rather than sitting back and saying 'there are facts of the matter about who is disordered and who isn't, about what is true and what is false of the disorders that science will discover quite independently of our interests / values'. That is mostly what I want to say, I think.

Ian Hacking has interesting stuff to say about this with respect to 'abuse' and the phenomena that gets to count as 'abuse'. One of the things that he maintains (controversially, but I find somewhat plausibly) is that whether or not an individual is harmed by phenomena x is at least partly determined by whether society considers x to be abuse. In other words - if we classify a phenomenon as 'abuse' then a consequence of this is that someone was 'victimised' or harmed. This is the kind of thing I want to address in the last chapter. I'm particularly interested in the situation of such disorders as schizophrenia (for instance) being defined as chronic such that if a person manages to function normally then we conclude misdiagnosis rather than concluding that people with schizophrenia can recover after all. I then worry (one hell of a lot) about the consequences of our telling a person that they have a diagnosis of schizophrenia (in particular whether the act of classifying

them makes it more likely that they won't recover). There is a literature on this with respect to intelligence (with respect to teachers treating students differently and also with respect to self conception). Self fulfilling prophecy.

1.3 What are the major issues here?

What is the distinction between the 'worried well' or those seeking 'self improvement' and those seeking help for a disease or a disorder? (We care about this because of who has the right (defeasibly) to free / subsidized treatment. What sorts of treatments should be free / subsidized (e.g., medication? talk therapy? behavioural intervention? social intervention?) Who should treat people with mental disorders? (Which field? What should the curriculum be for each field?) Should there be treatment parity (e.g., equal reimbursement / assistance for mental and non-mental disorder? Criminal probably needs to come up here too. Maybe something about moral responsibility etc to lead into the next chapter.

1) Attempts to justify psychiatry's status as a specialist branch within medicine where mental disorders are construed as certain kinds of physical (read biological) disorder, 2) Attempts to gain parity with medicine with respect to health insurance reimbursement for treatment 3) Attempts to show that psychiatry has progressed with respect to better (drug) treatments for mental disorders.

There has been a lot of controversy over how we should define the medical notion of disorder such that it picks out the appropriate people and conditions. While the issue first came up with respect to medicine, most of the recent debate has been driven by attempts to either justify or undermine psychiatry's status as a specialist field within medicine. On one side of the debate theorists have maintained that the same notion of disorder is in play in psychiatry and in general medicine. On the other side of the debate theorists have maintained that the psychiatric notion of disorder is importantly different from the medical notion. This issue is typically regarded as being important because which individuals (and conditions) count as being disor-

dered has implications for which individuals have some claim to treatment that is either publicly funded or covered by health insurance. The intuitive idea is that if someone is disordered then they have some claim to treatment whereas if someone is not disordered then while medical intervention might well improve their life they have no claim to it[4].

A related issue is that of what kind of treatment people get. While mental disorders are the subject matter of psychiatry they are also the subject matter of clinical psychology, for example. Psychiatry has had a history of association with psychoanalysis but it has been argued that psychoanalysis is only suitable for people who aren't psychiatrically disordered, however[5]. In practice psychiatrists are distinguished from psychologists in that psychiatrists have medical training and prescribe medication whereas clinical psychologists have a psychology background and provide therapy. Most of the anti-psychiatry critique has been to counter the notion that people who are mentally disordered should receive medical treatment. Some critics further maintain that psychiatry should not have the authority to confine and medicate people against their will and / or that the insanity defence should be abolished. While there is controversy over the precise commitments of the medical model the general notion is that there is some kind of dysfunction or disorder with the individual. Most of the anti-psychiatry critique of psy-

[4]This distinction is meant to capture a difference between 'problems in living and 'disorder. The intuitive idea behind the distinction is that there is a difference between some people having a disorder and having some claim to being in a non-disordered state, whereas people who dont have a disorder yet would like to improve their functioning to become super-functional don't have such a claim. One might think this distinction is merely a way of prioritising our finite resources and that if resources were infinite 'problems in living would have an equal claim to treatment. In response, we don't have infinite resources.

[5]There aren't outcome studies on psychoanalysis to show that it is an effective intervention for people with psychiatric disorders. Psychoanalysis has morphed into briefer forms of psychodynamic therapy which has been tested against other varieties of therapy such as CBT. In practice psychiatrists tend to provide medication, however, while therapy is more often provided by clinical psychologists. While there is some cross-over with psychologists being granted limited prescribing rights in some states and psychiatrists providing some therapy in in-patient and out-patient settings a medical background wouldn't seem required for therapy and yet would seem required for differential diagnosis, prescribing, and effective monitoring of side-effects of medication such as diabetes, heart attack etc

chiatry has focused on the consequences of regarding mental disorders to be certain kinds of physical disorders so that the medical model can be equally applied to them. Anti-psychiatrists have sometimes even maintained that there aren't any such things as mental disorders. This claim is best understood as being a metaphysical thesis rather than a sceptical thesis, however. Part of the anti-psychiatry critique is that psychiatry has more to do with treating people who make us feel uncomfortable because they violate certain kinds of social and / or moral norms rather than treating people because we have good grounds for thinking that they have dysfunctional biology. As such, psychiatric disorders are thought to be problems with society rather than problems with individuals who are diagnosed as having them.

The debates are complex and there seem to be a number of distinct issues in the vicinity. Some of those issues seem to be conceptual (how should we define disorder) and some of those issues seem to be empirical (what is the most effective treatment for mental disorder?) In order to bring some clarity to this complex of issues theorists have attempted to carve off the supposedly non-normative issue of the nature of mental disorder from the supposedly normative issue of who we should and should not treat and how we are best to treat them (e.g., social, psychological, or medical intervention). The thought is that the nature of mental disorder can be grounded firmly in the sciences such that psychiatry's status as a branch of medicine is secured. An assumption of this distinction is that there are facts about who is and who is not disordered that are quite distinct from whether the person is entitled to treatment. A problem with this strategy is that it runs the risk of divorcing itself from why the debate matters, however. While theorists often motivate their concern with the nature of mental disorder by saying that it makes a difference for treatment they then proceed to set the latter issue aside. In doing so we may well wonder what the concept of mental disorder that they defend has to do with the concerns that motivated us to care about the concept of disorder in the first place.

One thing that I do need to be clear on at the outset is that mental disorder is a notion that comes up in a variety of contexts. There is the legal notion

of insanity where a diagnosis of a mental disorder is necessary (though not sufficient) for an insanity defence. There is the folk notion of crazy where certain kinds of (yet to be specified) social and / or moral norm violations would seem to play a necessary (or possibly necessary and sufficient) role. In what follows I shall only be interested in the psychiatric notion of mental disorder, however. I don't see any reason why we should expect the psychiatric notion to coincide perfectly with these other notions, and further work would be required in order to explicate the relationship between the psychiatric notion, the legal notion, and the folk notion.

Fulford maintains that the main reason for the prevalence of the values out view is that there have been many advances made in the natural sciences whereas there seem to have been comparably few advances made in ethics. He states that the values out line is the view that psychiatry can in some interesting sense be shown to be reducible to medicine and that medicine in some interesting sense can be shown to be reducible to biology and that biology can in some interesting sense be described as purely causal facts. If psychiatry can be shown to be thus reducible then the scientific foundation for psychiatry is assured. This brings us to the grounding project

Chapter 2

The two-stage view: Scientific discovery vs normative theory

Chapter Introduction

In this chapter I will focus on the division of labour between scientists and non-scientists (including normative theorists and conceptual analysts). The form that this debate has taken has been one over one versus two stage views of mental disorder. In this chapter I'll thus begin by introducing the fairly standard motivation for a two stage view. I'll then consider alternatives to it - views that maintain that only one of the stages is sufficient for mental disorder. In the next chapter (or possibly at the end of this one) I'll introduce the most popular and systematically defended view on offer: That provided and defended at length by Jerome Wakefield.

2.1 Paradigmatic cases and the two-stage view

For as long as we have been interested in helping those with what are now considered paradigmatic mental disorders we have been interested in helping those with what are now considered paradigmatic non-mental disorders. Indeed, the distinction between mental and non-mental disorders is a rela-

tively recent one that arose around the time that other medical specialities developed in the 19th century (Shorter, 1997, p.1). Before medicine became systematized as an institution people sought help from a variety of sources and treatments varied from social support, to ingestion of substances, to prayer.

Imagine a tribe of hunter-gatherers where each individual contributes towards the hunting or gathering of food on a daily basis. Consider the following four cases:

- 1. During a hunt one of the members of a tribe is accidentally stabbed in the leg by a spear. When the spear is pulled out the skin is open and there is blood. When the person attempts to walk they scream and cease in their attempts.
- 2. One of the members wakes up in the morning and when the person attempts to walk they scream and cease in their attempts. The leg doesn't look any different to what it looked like yesterday when the person participated successfully in the collection of food.
- 3. As above except in this case the persons prior participation in the collection of food was unsuccessful and the person has previously expressed reluctance to participate.
- 4. As above except in this case the person refuses to attempt to walk.

I think most will find it plausible that the first case is the clearest case of the presence of bio-medical disorder whereas the fourth case is the clearest case of the absence of bio-medical disorder. That being said, a person might maintain that each of these cases is a case of disorder, or that none of them are. It would appear that these theorists would need to make a case for their theory, however. Theoretical considerations might motivate us to revise our intuitions (or to dismiss them as wrong) but the burden of proof would be on the revisionary theorist.

When it comes to determining which if any of the above cases are cases of bio-medical disorder the following considerations seem to be relevant.

- 1. The presence of physical abnormality / dysfunction / defect. What this is trying to capture is that in the cases where the skin is open and there is visible blood we tend to have the intuition that there is something physically bio-medically wrong with this person.
- 2. The presence of suffering, pain, distress. What this is trying to capture is that the person seems perturbed psychologically and that this is limiting their normal activities[1].
- 3. We have some kind of duty or obligation to assist them at cost to ourselves / society since they would be better off without their condition.

And thus we arrive at a fairly generic version of the two-stage view. According to the two-stage view there is firstly an objective aspect to disorder (provided by the first condition) and secondly a normative aspect to disorder (the second and third condition). There are a number of different particular versions of the two-stage view as different theorists attempt to cash out the features of each stage in slightly different ways.

2.1.1 Motivating the two-stage view

The main controversy that has been inspired by the two-stage view is controversy over the supposed norm independence of the malfunction condition. In particular the controversy is over whether the malfunction condition is sufficient to ground psychiatry as a branch within medicine or whether the malfunction condition is implicitly thick in the sense of being implicitly normative. Fulford characterises a naturalisation project that theorists have where they maintain that psychiatry can be grounded in medicine and that medicine can be grounded in biology. Fulford (2000, p.78) states:

> The philosophical project of naturalization in biology, medicine, and psychiatry has been concerned mainly with five key terms:

[1]This aspect might be phenomenological. There is a bit of a continental literature on this and on the experience of mental illness that I should probably mention.

> function, dysfunction, disease, illness, and disorder. The meaning of these terms, moreover, most authors recognize, are linked. The details vary, but broadly speaking they are taken to form a logical cascade. In this naturalization cascade, as I will call it, disorder includes disease and illness, illness (the experience of illness) is defined by reference to disease, disease by reference to dysfunction, and dysfunction by reference to function. (I give some examples in a moment). The importance, therefore, of biological function statements to the naturalization project is that they appear to provide a value-free scientific foundation on the basis of which the other terms in the naturalization cascade can be built up. Most authors recognize that values must come in at some point in the cascade: if not with dysfunction, then with disease; if not with disease, then with illness; if not with illness, then with disorder. But provided biological function statements are value free, the naturalization project, it is widely assumed, can at least get underway.

He continues:

> Medicine has been successful precisely through its identification with science. No one wants to be a loser, therefore. Everyone wants to join the winning team. Psychiatrists (such as Kendell 1975) want to naturalize mental illness in terms of disease so that they can join the medical team of medical science; medics (such as Campbell, et al. 1979) want to naturalize disease in terms of dysfunction so that they can join the winning team of biological science; biologists (Allen, et al. 1998) want to naturalize dysfunction in terms of function so that they can join the winning team of natural science; natural scientists, on this model, are the winning team (Fulford, 2000, p.79)

In order to understand the motivation for a two-stage view it is worth our considering each stage individually. First, we shall look at the scientific. Secondly, we shall look at the normative. Or the other way around, I can't

figure this one out. The main issue here is understanding why people think that a view that is entirely scientific or entirely normative is insufficient.

Murphy maintains that the Orthodox Program of Conceptual Analysis is committed to both a two-stage view of mental disorder and a particular view of the role of conceptual analysis. Ill begin by offering an account of the two-stage view and then turn to the more general issue of the view of conceptual analysis in the next section.

According to the two-stage view there are two individually necessary and jointly sufficient conditions for mental disorder. Firstly, there is malfunction, and secondly, the malfunction has harmful consequences for the individual and / or society. The clinicians handbook The Diagnostic and Statistical Manual of Mental Disorders endorses the necessity of the first condition when it states that whatever its original cause it must currently be considered the manifestation of a behavioural, psychological, or biological dysfunction within the individual. While Wakefield differs from the DSM by maintaining that inner rather than purely behavioural malfunction is required it is clear that Wakefield and the DSM are similar in regarding malfunction to be necessary for mental disorder. The two-stage view has been extremely influential, partly because it promises, by way of the objectivity of the first condition, to ground psychiatry firmly on a scientific footing. The notion is that scientists can investigate malfunction independently from our normative assessment of harm. It is partly because malfunction is regarded as objective that the two-stage view has been embraced by the majority of psychiatrists.

2.1.2 Disordered individual's vs kinds of disorders

It might be worth distinguishing two different issues at this stage. Firstly, there is the issue of the boundary between disorder and non-disorder. This might amount to the question 'do medical disorders form a natural kind' (or something approximating it. We could also ask the question 'do psychiatric disorders form a natural kind' (or something approximating it. We could also ask the question 'does a particular kind of psychiatric disorder form a natural

kind' (or something approximating it. It might be that the answer to one of these is different from the answer to another of these. It might also be that while there is a normative aspect to one of these (e.g., whether an individual has a disorder or not) there is a completely non-normative story to be given about particular kinds of disorders. There seems to be this intuition that the difference between schizophrenia and bi-polar (if there is one) is solely a matter for science and our normative considerations don't play a role at all. Whereas when it comes to the issue of whether schizophrenia or bi-polar are neurological compared to psychiatric might be an issue that is partly decided by pragmatics, and whether a particular individual with schizophrenia or bi-polar is disordered or not, or whether schizophrenia or bi-polar really are disorders or not might well have a normative aspect. I will have much more to say about this when I come to the section on natural kinds. Especially with respect to differentiating types and tokens.

2.2 Alternatives to the two-stage view

2.2.1 Scientism

In the 1960's psychiatry came under pressure from without and within to justify the decision to include some conditions as disorders while excluding others. From the outside political campaigning of gay rights activists to get homosexuality removed from the DSM put pressure on what considerations were relevant for inclusion or exclusion of conditions as disorders. From outside psychiatrists had many other examples of what they were tempted to regard as 'abuses' of psychiatry. In Russia political dissenters had been regarded as having 'sluggish schizophrenia' solely in virtue of their political dissent, and that was thought to justify their being involuntarily institutionalized and treated. At least one psychiatrist in the USA had argued that slaves in the American south who desired to escape from their owners were suffering from the disorder of 'draeptomania'. While draeptomania was never included in the DSM psychiatry came under increasing pressure to justify why some conditions were included while others were excluded. In

particular, the concern was that there was little more to our regarding an individual (correctly) to be disordered than our judging the individual to be violating certain kinds of (yet to be specified) social and / or moral norms. In the face of this concern the APA attempted to define mental disorder in the DSM III and in particular, to say something about how biological or scientific facts more generally played a role in our rightly regarding an individual to be mentally disordered.

Scientism arose as a defence of psychiatry's place as a specialist field within medicine. According to scientism there are objective facts about who is and who is not mentally disordered that are biological and for the natural sciences to discover. On this view the relevant facts for determining disorder are scientific. Hence we have geneticists, neuroscientists etc attempting to discover those relevant facts so we are better able to diagnose.

While theorists might explicitly reject one stage objectivist views as inadequate this view does seem to be something that people implicitly hold. Newspaper headlines often proclaim such things as 'science has discovered the biological basis of schizophrenia' and while such claims are surely premature the thought behind this seems to be that there is such a biological basis to be found. This seems to tap into the intuition that a lot of people have that whether a person has a disorder or not is something that is discoverable by science. Diagnostic tests. Doctors as scientists reading off the data. Go to the doctor to see what (if anything) might be wrong with you.

It seems fairly intuitive to many people that there are objective facts about bio-medical disorders that are discoverable by the natural sciences. Hence we have a handle on why it is that geneticists, neuroscientists etc are studying the phenomenon and why it is that we think we are making scientific progress in understanding the phenomenon.

Before I do this I think that it is worth saying that while theorists might explicitly reject one stage objectivist views as inadequate this view does seem to be something that people implicitly hold. Newspaper headlines often proclaim such things as 'science has discovered the biological basis of

schizophrenia' and while such claims are surely premature the thought behind this seems to be that there is such a biological basis to be found. This seems to tap into the intuition that a lot of people have that whether a person has a disorder or not is something that is discoverable by science. There also seems to be an intuition that whether an individual has a particular kind of disorder or not is something that there are objective facts about. Particular psychiatrists or psychologists could get the diagnosis wrong. They could also be mistaken in regarding an individual to have a disorder when they don't or they could be wrong in regarding an individual not to have a disorder when they actually do.

Strong scientism

The standard view of bio-medical and mental disorder is the medical model. According to the medical model mental or psychiatric disorders are a subset of non-mental or bio-medical disorders more generally in the same way to how orthopaedic, paediatric, and cardiology disorders are. Indeed, psychiatry is a specialist field within medicine just as how orthopaedics, paediatrics, and cardiology are. This is a fairly weak version of the medical model - one that is typically also endorsed by allied health professionals such as counsellors, social workers, and psychologists. The idea seems to be that there are scientific facts about bio-medical disorders that are discoverable by scientists and that mental disorders are just like bio-medical disorders more generally in relevant respects (including those to do with prioritizing it as a leading health issue).

A stronger version of the medical model according to which the above facts entail that psychiatry should be authoritative when it comes to diagnosing, classifying, and treating mental disorders is controversial and tends to be rejected by professionals outside medicine and psychiatry, however.

Weak scientism

The weaker version. Only the anti-psychiatrists seem to deny it. They hold a strong scientism about bio-medical disorder, however.

2.2.2 Normativism or evaluationism

Normativism or evaluationism is basically a denial of the scientific aspect outlined above. It is a denial that even the weak medical model is true of mental disorder. On this view mental disorders are radically different from bio-medical disorders insofar as the scientific aspect holds true of bio-medical disorders but not mental disorder.

It clearly won't do to leave it at that, however, as criminal conduct is also a violation of social and moral norms and yet criminals don't thereby come to be regarded as mentally disordered. It is also common to distinguish between socially and / or morally deviant character traits like anti-sociality, shyness, laziness etc and mental disorders. We seem to have intuitions that these are separate kinds of social and / or moral norm violation and thus more must be said about what kinds of social and moral norm violations are relevant for mental disorder. There is also controversy as to whether symptoms like delusions and hallucinations are disorders even when they are endorsed by social and moral norms. While we might decide not to treat delusions and hallucinations if the person is unwilling and society is not condemning this doesn't show us that the person doesn't have a cognitive and / or neurological disorder / dysfunction. It might be that there is something objectively wrong with the person who has such symptoms regardless of whether they are thought to violate certain kinds of social and moral norms. While norm violation might come into issues to do with involuntary treatment it might well be a separate notion from that of disorder.

Normativism / evaluationism can take one of two broadly different forms. It is to these that we shall now turn.

Non-eliminativism

Non-eliminativist versions hold that there are such things as mental or psychiatric disorders but the nature of them turns out to be radically different from what we had supposed. The idea here is that mental or psychiatric disorders are different from bio-medical disorders insofar as the medical model (either

the strong or weak version) is basically accepted for bio-medical disorders but not for mental or psychiatric disorders. On this view mental disorders are not a subset of bio-medical disorders after all. They turn out to be different. 'Mental' is not a simple modifier the way that 'immunological' or 'cardiovascular' or 'paediatric' are.

While this view has traditionally been associated with the view that psychiatry (as a field of medicine) should be abolished it is important to note that this need not be the case. If mental disorders are radically different from the subject matter of the rest of medicine this might be grounds for saying that psychiatry should not be a field of medicine - perhaps psychiatrists should alternatively be found in psychology or sociology departments? Alternatively psychology or counselling. Just trying to separate the notion of scientism from the notion of treatment.

Eliminativism

We can distinguish roughly two different kinds of eliminativism. The first is eliminativism about mental disorders. On this view there aren't any such things as mental disorders. In order to understand where this view is coming from it is worth considering other things that science has investigated before concluding that there isn't any such thing. Firstly we have witches, secondly we have phlogiston. We also have the discovery of black swans (despite all swans are white previously being cited as a true universal generalization). The thought is that witches were thought to have a bunch of properties that enabled us to more or less identify them and thus to learn about them. The thought was that while there were indeed women and some of them had cats and beards and lived a solitary lifestyle. And indeed some of them floated when we attempted to drown them. And yet these individuals turned out not to have special powers. And thus we conclude that since having special powers (and using them for malevolence) is incredibly central to the concept of witch-hood that the appropriate thing to conclude here is that there aren't any witches. And thus since there aren't any witches we should probably stop drowning innocent (insofar as they aren't using special powers

for malevolence) women. Whether or not they have black cats. And live by themselves. And float.

Or consider phlogiston. The wonderful magical heat fluid that flows from one substance to another that explains how... Heat gets transferred from one substance to another. We could have concluded that phlogiston was a wonderful magical heat fluid that didn't weigh anything (since cooler items don't weigh more). Though of course ashes weigh less... Anyway... The point is that we concluded that there wasn't any such thing as phlogiston (the fluid stuff) instead there was just the regular kind of stuff moving. And of course there aren't any beliefs or desires according to neuroscience. And there aren't any tables and chairs according to subatomic physics. Weak eliminativism is theoretical. There aren't any such things as mental disorders. It isn't particularly committed to anything at all with respect to how we should or should not proceed with treatment, however. Strong eliminativism maintains that since scientism turns out to be false psychiatry as an institution is illegitimate.

The dismissal of normativism

Normativism isn't a very popular view - but it does represent the main critique for scientism and also for two-stage views (that we will get to) that maintain there is a role for both scientism and normativism. The main arguments against normativism are that it fails to capture our intuitions with respect to certain cases and the implicit assumption is that our intuitions are correct and must be respected.

For instance, normativism (the argument goes) seems committed to the view that when- ever people are violating norms they are disordered. So, for instance, when psychiatry judged individuals to be suffering from sluggish schizophrenia due to their political dissent those psychiatrists were correct. Not all normative violations are (intuitively) mental disorders. E.g., oddness, laziness, criminality etc.

Secondly, normative violation seems to vary cross-culturally. What is and

isn't a violation varies. But whether or not someone has cancer or HIV doesn't vary cross-culturally[2]. At this point the burden of proof seems to lie with the theorist who maintains that normative violation is necessary and sufficient for disorder. What kind of normative violation? Are health and/or psychiatric norms ones that are familiar to us - just put to special use here? Or are these a distinctly different kind of norm than those that are more familiar to normative theorists?

The two-stage theorist needs an account of the relevant kind of normative violation too, of course. The issue seems to be more significantly pressing for the theorist who maintains that normative violation is not only necessary but also sufficient for mental disorder. In the two- stage view literature the normative violation aspect often seems to serve as a placeholder. The debate has focused on offering an account of the scientific or non-normative aspect. Skeptics maintain that the non-normative aspect turns out to be normative after all while two-stage theorists defend a non-normative aspect. There is much work that remains to be done on the normative aspect for the one stage normative theorist and for the two stage theorist alike.

There are two different ways we can go here. Firstly, we can consider that instead of prevailing social norms setting the relevant standard, there are some idealized norms and when individuals violate those norms that is what is relevant for our being appropriately justified in regarding an individual to be disordered. Secondly, we can maintain that not just any kind of norm violation is relevant, rather than norm violations must be of a certain kind. Both of these moves is to make a distinction between our actual judgement that someone is violating the norms relevant for fixing disorder and facts that are at least potentially independent of the judger that fix whether the judger is correct in their judgement.

The problem of relativism is something that comes up fairly standardly in ethical theory. Ethical theorists are typically concerned with theories or

[2]Of course prevalence rates can vary cross culturally. One can't simply cure a disease by moving countries, however.

norms that transcend the norms currently embraced by any particular society or culture. They thus attempt to promote tolerance of difference while still allowing us to critique certain actual and possible moral or social practices as unwarranted. We typically do want to allow that some societies and / or cultures have social and / or moral norms that are unwarranted or illegitimate or criticizable in some way. Utilitarianism, for example, allows that a particular act that maximizes utility in one situation or culture may be quite different from the particular act that maximizes utility in another situation or culture. It seems that a similar move is available to the normative view of mental disorder. Instead of maintaining that mental disorders are in violation of particular current norms they are able to make a comparable view in maintaining that they are in violation of these norms which would be held by a sufficiently enlightened or otherwise idealized society. Whether there will be one unique view remains to be seen. Whether there will be cross cultural variations also remains to be seen (e.g., utilitarianism).

The dismissal of scientism

Statistical abnormality isn't sufficient because some are positively valued whereas others are negatively valued and so whether we value something or not seems to be relevant. This has been taken to be motivation for values but the statistical notion has come to be dismissed anyway.

Perhaps if we get the scientific aspect right there won't be the need for a normative / evaluative aspect. While it is often thought that the more the merrier (that inclusive is good and that more ideas are better) it is important that we not multiply components beyond what is strictly necessary. Grounds for dismissal of Freud (though Plato provided a good defence for a tripartite division of the soul / mind).

2.3 Theoretical motivation for the two-stage view

The standard view of psychiatric or mental disorder is that it is a subset or particular kind of bio-medical disorder[3] The thought is that the term 'disorder' is shared between psychiatry and other fields of medicine such as neurology, cardiology, and paediatrics. The distinction between psychiatry and other branches of medicine such as neurology is meant to be a function of the 'mental' modifier[4]. The main current approach to bio-medical disorder in general and psychiatric disorder in particular is a two-stage view. According to the two-stage view there are two individually necessary and jointly sufficient conditions for bio-medical disorder[5]. Firstly, there is a dysfunction, often thought to be within the individual, that is to be discovered by science. Secondly that dysfunction is thought to result in harm to the individual and / or to society where harm is a normative notion. Firstly, there is a dysfunction, often thought to be within the individual, that is to

[3]It might seem that there is dissent here from the American Psychological Association in that it threatened to sue the American Psychiatric Association if they stated that mental disorders were 'biological' in the *Diagnostic and Statistical Manual of Mental Disorders* IV. It might help to distinguish between the view that the most effective interventions for mental disorders are medication and / or surgical interventions provided by doctors compared to the view that mental disorders are a certain kind of bio-medical disorder, but where psychological interventions (e.g., psychotherapy) may be as effective or the most effective for at least some conditions. Clinical psychologists can thus agree that mental disorders are biological at base so long as the biological base is expanded to include psychological or mental disorders. Further consideration of this issue would take us too far afield. It is also worth stating that while some theorists have concerned themselves with distinguishing between 'disorder' (in the bio-medical sense) and 'disability', 'malady', 'treatable condition', and 'disease', we will follow recent convention in using 'disorder' as a stand-in for all of these related phenomena.

[4]The distinction between psychiatry and neurology is about as problematic as the distinction between the mind and the brain. While none of the accounts of the distinction on offer in the philosophy of mind or the philosophy of cognitive science seem to be able to capture what is going on with our division of psychiatric and neurological disorder this is not an issue that we shall take up here.

[5]While the DSM doesn't put the matter so succinctly the list of disjunctive conditions provided by the DSM can be analyzed into these two distinct components though the DSM states explicitly that it is not attempting to offer a necessary and sufficient condition analysis.

be discovered by science. Secondly that dysfunction is thought to result in harm to the individual and / or to society where harm is a normative notion.

The main virtue of the two-stage view is that it seems able to respect two different sets of intuitions that we have about the role of science, and about the role of normative, evaluative, or possibly even ethical theory. On the one hand we have the intuition that science has a role to play in the discovery of objective facts about disorder[6]. On the other hand we have the intuition that norms have a role to play in whether an individual is harmed by their dysfunction or whether their dysfunction is neutrally or positively valued by themself and / or society. This harm is thought to have normative implications along the lines of rights and duties to treatment. The two-stage view is thus meant to provide a middle ground between a scientism which maintains that there is no role for values or normativity and theorists who maintain that there is no role for science as our judgement that a person is disordered is no more than our judgement that they are violating some kind of (yet to be specified) social and / or moral norms[7].

While unpacking the notion of harm is at least as problematic as unpacking the notion of function and dysfunction the majority of the debate over the two-stage view has focused on either attempts to offer an analysis of the

[6]My thought was that we can come back with their having a role to play with respect to finding out the causal processes that lead to the production / in discovering the most effective interventions. Whether facts about dysfunction is something that they do really is unclear to me.

[7]This distinction is frought. Wakefield runs together objective / to be discovered by science / mind- independent / culturally invariant and subjective / normative / culturally specific / mind-dependent. I don't really know why he does this. I think that one needs to be careful because some theorists do believe that there are objective facts about norms and that there is something along the lines of a 'final science' equivalent. I don't know whether the final normative analysis will tell us which behavioural symptoms clusters are appropriately regarded as psychiatric deviations rather than breeches in some other kind of normativity or not. It does seem clear that ethical theory provides us resources to critique past bad psychiatric practices (i.e., for their having bad norms. I personally don't want to go this way but I'm surprised that nobody has tried. Similarly I think that the evolutionary notion of dysfunction might be able to be expanded such that it allows for cultural variation and basically encompasses harm (e.g., the phenotype seems to need to be harmful to the organism on average in order for the phenotype to be regarded as dysfunctional. That clearly needs a lot more thought.)

dysfunction condition, or attempts to enumerate problems with the analyses of dysfunction that have been offered thus far. The problem of offering a naturalistic or scientifically respectable account of biological function and dysfunction has long been a concern for philosophers and for philosophically inclined biologists. In the 1960's a number of philosophers attempted to naturalize talk of function and dysfunction in biology by appealing to evolution by natural selection as the naturalistic process that fixes them. The thought is that if talk of 'function' and 'dysfunction' in biology can be successfully translated into talk of evolutionary functions and dysfunctions then biologists use of the terms are unproblematic from a scientific point of view. Jerome Wakefield has appealed to work done by these theorists in his arguments that the evolutionary notion of function and dysfunction is the relevant notion for psychiatry and general medicine.

According to the two-stage view the one stage theorists both get something right in the sense that both science and normative factors play a role in determining disorder. Dysfunction isn't thought to be enough, however, as intuitively a person can be dysfunctioning in a biological sense and yet due to contingencies in their environment that doesn't result in harm. Conversely, a person may be harmed and yet we have the intuition that not every problem in living is a bio-medical disorder. In order to respond to this objection one would need to consider either biological dysfunction or harm in more detail in order to show that either there isn't really a bio-dysfunction or harm after all or tell some story that would persuade us to revise our intuitions about this.

If we rightly understand both biological dysfunction (in the sense that is relevant for psychiatry) and normative violation (in the sense that is relevant for psychiatry) then it might turn out that they are co-extensive. It surely seems that they aren't to be identified - but without more of an account it really is hard to know. It is important to note that what is distinctive about science doesn't seem to be a function of the subject matter (we can do a science of norms with respect to what norms people actually adopt or with respect to how useful certain norms are in fact with respect to some goal

that people agree is worth obtaining) though naturalization of the categorical imperative or what people should adopt (simplicitor or regardless of their other goals etc) is frought.

While there is much work to be done on rightly understanding the notion of normativity that is in play (is it social norm violation? moral? something else?) this is not an issue that I shall take up here. What I wish to propose instead is to focus on the notion of biological dysfunction and in particular focus on it with respect to how useful or accurately it seems to cast the role of the biological sciences in helping us understand disorder. There are many different biological notions of dysfunction (depending on how one individuates concepts) and it is worth getting at the role (and limits of science) with respect to the dysfunction criterion in particular.

It is a little worrisome how much our intuitions really do play a significant role. What else are we supposed to do, however? Variation in different peoples intuitions and what seems obvious to one doesn't seem obvious to another. While we have seen the two one stage views on offer and considered ways in which one might try and develop them in order to respond to the objection. The most popular version of the two-stage view maintains that biological dysfunction is necessary but not sufficient for disorder. In addition to biological dysfunction the products of that must result in harm to the individual and / or to society. This is supposed to capture a middle way between our intuitions that science plays a role and that norms play a role. So this is the account of the division of labour. Scientists discover biological dysfunction and normative theorists consider the harm.

Scientists discover a dysfunction then normative theory is decisive. Norma tivists describe a normative violation and scientists are decisive with respect to dysfunction. For the necessity of the evaluationist criterion the thought is that it is not enough to have a dysfunction in the factual sense, that dysfunction must result in harm. stage views of mental disorder attempt to carve a middle way between both of the one stage approaches. The thought here is that while there are objective biological facts that are relevant for determining who is and who is not disordered there is a normative aspect

as well. One strategy is to attempt to show that the normative aspect of psychiatry is one that is shared with general medicine so insofar as mental disorder is partly normative this won't undermine psychiatry's status as a medical field so long as medicine shares a similar normative aspect. The concern here seems to be to show that mental disorders are like non-mental disorders rather than psychiatry being more like law than like medicine. Most of the critique has come from anti-psychiatrists so the concern is to justify psychiatry's status as a branch of medicine and to justify the medical (biological) model of mental disorder.

There is also a concern to show that medicine is grounded in the natural sciences, however. While it seems less controversial that there is a biological aspect to non-mental disorders. Two stage views of mental disorder attempt to carve a middle way between both of the one stage approaches. The thought here is that while there are objective biological facts that are relevant for determining who is and who is not disordered there is a normative aspect as well. One strategy is to attempt to show that the normative aspect of psychiatry is one that is shared with general medicine so insofar as mental disorder is partly normative this won't undermine psychiatry's status as a medical field so long as medicine shares a similar normative aspect. The concern here seems to be to show that mental disorders are like non-mental disorders rather than psychiatry being more like law than like medicine. Most of the critique has come from anti-psychiatrists so the concern is to justify psychiatry's status as a branch of medicine and to justify the medical (biological) model of mental disorder.

Theorists who are inclined to the medical model accept that mental disorders are certain kinds of bio-medical disorders. Theorists who offer a one-stage normative account of mental disorder typically have a different view of the nature of bio-medical disorder, however. The thought is that in seeing whether mental disorders are kinds of biomedical disorders or not one needs to have a view on both the nature of mental disorder and the nature of bio-medical disorder more generally. With both those views in place on is then in a position to see whether mental disorders form a subset of biomedical disorders or

not. One thing that can (and has) gone wrong with the debates is that theorists who maintain that mental disorders are not biomedical disorders have a different view on what it takes to be a biomedical disorder than theorists who maintain that mental disorders are biomedical disorders.

No	Bio-medical	mental	are they the same?
1	One-stage normative	One-stage normative	yes
2	One-stage normative	One-stage objective	no
3	One-stage normative	Two-stage	no
4	One-stage objective	One-stage normative	no
5	One-stage objective	One-stage objective	yes
6	One-stage objective	Two-stage	no

The above table shows the variety of positions that one could take on both the nature of biomedical disorder and the the nature of mental disorder. The 'BioMedical Disorder' column provides three different positions that one could have on the nature of Biomedical disorder. One could hold a one-stage normative view of them (though I don't know that anyone has done so). One could hold a one-stage objective view of them like Boorse and the majority of theorists who maintain that mental disorders are not kinds of biomedical disorders. One could adopt a two-stage view of them - as the majority of theorists who accept the medical model do.

The second column provides a list of the different positions that one could adopt on the nature of mental disorder. Once again we have a one-stage normative view, a one-stage objective view, and a two-stage view. The two most common views out of the nine different combinations is the fourth view - that held by the majority of theorists who maintain that mental disorders are not a kind of biomedical disorder. The most commonly held view for theorists who maintain that mental disorders are a kind of biomedical disorder is position nine - though position five is also possible (and Boorse might be

best characterized here).

What is interesting about this table is the number of positions on it that are unoccupied. Nobody attempts to argue that psychiatry is more objective than biomedicine more generally - and I don't know of anyone who has a completely normative account of bio-medical disorder more generally.

Non- Normativists maintain that there are non-normative facts that make it the case that an individual is disordered. Theorists differ as to whether the relevant facts are facts about the inner nature of the individual, the individuals behaviour, the individuals behaviour in relation to the persons society, or some combination of the above. Different theorists have thus located the facts at different levels of analysis. Relevant levels could include genetic, neurological, cognitive psychological, behavioural, sociological, or some combination of the above. Typically the relevant facts are thought to be facts about malfunction. The notion of malfunction that is in play here is regarded as objective or non-normative in the sense that malfunction is thought to be fixed by facts about proper function that are determined by breakdown of the effects that are responsible for the presence of the mechanism in current populations. Normativists maintain that the relevant facts are essentially normative in the sense of being determined by our values. The relevant facts are taken to facts about the individuals behaviour being abnormal, aberrant, disvalued, or harmful to the individual or to someone else, where facts about these are thought to be irreducibly value laden. Normativists typically refer to examples in the history of psychiatry to support their normativism. Examples include political dissenters in Russia who were regarded as having sluggish schizophrenia and were involuntarily institutionalised and medicated. Another example is homosexuality that was only taken out of nosology in the 70's, and drapetomania that was a suggested diagnostic category for slaves who attempted to escape their owners. Normativists maintain that examples such as these show us that whether or not someone is regarded as mentally disordered is determined by whether their behaviour is taken to contravene the values of society. Normativists also seem to appeal to facts about what a society does value rather than facts about value

in a more general sense where societies values may or may not reflect these universal normative facts.

I now want to distinguish between simple views that maintain that either non-normative or normative facts but not both determine whether an individual is mentally ill, and complex views that maintain that both kinds of facts are relevant. Simple non-normativism is typically acknowledged to be inadequate as we can imagine cases where an individual has an objective dysfunction and yet where that dysfunction is beneficial to that person and / or to society. One could have an inner malfunction, for example, that resulted in improved performance on some cognitive task and in this case we would not consider the person to be mentally disordered. One could attempt to defend simple objectivism on the grounds that this could not arise from the relevant kind of dysfunction. In order to do this one would need to specify in more detail what the relevant kind of dysfunction is so that one could rule out the above possibility without appealing to normative facts. Simple normativism is also typically acknowledged to be inadequate for a couple of reasons. Firstly, if one maintains that it is the norms that are endorsed by current society that are relevant then we would be unable to criticise political dissenters being classified as mentally disordered so long as their behaviour was not in accordance with the norms of their society. Secondly, even if one takes the relevant normative facts to be universal and hence possibly different from the norms of any society it would still seem possible for a persons behaviour to be harmful, aberrant, or abnormal in a normative sense and yet for the individual to not be mentally disordered. One could attempt to defend simple normativism on the grounds that this could not arise from the relevant normative facts. This would, however, require one to specify in more detail the relevant kinds of normative facts.

The most influential view of mental disorder is probably Jerome Wakefield's Harmful Dysfunction (HD) analysis where he maintains that there are two individually necessary and jointly sufficient conditions for mental disorder. The first condition is an objective notion of malfunction and the second condition is a normative notion of harm. While Wakefield doesn't attempt to

state what makes mental disorders mental or psychiatric as opposed to non-mental or neurological his analysis of the concept of disorder has been very influential. I shall return to unpacking his notion of objective malfunction in a later section but first I want to consider the role that conceptual analysis is supposed to be playing with respect to fixing the facts that determine whether or not an individual is mentally disordered. Conservative and Revisionist While I have been talking about facts that may be relevant to determining that an individual is mentally ill the debate has typically been presented as a debate about our concept of mental disorder or what we must believe about a person in order to justifiably classify them as being mentally ill rather than as being a debate about the kinds of facts that determine whether they are in fact mentally ill. I am less interested in our concept of mental disorder and present nosology and more interested in the nature of mental disorder and the different categories of mental disorder, however.

Conservatives maintain that it is worthwhile to engage in conceptual analysis as conceptual analysis fixes the kinds of facts that are relevant for determining whether an individual is mentally disordered and / or what categories of mental disorder there are. One could be conservative because one believes that our concepts do map on to categories in nature in virtue of being true to the facts about these things. One could also be conservative yet an eliminativist because there simply aren't the relevant kind of facts. Revisionists maintain that we can revise our concepts in the light of scientific discovery. The notion is that our concept of mental disorder could be false to the facts about mental disorder and that in this case the appropriate thing to do would be to revise our concepts. Revisionists maintain that the facts that are relevant for mental disorder are an empirical matter to determined by scientific investigation rather than by a-priori analysis of our concept of disorder or rather than by surveys of peoples intuitions as Wakefield attempts to muster intuitive support for his harmful dysfunction analysis of the concept of mental disorder. A revisionist would maintain that the DSM is attempting to provide a nosological system that maps onto different categories of disorder that are to be found in nature and that it is an empirical matter whether the

current diagnostic system is true to the facts and if it is found to not be true to the facts then the current diagnostic kinds would need to be revised. It seems to me that the revisionist line is more plausible. One might consider that there are prototypical cases of mental illness and that our concept of mental disorder was formed so as to refer to these prototypical cases.

Our concept thus seems have been formed on the assumption that these prototypical cases have something in common. Whether the prototypical cases actually do have anything in common that makes them a category is an empirical matter, however. While conceptual analysis might play a role in determining when we choose to apply and withhold the concept of mental disorder or in whether this person does or does not qualify as meeting present diagnostic criteria for a certain kind of illness it seems plausible to me that our concept of mental disorder and of kinds of mental disorders can and indeed should evolve in light of scientific investigation as to what the cases have in common. Another issue that arises is whether mental illness is best thought of as categorically different from non mentally disordered or whether it is simply a matter of degree. This would also seem to be an empirical matter in the sense that the empirical facts will determine this issue rather than a-priori conceptual analysis, however. While the conservative project involves analysing our concept of mental disorder and perhaps an allowance for making revisions on the basis of considerations such as internal consistency the revisionist project takes paradigmatic cases of people who we regard as mentally disordered and attempts to assess empirically what, if anything, these people have in common. One needs to distinguish between our concepts on the one hand and the categories to be found in nature on the other.

Chapter 3

Wakefield's harmful dysfunction analysis

Chapter introduction

In this chapter I will focus on the two-stage view that has been presented and defended re- peatedly by Jerome Wakefield (e.g., 1992a, 1992b, 1993, 1999, 2000a, 2000b, 2003, 2004). Wakefield's view is of particular interest to us for two main reasons. Firstly, for his explicit intention to take work that has been done in philosophy on the word-concept-world connection and apply it to the issue at hand. Secondly, because it has been particularly influential. He has developed and extended it significantly in response to critiques and there is much to be learned in understanding why he has come to cash it out the way he presently does.

3.1 The Harmful dysfunction account

Wakefield maintains that there are two individually necessary and jointly sufficient conditions for bio-medical disorder.

- **Stage One** There is a failure in the evolutionary function of a mecha-

nism

- **Stage Two** This failure results in harm to the individual and / or society

The first condition is typically regarded as objective in the sense that there are facts about functions and malfunctions that are independent of our normative assessment and are to be discovered by the biological sciences. Malfunction is regarded as insufficient for disorder, however, as it seems possible for someone to have a malfunction that has beneficial consequences for the person and / or society and this would not count as a disorder.

The second condition is typically regarded as normative in the sense that we need to assess whether the consequences of the malfunction are harmful to the individual or society and only the malfunctions that are harmful are disorders.

One advantage of the two-stage view is that the first stage is seen as setting the scientific foundations of the study of disorder such that the scientists can learn about functions and malfunctions and identify them independently of our normative assessment of harm. While harm might be dependent on the values, norms, and activities of particular cultures malfunction is universal. The sciences can thus get on with discovering the objective facts about function and malfunction independently of our assessment of the normative consequences of the malfunction. There are facts about the individual malfunctioning that are necessary for mental disorder and as such mental disorder is not solely a matter of the individual violating norms.

He thus presents a picture where there is a division of labour between scientists on the one hand (where they are characterized as being engaged in the project of discovering dysfunctions) and normative theorists on the other (who consider the notion of harm). Wakefield's account of mental disorder is the same as the above except that the failure of evolutionary function is specified to be a failure in the evolutionary function of a *mental* mechanism. Mental disorders are thus thought to be certain kinds of bio-medical disorders.

While the view is simple in offering a mere two conditions that are thought to be individually necessary and jointly sufficient there is a lot going on in each of those clauses. I will now turn to Wakefield's argument for the failure being in evolutionary function since this clause has generated most of the controversy. Wakefield (2000a, p.39) maintains that conceptual analysis reveals that certain types things that we take to be effects can be explained in terms of a common underlying causal process. He maintains that the relevant concept of function is a shared concept based on prototypical examples of non-accidentally beneficial effects like sight, and on the idea that some common underlying process must be responsible for such remarkable phenomena The notion here seems to be that certain effects like sight or behaviour are adaptive, and that there must be a common explanation for the presence of those effects.

He maintains that the second step to biological function is the modern discovery that the essential process referred to in the conceptual analysis is natural selection (2000 p. 39). While a-priori God could have been the relevant process that fixed the function of effects like sight and behaviour it turned out as a matter of empirical discovery that the relevant process to fix the function is evolution by natural selection. The thought is that natural selection is the process that has resulted in eyes being fairly standard in our species in virtue of eyes enabling us to see. The biological function of the eye is therefore that it enables us to see. When the eye does not enable the person to see then we can say that the eye is malfunctioning. He thus offers an account that is in line with the two-step process of Kripke, Putnam, and others by maintaining that conceptual analysis helps us delimit the phenomena that we are interested in, and that scientific investigation reveals the essential nature of the phenomena. What all mental disorders are thought to have in common on Wakefield's view, is that they are failures of evolved functions that result in harmful behaviour. Wakefield thus maintains that the notion of biomedical function is objective and that mental disorders are failures of evolved functions.

3.1.1 The argument for an HD account of mental disorder

Wakefield's argument for evolutionary dysfunction being necessary for mental disorder may be reconstructed as follows:

- **P1** It follows from our concept of mental disorder that there is a dysfunction to a mental mech- anism (in some pre-theoretic sense of dysfunction) that results in harm to the individual and / or to society.
- **P2** It follows from our pre-theoretic notion of dysfunction that there is an historical process that fixes biological functions and dysfunctions. The nature of that process is to be discovered by science.
- **P3** Scientists have discovered that the relevant historical process for fixing biological functions and dysfunctions is evolution by natural selection.
- **C** Bio-medical disorders are evolutionary dysfunctions of a mental mechanism that result in harm to the individual and / or to society.

3.2 Commentary on the argument

3.2.1 Parsing premiss one

Dysfunction

Wakefield thinks that this premiss is a-priori or that it simply follows from reflection on our concept of disorder. He thinks that it is an intuitive and obvious truth that 'there is something wrong with people who are disordered' in some pre-theoretic or common-sense sense of 'should'. His use of 'dysfunction' here is supposed to be an uncontroversial analysis of 'wrong with'.

Dysfunctioning mechanism

The idea here is that the dysfunction occurs to a mechanism where a mechanism is (very roughly) a component or a part of a person. This is to say that for Wakefield persons cannot be dysfunctioning and neither can their behaviour. Rather, there is some component of them (genetic, cardiac, neurological etc) that is dysfunctioning. Wakefield's notion of a mechanism is fairly relaxed. He maintains that there can be relevant mechanisms on different levels of analysis. In the general medical case he considers mechanisms at the organ level and at the level of the cell. In the psychopathology case he considers mechanisms at the level of neurology and at the level of cognitive psychology. This can be contrasted with the way that Wakefield conceptualizes 'harm' where harm is to persons.

Mental mechanism

Wakefield has both an account of mental and non-mental, bio-medical disorder. The only difference in the accounts is the presence or absence of the 'mental' part of the mechanism. Much work has been done on the notion of a mechanism. I don't want to get lost in this debate here. The idea, however, is that there is a breakdown (a dysfunction) in a component or a part which results in (or causes) the problematic output. Wakefield has defended the idea that the relevant mechanism must be internal to the person. It can't be that behaviour or morphology is dysfunctioning - for Wakefield the phenomenon that is of interest to us (the behaviour or the morphology) is rather caused by inner dysfunction.

The clincian's handbook The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) maintains that dysfunction is a necessary condition for mental disorder but they allow the dysfunction relevant for mental disorder to be behavioural rather than maintaining that it has to be an inner cause of the harmful behaviour. Wakefield attempts to defend the notion that the relevant malfunction needs to be internal to the person by appealing to our intuitions. The example he provides is a case where a person meets the DSM criteria for reading disorder and yet this is not due to inner malfunction,

rather it is due to the fact that he has never been taught how to read. He maintains that intuitively this person is not mentally disordered, whereas the person with the inner malfunction who is likewise prevented from learning how to read does have a mental disorder. Wakefield criticises the DSM for failing to draw a distinction between 'mental disorders' that are caused by inner malfunction and 'problems in living' that are caused by poor person-environment fit. He maintains that after some thought people typically agree that only the behaviours caused by inner malfunction are instances of mental disorder. Along the same vein, Woolfolk and Murphy offer the example of a fire detector that is placed too close to the stove. Because of the placement the fire detector gives off a large number of false positive responses. Wakefield maintains that the fire detector isn't malfunctioning as the problem is one of being in an alien environment.

One of the virtues (appeals) of Wakefields INNER malfunction assumption is that it promises to differentiate between disorder and problems in living. This distinction does indeed seem to be important and it would be nice if we could have an account of it but I have my reservations about Wakefield capturing it because the notion of function is indeed problematic. Murphy is correct to observe that Wakefield does indeed seem to be capturing an intuition that we have that CAUSE of the behavioural symptoms is important. Some kinds of causes (play acting, attempt to get gain, drug induction etc) seem to be exclusion criteria for a person having a certain disorder even if they display the behavioural symptoms. Murphy is also correct to note that we don't need to assume that the relevant cause is a malfunction in order to capture the intuition that a certain kind of cause is important. Precisely what more we say about which causes are relevant and which are not will depend on how things turn out. The notion seems to be that those who are play acting (etc) are importantly different from the other cases. Maybe that they are not the typical cases (if all instances were play acting would we conclude that there is no such thing as mental illness or would we conclude that the nature of mental illness was that it was a play act? Depends whether we take it to be more revisable that mental illness is due to non-intentional causes or

whether we take it to be more revisable that those people are mentally ill).

Harm to the individual and / or to society

The idea is that the dysfunctioning mechanism causes the morphology or the behaviour and that morphology or behaviour goes on to result in harm to the individual and / or to society. The notion of harm that is in play here is thought to be normative. Whether or not an individual is harmed is thought to vary depending on details about their society.

Not terribly much is said about this. The motivation for including it is merely that inner dysfunction in the absence of harm doesn't seem sufficient for disorder. For example, dysfunction in the absence of harm isn't disorder. Mozart's musical genius. Insofar as a dysfunctioning mechanism results in consequences we find favourable or beneficial the person isn't disordered. For instance, I might have a broken bit which enables me to do something that is positively valued. The fact that it is positively valued means I don't have treatment.

It also shows us that the relevant (so it is argued) notion of harm is different from the notion of dysfunction. It also shows us how dysfunctions can be positively or negatively valued - in themselves they are neutral. Wakefield doesn't say a great deal about the notion of harm. It is clear that for Wakefield the notion of harm is a notion that is supposed to cover the normative aspect of mental disorder - but he is more interested in showing psychiatry to be grounded in the natural sciences than in offering an account of that normative aspect. The notion seems to be that behaviour that is harmful in one society may well not be harmful in another.

3.2.2 Parsing premiss two

Caused by an historic process

The idea here is that a-priori it follows from our concept of dysfunction (or indeed of function) that whether or not a thing has a function (or is dyfunctioning) is determined in some way from the history of the thing. This

is something that I will look into more on dysfunction (e.g., on propensity, or forward looking accounts)

3.2.3 Parsing premiss three

This third premiss is meant to be an empirical, or a-posteriori premiss. Wakefield thinks that science has discovered something interesting: It turns out to be the case that in our world the relevant historical process for fixing function and dysfunction is evolution by natural selection. He thinks that a-priori it seemed to us that things could have gone differently (in the same sense as a-priori it seemed to the ancients that the morning and evening star were two different entities). As a matter of empirical fact, this is the way things are, however.

3.3 Wakefield's general strategy

3.3.1 The causal-historical theory of reference

Wakefield draws an explicit analogy between his approach to 'mental disorder' and the causal- historical approach to natural kind terms such as 'gold' and 'water' that was defended by theorists such as Kripke and Putnam in the 60's[1]. Since then a popular view in semantics is that there are two aspects to meaning; what we may (roughly) call a 'primary intension' or an 'A intension' or a 'description' or a 'meaning' on the one hand, and what we may (roughly) call a 'secondary intension' or a 'B intension' or a 'real nature' or a 'referent' on the other.

The primary intension is thought to consist in something along the lines of a description or a list of features that are cognitively significant and that form part of the meaning of the term / the content of a concept. In the case of 'water' / WATER the A intension consists in something along the lines of

[1]Wakefield doesn't discuss some of the more modern controversies within the two-dimensional semantics framework such as the nature of the a-priori, issues of concept individuation etc. I'll gloss over the details here.

the colorless, odorless, potable, drinkable stuff that falls from the skies and fills the lakes etc. In the case of 'gold' / GOLD the A intension consists in something along the lines of the yellowy, shiny, malleable, valuable metal. Now, while it is thought to be contingent that the terms 'water' or 'gold' have the A intension that they have, it is thought that in order to grasp the concept of WATER or GOLD one does need to grasp the A intension. That is what it is to understand the meaning of the terms or to have grasped the relevant concept. As such, it it thought to be a-priori, or a conceptual or analytic truth that the A intension of water or gold is the description that is listed in the A intension.

Kripke and Putnam went on to argue that while this is one aspect to meaning, there is also another aspect to meaning - reference - that served to link the term or the concept on to something that mind-independently exists in the actual world. The B intension is thought to be discovered by science by way of their discovering what the realizers or the A intension are in our world. In the case of 'water' scientists discovered that the colourless, odourless, potable, drinkable stuff that falls from the skies and fills the lakes etc around here was H_2O. In the case of 'gold' scientists discovered that the yellowy, shiny, malleable, valuable metal around here has atomic number 79. The notion then is that certain kinds of terms - natural kind terms - function to track the reference or the B intension. So in Putnam's famous twin-earth scenario if there is a world (not the actual world) in which the watery stuff (the A intension) turned out to be XYZ, then the watery stuff on that world would not be 'water', water, or water. Conversely, if it turned out that if there is a world (not the actual world) in which H_2O is black and tarry then the correct way to describe the world is that their 'water', WATER, or water is black and tarry. This is because 'water', water, and water is necessarily or essentially H_2O *given that* 'water' functions as a natural kind term (which is to say *given that* 'water' tracks the B intension) and that H_2O is the B intension / nature on this world.

3.3.2 Black box essentialism

We are now in the position to see that Wakefield attempts to ground mental disorder in evolutionary dysfunction by employing a similar strategy that he calls 'black box essentialism'. The first premiss in the reconstruction of the argument consists in something that is meant to follow conceptually or analytically from our concept of mental disorder. The notion is that in order to grasp the notion of mental disorder one must grasp that there is something wrong or dysfunctional about a person who has one. The second premiss of Wakefield's argument is also meant to follow conceptually or analytically from our concept of mental disorder. Wakefield maintains that it simply follows from our concept that the nature of the bio-medical dysfunction that he arrived at in premiss one is something that is for science to discover. This is, in effect, to treat the relevant sense of 'bio-medical dysfunction' to be a natural kind term whose essential nature (reference or B intension) is to be discovered by science.

The third premiss consists in an empirical claim that is meant to be revisable in the face of future empirical evidence. The notion is that as the best current chemical theory holds that water is H_2O and that gold has atomic number 79 that the process for fixing biological functions and dysfunctions is evolution by natural selection. The conclusion thus follows analytically from the premisses: *Given that* our notion of BIO-MEDICAL DISORDER entails BIOLOGICAL DYSFUNCTION (as asserted in premiss one); and *given that* BIOLOGICAL DYSFUNCTION is a natural kind term (which is to say that it tracks the B intension as asserted in premiss two); then given that science tells us that BIOLOGICAL DYSFUNCTIONS are fixed by evolution by natural selection (as asserted in premiss three); it follows analytically or conceptually from those premisses that BIO-MEDICAL DISORDERS are (at least) failures of evolutionary function.

Wakefield (2004, p.79) maintains that the Harmful Dysfunction analysis of the concept of mental disorder is Black Box Essentialist.

> the proposed concepts are essentialist because category member-

> ship is ultimately determined not by observable properties (e.g., for water, clear, thirst-quenching liquid) but by the hypothesized theoretical property of "inner nature that explains observed features (for water H_2O). The proposed concepts are black box because, rather than defining concepts by specific theoretical properties (e.g., "water is H_2O"), such concepts postulate a theoretical explanatory structure and remain agnostic on its identity, which may be unknown (e.g., "water is anything that has the same substance-essence as the clear thirst-quenching liquid in the lakes and rivers)

Wakefield thus treats the concept of mental disorder as a natural kind concept in the way that water and and gold are. He maintains that the identification of mental disorders with harmful dysfunctions proceeds in three stages: Firstly, it is a-priori to our concept of mental disorder that disorder is a dysfunction. He elaborates on this in stating that:

> a disorder exists only when an internal mechanism is dysfunctional, specifically in the sense that it is incapable of performing one of its natural functions (Wakefield, 1999, p.375).

He also maintains that at this stage of the analysis, natural function is used in an intuitive sense that has existed for millennia, not in a technical evolutionary sense.

In the second stage, Wakefield maintains that the seemingly anthropomorphic notion of the function of a biological mechanisms is analysed in straightforward scientific causal terms The language of function is used to indicate that certain effects of biological mechanisms are so complex, beneficial, and intricately structured that they cannot be accidental side-effects of random causal processes but, like the intentionally designed functions of artefacts, must somehow be part of the explanation of why the underlying mechanisms exist and are structured as they are Assertions that certain effects of a mechanism are useful do not offer any explanation of the mechanism; the usefulness could be due to chance. In contrast, function attributions implic-

itly make an explanatory claim, namely, that the mechanism is the way it is partly because of its usefulness. Disorders, then, are failures of mechanisms to perform their natural functions, where natural function is understood in the aforementioned explanatory sense.

He then maintains that strictly speaking, these two steps complete the conceptual analysis of disorder. However, this analysis does not explain how an effect (e.g., pumping, seeing) could explain its own cause (the heart, the eyes), nor does the analysis provide a criterion by which one can scientifically distinguish natural functions from other effects in a manner more precise than that afforded by common-sense intuitions. The analysis inevitably leads to the question, what kind of underlying process could possibly be responsible for such seeming design in natural systems without any designer? To answer this question, there needs to be a scientific theory of how such explanatory effects can come about. The attempt to answer this question leads to a third step in the argument: Evolutionary theory provides the only plausible scientific account that presently exists of how the natural functions of a mechanism can explain the existence and structure of the mechanism The third, theoretical argument leads to the conclusion that disorders are failures of mechanisms to perform functions for which they were naturally selected.

Wakefield is thus led to identify mental disorder with a failure of an internal mechanism to perform its evolutionary function. Our concept of water is such that it is transparent, potable etc. Our concept of water is also such that water is a substance. Best scientific theory then tells us the underlying property of the substance that is responsible for the properties that featured in our concept. The essential property of water is thus the property that the scientists have discovered. Wakefield similarly thinks that our concept of mental disorder is such that it is a harmful dysfunction. Our concept of harmful dysfunction is also that there is a causal process that fixes the functions and dysfunctions. Best scientific theory then tells us that the underlying causal process is evolution by natural selection. The essential property of mental disorder is thus the property that scientists have discovered.

The "black box essentialist account I (Wakefield, 1997, 1999b, 2000) present

is one flavour of such essentialist views; the name and a few nuances are mine but the basic ideas are derived from the noted philosophers. The proposed concepts are essentialist because category membership is ultimately determined not by observable properties (e.g., for water, clear, thirst quenching liquid) but by the hypothesised theoretical property or "inner nature that explains the observed features (for water, H_2O. The proposed concepts are black box because, rather than defining concepts by specific theoretical properties (e.g., "water is H_2O), such concepts postulate a theoretical explanatory structure and remain agnostic on its identity, which may be unknown (e.g., "Water is anything that has the same substance-essence as the clear thirst quenching liquid in the lakes and rivers). This essentialist definition uses the prototype properties not as universal criteria for the construct but only to indirectly refer to its essence. Thus, the definition allows things very different from the prototype set, such as ice, steam, or H_2O atoms floating in space, to be water. The description based on the prototype sample allows us to fix the reference of the construct term by a closed concept, and all other propositions remain "open but are not part of the concept (Wakefield, 2004, p.79). According to black box essentialism:

> Roughly, schizophrenia might be defined as follows: Take as the prototypical set those who Bleuler originally picked out as clear cases of schizophrenia, when he defined the concept; an individual then falls under the concept of schizophrenia if he or she possesses the underlying psychopathological structure that was shared by most of that prototypical set and explains the symptoms that led to their being placed in the set. (Wakefield, 2004, p.81)

On scientific discovery:

> But isn't it possible, however remotely, that temperature could turn out not to be mean kinetic molecular motion after all, or that water could yet turn out not to be H_2O? We could wake up tomorrow, for example, and find that chemists had discovered that their instruments had been systematically mis-calibrated and their instruments readings misinterpreted, and that the liq-

> uid in the familiar lakes and rivers has as its molecular structure not H_2O but, say, XYZ. If that happened, we would surely not conclude that there is no water in the lakes and rivers Rather, we would conclude that water is not H_2O after all, but XYZ showing that even theoretical reductions get their legitimacy from whether they in fact match out pre-theoretical concept. (Wakefield, 2004, p.81)

Modal intuitions

At this point one might well be wondering how much mental disorder is like water or gold. An analysis that is perhaps a little closer to home is David Lewis with 'Mad Pain, Martian Pain'. Lewis argues that since pain in us (human beings, higher animals) turns out (or probably will turn out if science continues on its business) to be brain state x (where x is a placeholder for whatever it is that it turns out to be) that pain is to be identified with brain state x. One upshot of this is (of course with an identity) that if we then discover Martians without a nervous system then they can't have pain. This is just to say that when the descriptive features come apart from the reference the dennotation tracks the referent. I will have more to say about this in the chapter on natural kinds.

Also artefacts. What is thought to be essential about pens is that they are designed by agents with certain intentions. Or (sometimes) that they are in fact used by agents for a certain purpose. I will discuss the notion of different kinds of kinds later. In particular, it is an open question whether there are natural kinds (disorder, mental disorder, schizophrenia). Or (putting things another way) it is unclear what kinds of things they will turn out to be.

While Wakefield does attempt to offer a rough analysis of the concept of biological function his analysis is very rough indeed and it is probably fair to say that it raises at least as many issues as it helps illuminate. The general approach is familiar to us, however, from Dretske, Millikan, and Neander's work on biological function. Very roughly, we can say that

- M has the function of causing behaviour B* iff
- 1) M has been naturally selected in virtue of causing B*

Critics have rightly pointed out that M could have been selected for B in our evolutionary past, but be maintained in current populations in virtue of causing C. One example of this would be that the mechanism that subserve language were selected for one function in our evolutionary past, and yet they seem to have an acquired function of subserving language now so that if language was impaired due to their failure this would be a genuine instance of malfunction. Wakefield responds to this objection by clarifying the role of evolutionary history by natural selection:

> an effect is a function only if it plays a continuing role in explaining the maintenance into the present generation (i.e., continued existence) of the mechanism in the species. A former function that ceased exerting selective pressure long ago is not currently a function because it has no role in explaining current species-typical structure. (Wakefield, 2003 p.979). dysfunction as factual.

Thus Wakefield's revised view thus seems to be that:

- M has the function of causing behaviour B* iff:
- 1) M is maintained in the population (by natural selection) in virtue of causing B*

The biological notion of function is thus thought to be fixed by objective facts about the mechanisms and facts about evolution by natural selection.

Thus, according to Wakefield a clinician is justified in maintaining that X is mentally disordered iff:

- 1) The clinician judges that according to the best theory of B*, B* is caused by a mal-functioning mechanism

the HD analysis is an analysis of the concept of disorder, not a theory of the mechanisms or dysfunctions underlying disorders (Wakefield, 2003, p. 978

2003 dysfunction as factual).

Behaviour vs inner mechanism

Wakefield differs from the Clinicians handbook The Diagnostic and Statistical Manual of Mental Disorders (DSM) not only in maintaining that natural selection fixes bio-medical function, but also in maintaining that mental disorder is the result of malfunctions in mechanisms that are internal to the individual.

The clinicians handbook The Diagnostic and Statistical Manual of Mental Disorders (DSM) concurs with Wakefield that dysfunction is necessary for mental disorder but instead of maintaining the dysfunction must be within the individual they consider behavioural dysfunction as well. The DSM asserts that in order to diagnose mental disorder it must currently be considered a manifestation of a behavioural, psychological, or biological dysfunction in the individual DSM xxxi. Wakefield criticises the DSM as failing to distinguish between problems in living, where the dysfunction may be purely behavioural, and mental disorder, where the dysfunction is the cause of the behaviour. failure to be capable of the action of reading is a disorder when and only when the failure is due to an underlying dysfunction, and it is not a disorder when the failure is due to a non-dysfunction such as lack of education (p. 18-19 aristotle as sociobiologist).

There is a proliferation of different theories of mental disorders which has led the DSM to attempt to specify mental disorder in a theory neutral way. They thus stick to listing observable behavioural symptoms and they try to refrain from commenting on underlying processes or causes of mental disorder.

Firstly, they don't specify the notion of function or malfunction, and secondly, Wakefield criticises the DSM for allowing there to be behavioural dysfunction in the absence of dysfunction within the individual. He maintains that we distinguish between mental disorder and problems in living and this distinction is captured by whether one has an inner dysfunction or not. The DSM allows the dysfunction to be behavioural or within the person and hence the

DSM fails to capture the mental disorder / problems in living distinction.

It might be the case that the difference between Wakefield and the DSM is merely terminological. The DSM might be working with a more liberal notion of mental illness, whereas Wakefield is dealing with a narrower conception.

An important thing to ask at this point is what does the notion of mental disorder do for us? Mental disorder and responsibility. Mental disorder and treatment. These seem to come apart, however. What else is it supposed to do? Scientifically interesting kinds Taxonomy. The DSM attempts to provide criteria for mental disorders that are theory neutral and thus largely consist of behavioural symptoms. Wakefield maintains that as a result of this failure to consider aetiology or cause the DSM is over-inclusive and people may meet DSM criteria for mental disorder even though intuitively they are not mentally disordered. One example of this is the diagnostic category of reading disorders where people meet criteria or not based on their reading ability. Wakefield maintains that it is important whether the person meets criteria on the basis of inner mechanism malfunction or whether the person meets criteria on the basis of insufficient instruction. He maintains, fairly intuitively, that only the former are mentally disordered and the DSM thus fails to distinguish between mental disorders and problems in living that are not due to mental disorder.

3.3.3 Malfunction of a person?

Wakefield maintains that whether an individual is mentally disordered or not is determined by whether their harmful symptoms are due to the presence of internal mechanism malfunction. He maintains that we should judge that a person is mentally disordered only when according to the best theory of their symptoms the symptoms are caused by objective malfunction within the individual. This involves the best theory we have of the sciences of the mind and he puts special weight on evolutionary psychology with respect to fixing the function of inner mechanisms. These two things can come apart as when we do not judge that a behavioural symptom is due to internal malfunction

even though it is, and when we judge that a behavioural symptom is due to internal malfunction even though it isn't. As Wakefield is concerned about providing a conceptual analysis of the concept of mental disorder he attempts to offer examples that are both in line with his analysis of mental disorder and our intuitive judgements as to whether the person is mentally disordered.

Chapter 4

Dysfunction I

4.1 Chapter introduction

In this chapter I will consider two accounts of function. The first is the teleological approach that is standardly though to be normative, or to appeal to the intentions of intelligent agents. The second is the mathematical approach that was the first step from teleology to science. My aim in introducing these notions is to show that there are a number of different ways in which we can characterize talk of function and dysfunction or, alternatively, that there are a number of different conceptions of function and dysfunction. This casts doubt on Wakefield's argument that the evolutionary notion is the only serious contender.

4.2 Teleology

Teleological accounts share the feature of appealing to a telos, intention, or purpose of an agent. For example, Aristotle thought that the function of a human being was rationality. The idea (roughly) is that when we ask what it is that is essential to being human we need to look for a property that humans have that other species lack. Aristotle thought that man (or humans) were rational agents. He thus thought that what is essential to humans is that we

are rational agents. The function of a human was thus thought to be reason.

The Aristotelian view might be thought to have good prospects for psychiatry insofar as it seems fairly intuitive that mental disorders (at least some of them) are disorders of irrationality. Irrationality seems to play a significant role in extra-scientific concerns that we have - such as with respect to involuntary confinement for psychiatry and the insanity defence for law.

Christopher Megone maintains that two-stage views of psychiatry fail to show psychiatry to be grounded in the sciences because the notions of 'function' and 'dysfunction' turn out to be evaluative. He thinks that the relevant notion of function and dysfunction is Aristotelian[1]. Megone considers this notion of function and dysfunction to be normative or evaluative in a way that renders the two-stage view unable to separate out matters of empirical fact from matters of normative value.

4.2.1 Aristotle

More recently Christopher Megone has defended a view of disorder that is essentially evaluative. His problem with the two-stage view is that he is not persuaded that a non-evaluative account of dysfunction can be given and hence he maintains that the two-stage view smuggles values in the back door in the supposedly non-evaluative notion of dysfunction. Megone's argument for this is inspired by Aristotle's notion of function and the Aristotelian notion of the function of different kinds of people. While naturalists about function and malfunction are unlikely to be persuaded by Megone's appeal to Aristotle's teleological notion of function and malfunction I do think that there are some legitimate concerns about the normativity of evaluative aspects of the relevant notions of function. I now wish to consider some of the features of the natural world that have been appealed to by those who endorse a two-stage view of mental disorder. I shall show that all of the following strategies are inadequate to ground the relevant notion of dysfunc-

[1]I am more interested in Megone's account and assessing it's plausibility than I am interested in whether or not Megone has correctly interpreted Aristotle.

tion (abnormality, illness, malady, etc) as non-normative or evaluative before going on to offer an alternative account in the next section.

The teleological notion of function is commonly thought to have originated with Aristotle. Aristotle argued that the essence of a person was that persons are rational animals because being a rational animal is the only property (or properties) that persons have that nothing else has[2]. He went on to argue that since being a rational animal is the essence of a person, being a rational animal is required in order for an instance to be a good instance of personhood or a functioning or flourishing person. For Aristotle, being a good instance of personhood amounts to the instance well approximating the ideal form of personhood - in this particular example, approximating the ideal form of rational animality.

The Aristotelian notion of function is commonly thought to be forward looking in the sense that the good for an instance is fixed by the ideal form or essence. It is also thought to be prescriptive or normative in the sense that it would be better for a person to develop their rational and physical capacities since these are an important part of what it is to be a good person, what is good for a person, and what is required in order to be a good instance of personhood since these amount to the same for Aristotle.

(Not sure whether it is worth mentioning Megone's view... It seems to me that Aristotle was attempting to offer an analysis of 'good' in terms of the notions of 'flourishing' / 'health' for a 'kind' or 'type' (e.g., acorns). Megone is doing something extremely weird in attempting to ground the notion of 'flourishing' / 'health' in the notion of 'good' for a 'kind'. It is looking like a pretty tight circle to me where each notion in the circle is problematic and where no independent specification is offered... I simply don't see how this kind of account is supposed to get up off the ground in the first place... Maybe as an analysis of 'good' - which is what Aristotle was trying to do with it - Moore aside - but it seems to have pretty crap prospects as an account of health.)

[2]I will deal with problems individuating properties in the section on kinds of kinds.

It is an interesting idea to take the notion of health as primitive... But this won't do if we want an account of health. Maybe we don't really care about an account of the notion of health. Maybe we need to return to the questions of interest and none of this thus far is particularly relevant to that, at all.

While there are different ways of characterizing what it is that makes an instance a member of one kind as opposed to another a fairly standard view is that functions and dysfunctions are properties of individuals that are conferred on the individuals in virtue of their membership in a kind. Aristotle seems to have this intuition and I will go on to show how the other theories also attempt to fix functions and dysfunctions in the relation that an instance has to an ideal, kind, type, or form[3]. While Aristotle has the argument from the only property/s that a kind has to the function or good of a kind he doesn't provide an independent criteria for types. As such the issue arises what justifies our regarding an instance to be a dysfunctioning instance of one type rather than a functioning instance of another?

One feature of Aristotle's view is that there are mind-independent facts about what forms there are that determine kind membership. As such, kinds aren't merely arbitrary mereological fusions. Aristotle's view of kind-hood is also fairly broad, however, in the sense that it isn't merely a theory of what determines natural kind membership, rather it is a theory that determines kind membership quite generally. Aristotle thought that there were ideal forms of geometric shapes, for instance, where the kind of shapes in the actual world was fixed by their resemblance to an ideal. He also considers the form of a person and here we can see that he is interested in more than natural kinds as he could well grant that the essence of a person is different from the essence of homo sapiens in the sense that there may well be homo sapiens that lack rationality to the extent that they are not persons. It is also important to note that Aristotle has an essentialist view of kind membership;

[3]It is crucial whether we conceptualize mental disorders as themselves being kinds of kinds or whether we conceptualize mental disorders as being failures in an instants approximation of certain kinds of kinds. I will attempt a solution in the chapter on kinds of kinds.

an issue taken up in the chapter on kinds of kinds.

Once we have fixed the forms (which were mind-independent for Aristotle) then an instants membership in a kind was thought to be fixed or determined by the relation that the instant bears to the form. There are two different ways that we can characterize the relevant relation (depending on how one individuates properties). The first is to say that since similarity or resemblance comes cheap (one thing resembles every other thing in at least one respect it is the degree of similarity that determines that an instant is an instant of one kind compared to another. The second is to say that there is something about the *degree* of similarity that renders it different in kind such that the instant bears a relational property to its form that it doesn't bear to any other form. The first way of characterizing the relation has it as a matter of degree (such that an instant can more or less well resemble its form). This seems to capture the intuition that there can be better or worse examples of the forms such that functionality and dysfunctionality come in degrees. The second way of characterizing the relation captures the intuition that there is some essence that the form has that is exemplified by instances that are of the kind that no other instances have. This would have it such that there is a fact about whether an instant is a member of one kind compared to another. We can see that there is a tension here that I will go on to show is also a tension for other views of function and dysfunction. The central tension is between the resemblance needing to be great enough such that the instant counts as an instant of the kind and yet weak enough such that the instant counts as a dysfunctioning instant of the kind rather than a functioning instant of some other kind.

Aristotle's view is top-down in the sense that he starts with persons. The systemic view similarly proceeds in an organizationally top-down way in that it attributes functions to components in virtue of the role they play in a greater system.

of starting from what is usually considered to be the uppermost level of the hierarchy that is relevant for anatomy and physiology. While I will discuss the hierarchy of levels of organization more in a subsequent section by way

of preview the levels in the hierarchy that are typically considered are (from bottom to top) the chemical, organelle, cellular, tissue, organ, organ system, organism. While Aristotle doesn't consider the organism from the perspective of homo sapiens he focuses on the essential features of a person instead. The essential features of a person are thought to be provided by what is distinctive about them and for Aristotle the distinctive feature was thought to be rational animality. Since rationality is forward looking (with respect to means ends reasoning at least) we can see that Aristotle focuses on the goal directedness of persons. The functions of components (e.g., the function of rationality) is to look out for what is good for the organism. This is a feature that seems to be shared with the systemic notion of function that similarly proceeds in an organizationally top-down way, at least (I will go on to show) at times[4].

Paley

Paley is best known for his argument from design. He thought it was intuitively obvious that there were certain features of the biological world that were so intricate and clearly adapted to fulfil some purpose or function in the world that the best explanation for those features was that the biological phenomena was created by an intelligent designer with that purpose or function in mind. While Paley's argument seemed most plausible before evolution by natural selection provided an alternative explanation to what seemed to be adaptive features of biological phenomena it is important to note that Paley's notion of function is commonly applied to artefacts. What is controversial about its use in biology is whether there is scientific utility in considering the biological world to be an artefact.

Paley's argument begins with the observation that if we found a watch on

[4]I have concerns that I'm sliding into Plato here. The relation between the good of the person and the good of the state is something that I'm hoping may turn out to be analogous to extending physiology to consider how the behaviour of a person in relation to its environment is an upper level of organization that needs to be added to the organizational hierarchy for the purposes of psychiatry at least. Whether that will take us into the realm of rationality remains to be seen - but a case can definitely be made for it.

the beach that is tracking the time then it would just seem intuitively obvious to us that the watch had been designed by an intelligent agent with the intention that it keep time. The intricate way that the parts work together in order to produce an obviously intelligent output strongly inductively suggests, according to Paley, that the watch was designed by an intelligent agent for the purpose or function of tracking the time. While Paley didn't discuss dysfunctions we can see how the account can cover this as if an intelligent designer created the watch with the intention that it track time then if the watch were to fail to track the time it would be dysfunctioning insofar as it wasn't performing its function as fixed by an intelligent creator agent.

Paley's argument continues with the observation that there are features of the biological world that similarly suggest that they have been designed by an intelligent creator or designer. Certain features of the biological world seem to be maintained by a complex arrangement of parts and those features seem to be so 'intelligent' or 'adaptive' with respect to tracking features in the world as to have been designed by an intelligent designer with a certain intention. Eyes just intuitively seem to be *for* seeing and seeing just seems to be so obviously adaptive and beneficial to the organism such that if they fail in this respect they seem to be malfunctioning. Paley thought that the obvious conclusion to draw from the presence of such obviously adaptive features was that they were designed by an intelligent designer (a creator God) with a certain intention in mind. We can conclude that according to Paley, the function of the eye is to see (or, eyes are *for seeing*) because that is what the intelligent designer (God) intended the eye to do.

In his commentary on the first premiss Wakefield maintains that according to our pre-theoretic notions of 'function' and 'dysfunction' we are uncommitted to what it is that fixes the relevant functions and dysfunctions. He states that it is perfectly consistent with our pre- theoretic notion that the functions and dysfunctions are fixed by the intentions of an intelligent designer or by a creator God. Wakefield then employs a sub-argument to lead to the conclusion that the *historic* process that was responsible for certain traits or features being present in present populations that is the process that fixes

the relevant functions, however. It seems to be to be far from clear that the historical process was the important feature of Paley's argument from design, however. It seems rather that the *obvious purposiveness* was the important feature. Some evolutionary accounts of function (propensity views) focus on the forward-looking aspects of evolution by natural selection rather than the causal-historical aspects in order to respect the teleological intuition and thus it is not clear how the pre-theoretic notion of function is essentially historical.

While the teleological notion of function might seem to be an unscientific notion of function it is important to see that it is often accepted by naturalistically inclined philosophers as (at least a partial) account of artefacts. Part of the story of the function of artefacts is thought to be the intentions or purposes of an intelligent designer (e.g., Putnam on twin earth pens). The common-sense view of mental disorders as failures of rationality (which would seem to have as much a claim to following conceptually from our notion of mental disorder as the dysfunction intuition does) might also be thought to have something to do with the notions of intention. Not the intentions of an intelligent designer to be sure, but the notion that an agents intentions can be evaluable against a standard of rationality (such as Bayesian norms of probabilistic reasoning). Evolution by natural selection isn't thought to be relevant for fixing Bayesian norms of probabilistic reasoning, but it seems that we shouldn't be too quick to write off the irrationality account of dysfunction even though it doesn't appeal to evolution by natural selection.

One important thing to note about the teleological notion of function before moving on to accounts that are typically thought to be more scientific or naturalistic is that rather than their being one teleological notion there seem to be different notions of function that are broadly teleological. One could thus have different teleological theories according to differences in which agent is meant to fix the functions, or according to the role that the intentions play in fixing the functions. This is important as I want to maintain that rather than there being just a few different notions of function on the table there are a whole bunch of notions. While they may be broadly carved up into the teleological / rational, mathematical / statistical, evolutionary, and sys-

temic, there are many different notions of function and dysfunction within these broad approaches. Seeing that this is the case problematizes Wakefield's thought that the relevant process for fixing functions and dysfunctions is something that follows unproblematically from our concept of bio-medical disorder.

4.2.2 Artefacts

While Paley appealed to the intentions of an intelligent designer for fixing the function of artefacts Putman also seems to grant that the function of artefacts (and indeed the feature that makes an artefact a member of its kind as an artefact) is the intentions of an intelligent designer. Putnam considers a twin earth scenario where we find a lump of matter that bears a structural or morphological similarity to a pen. Putnam argues that despite the superficial resemblance the lump of matter bears to a pen if the lump of matter was not designed by an agent with a certain purpose in mind (that it be a pen) then the lump of matter is not a pen. This way of capturing what is both essential to artefacts and what is function fixing for artefacts results in problems around how we specify or know precisely what the agents intentions are. If an agent makes a pen with the intention that it never be used as such then is it still a pen? If a factory mass produces them then are they pens? These aren't issues that I'll become bogged down in.

This is to focus on production. We could also focus on consumption. Structural similarity - resemblance. Insofar as people don't share Putnam's intuition that intentions are all important in fixing kind-hood membership and / or function and dysfunction or artefacts one might have a view whereby morphological similarity or resemblance is enough.

Putnam seems to be picking up on an aspect of Paley's view when he argues that the intentions of an intelligent designer are crucial for fixing kind membership for artefacts. He considers a lump of matter in some possible world that strongly resembles a pen with respect to morphological or structural features and concludes that insofar as the lump of matter was not designed

by an agent with the intention that it be used in certain ways the lump of matter is not a pen. Insofar as we share Putnam's intuitions we seem to be sharing the intuition that the intentions of an intelligent designer are crucial for fixing the kind of artefact and also the function of artefacts (and hence what would constitute a dysfunction). Historical - vs propensity. vs morphological structural similarity.

There is an alternative to this view that is based on structural or morphological similarity, however. On this view the intentions of the producer aren't as relevant or aren't the only aspect that is relevant for fixing function and dysfunction. On this view the use to which a consumer puts a lump of matter is either an additional factor or a more crucial factor. On this view if I take a lump of matter such as a rock and use it as a doorstop then the lump of matter counts as a doorstop in virtue of my using it as such. There seems to be something to this intuition as well. It is something that will be considered in later sections.

Insofar as we do not share Putnam's intuition (and I think a plausible case could be made for this line as well) there might be something to the intuition that there is at least a sense of function in which functions aren't fixed by the intentions of a designer so much as the intentions of a consumer. So, for instance, if I take a lump of matter like a rock and I use it as a paperweight then even though geological processes didn't fix that the function of the rock was to be a paperweight my actual using it as a paperweight might confer this function on it.

It is interesting to consider on these notions whether they have the resources to account for dysfunction. If a designer creates a chair, for instance, then can the chair malfunction? This line might help bolster the primary notion of function for artefacts being fixed by intentions. Insofar as we have worried about swamp pens and the like we might have the intuition that there is something plausible about either consumption or resemblance.

The notion of function has often seemed problematic to philosophers, however, because of the role that function talk has played with respect to at-

tempts to naturalize certain phenomenon. For example, a feature of intentionality (might be) that where there is representation there is the capacity for misrepresentation. One might similarly maintain that where there is function there is the capacity for malfunction. One might then be led to attempt to naturalize intentionality or representational content with respect to functions and malfunctions. If one could ground representations in functions and functions were grounded in purely physical properties and processes then one would have successfully naturalized representations. While there is controversy over how much representation can be analysed in terms of function the main suspicion has been that there is something going on with respect to functions. Either functions are natural features in which case representation cannot be reduce to them or functions are evaluative / normative in which case even if it was possible to ground representation in function function wouldn't be grounded in physical properties and processes.

4.3 Statistical

4.3.1 Arithmetic mean

One notion of function is a notion of statistical normality. While the statistical notion doesn't come up as frequently in psychiatry it does frequently come up in abnormal psychology where one reading of abnormal is the statistical notion of abnormality. The first thing to note about the statistical notion is that it seems insufficient as an analysis of mental disorder. While abnormal psychology might be construed as the study of psychological processes or behaviours that are statistically infrequent not all statistically infrequent psychological processes or behaviours are the subject matter of psychiatry. An IQ that is more than two standard deviations above or below the mean is abnormal, for example, but an IQ that is two standard deviations above the mean is not a mental disorder. Similarly, Mozart's musical ability was statistically abnormal in the sense that not many other people have comparable musical ability and yet he is not mentally disordered in virtue of his musical ability being statistically abnormal.

In response to these objections one could reply that not all statistical abnormalities are mental disorders because it is only the statistical abnormalities that result in harm that are mental disorders. If the statistical notion of abnormality is the correct analysis of malfunction then it might need to be supplemented with the harm condition as Wakefield maintained.

The main problem with the statistical notion of malfunction is that it rules out the possibility that mental illnesses could occur very frequently. With respect to the general medical notion of disorder it certainly seems possible that the majority of people could suffer from a medical condition such as infestation by parasites or high blood pressure or an infection. If mental disorders weren't similarly able to be experienced by the majority of the population then it would seem that the notion of disorder in play in psychiatry would be different from the notion that is in play in general medicine. Boorse was the main proponent of the bio-statistical view. According to Boorse the notion of function and dysfunction in science are mathematical in the sense that 'normal' is the mathematical average whereas 'abnormal' is standard deviation from the mean.

There does seem to be a notion like this that is in play. If we consider how we get fMRI they are averaged across different populations. E.g., if we want to consider the 'schizophrenic brain' then we average the data across a number of individuals.

One of the problems for Boorse's account is how we fix the relevant reference class. It seems that intuitively some will be too broad. For instance, if we wish to consider the normal or average human reproductive system then we are going to get a funny view of humanity by averaging data across males and females. Boorse thought that relevant features for determining the reference class were species, sex, age etc. It seems that what is going to be normal or average is going to alter as we alter the specification of which individuals to include in the population that we are finding the average of. One might worry about arbitrary classes. We need a non-arbitrary way of fixing which features are relevant for determining the reference class given that normality and abnormality will crucially depend on how we fix the relevant population

in the first place.

A-symmetry between 'valued' and 'dysvalued' deviations from the mean. Many think that this notion needs to be supplemented with an evaluative aspect. My main issue here is that the statistical notion seems to be fixed once we have the relevant reference class. The issue is more likely to hinge on whether there is an non-arbitrary way of fixing that. Needs to not be merely disjunctive.

One notion that might be appealed to is that abnormality is simply a statistical notion. Hypertension and mental retardation are typically offered as examples of disorders that seem to be defined by their being abnormal in the statistical sense. The first thing to note is that not all statistical deviations are considered abnormal or deviant in the relevant sense, however. Mozarts musical talent was surely statistically abnormal or deviant in the statistical sense but it doesnt seem that we would consider him the subject matter of clinical psychology or psychiatry solely in virtue of his statistical abnormality. Similarly, members of MENSA are statistically abnormal (part of the criteria for membership is that the person scores in the top 2? 5? Percent) and yet such people are not thereby considered deviant in the relevant sense. There is much controversy over whether mental retardation is merely a statistical notion or whether there are mechanisms that result in mental retardation. While intelligence is described on a bell curve there are clusters of people who are found at the low end of the range and it could be the case that certain mechanisms are responsible for the clustering. Hypertension is also a controversial example. It is unclear whether we are best to think of hypertension as a disorder that is defined in terms of heart rate at the high end of the statistically normal range or whether we focus our attention on these people because they are prone to disorders or dysfunctions. It seems clear that only some statistical abnormalities are relevant and it seems possible that in the cases that are statistically deviant their status as a disorder is dependent on something other than the fact that the conditions are statistically deviant. It would seem possible that the entire population could suffer from parasites or broken limbs, for example, yet the fact that such conditions

were statistically normal would not seem to change our intuitions about the conditions pathological status.

I think that statistical deviance does have an important part to play with respect to the relevant notion of function but that the story is a lot more complicated than the simple notion suggests. Firstly, it is not the case that all deviance is considered pathological. Sometimes it is merely one end of the range and not the other. In the cases where statistical deviation seems relevant it seems that something else is going on.

4.3.2 Boorse

Christopher Boorse offered one of the first attempts to naturalize functions and dysfunctions in his bio-statistical account. According to Boorse functions are species typical whereas the dysfunctions are species a-typical. Statistical accounts are sometimes given of 'abnormality' and 'normality' in psychology, and the thought is that 'normality' picks out something along the lines of the statistical mean and that abnormality can be measured in standard deviations from it. Boorse's idea is that the statistical notion might have a wider application in grounding biological dysfunctions in medicine.

While Boorse thought that the relevance reference class for fixing statistical functions and dysfunctions was the entire species we may question this aspect of his view. In particular, other features are often thought to be relevant for fixing the reference class such as biological sex and age. Bio-statistical theorists could thus differ according to how they think the relevant reference class for medical and psychiatric disorder should be picked out. Another dimension of difference would be on the relevant threshold for delineating functional from dysfunctional variants. If we attempt to measure degree of dysfunction according to the number of standard deviations from the mean then it seems that we are defining dysfunction as anything that is one or more standard deviation from the mean. Different theorists could set different values on the degree of variation from the mean is within normal or functional variation and the degree to which variation entails dysfunction. One feature

of the bio-statistical view is that it allows that dysfunctions can be a matter of degree and it allows for genuinely borderline, indeterminate cases.

Boorse might be thought to be a main proponent of such an account. According to Boorse disorders are biological dysfunctions where biological dysfunctions are thought to be statistically infrequent. The idea of cashing out disorder or abnormality in terms of statistical abnormality is a fairly popular way to go in clinical psychology. Some nice features of the view is that it makes it a matter of objective fact whether an individual is disordered or not. Some problems with the bio-statistical account include that it seems strongly counter-intuitive that whether an individual is disordered or not could have so much to do with how things are with others. It also seems intuitive that an entire population could be disordered. Also that only some statistical abnormalities are relevant - e.g., the musical ability of Mozart.

It is at this point that the majority of theorists take the lesson that while attempting to cash out dysfunction with respect to some biological notion is a good idea, the statistical notion isn't the way to go and so they dismiss Boorse. Indeed, developing an adequate view of the biological dysfunction has been the subject matter of much work in teleosemantics and it wasn't long before theorists gave up on purely statistical accounts in favour of evolutionary or historic accounts. So the general consensus seems to be that while Boorse was onto some- thing with respect to analysing medical disorder into biological dysfunction analysing biological dysfunction into statistical abnormality fails.

Coopers has a three part theory of mental disorder. Her 'unluckiness' criterion might be thought to be a development of Boorse's bio-statistical account where she develops the idea of 'unluckiness' as a relation that obtains across possible worlds rather than within a world. So while it seemed a counter-example to Boorse's view that we think epidemics involve a large number of people in the society with the same disorder (and thus the disorder is statistically frequent) Cooper's attempts to cash out a notion as a relation between worlds where a disorder is a state that an individual is 'unlucky' to have in the sense that they would not have that condition in nearby possi-

ble worlds. Cooper's unluckiness criterion thus might be thought to be an attempted defence of Boorse insofar as her unluckiness criterion attempts to develop the statistical notion in more plausible ways. Cooper's also has two other criterion, however, in part to block the move that some statistical infrequencies might be positively valued,. It might be the case that Mozart was unlucky to have his ability in the sense that he lacked it in close possible worlds and yet we still wouldn't consider his musical ability to indicate dysfunction. Cooper's thus can only be interpreted as offering a partial defence of Boorse.

Her idea of cashing out disorder in terms of 'unluckiness' also seems problematic in that it shifts teh problem to one of needing an objective criterion on closeness of worlds. We need some way of knowing which worlds are possible and how far away each world is from this world in order to use the framework to tell us whether a condition is unlucky in the sense that Cooper's takes to be relevant. I haven't heard a satisfactory account of closeness of worlds. Is there a close world in which I (a genetic female) am genetically male? Is that world closer than one in which I (suffering from a virus) do not suffer from the virus? What do the answers to this imply about a person who is genetically xxy and a person who has a virus? If genotypes are necessary for the identity of individuals (as Kripke seemed to think) then what sense can be made of genetic disorders?

Boorse maintains that disorders are biological dysfunctions where biological dysfunctions are statistically infrequent. We can get a handle on the statistical infrequency intuition via clinical psychology where abnormality is often introduced as a statistical notion. Some nice features of the view is that it makes it a matter of objective fact firstly, whether an individual is disordered or not and secondly how disordered that person is. Some problems with the bio- statistical account include that it seems strongly counter-intuitive that whether an individual is disordered or not could have so much to do with how things are with others which determine where one falls on a statistical distribution. It also seems intuitive that a whole population could be disordered on the same dimension as in times of epidemic, for example. It also

seems that only some statistical abnormalities are appropriately regarded as of the right sort to count as biological dysfunctions. The musical ability of Mozart was statistically infrequent, for example, but do not render him dysfunctional on any usual understanding of dysfunction.

It is at this point that the majority of theorists take the lesson that while attempting to cash out dysfunction with respect to some biological notion is a good idea, the statistical notion isn't the way to go and so they dismiss Boorse at this point. Indeed, developing an adequate view of biological dysfunction has been the subject matter of much work in teleosemantics and it wasn't long before theorists gave up on purely statistical accounts in favour of evolutionary or historic accounts. So the general consensus seems to be that while Boorse was on to something with respect to analysing medical disorder into the notion of biological dysfunction analysing biological dysfunction in terms of statistical abnormality fails.

Coopers might be thought to be offering an extension of Boorse in surprising ways. She has a three part theory of mental disorder. Her 'unluckiness' criterion might be thought to be a development of Boorse's bio-statistical theory where she develops the idea of 'unluckiness' as a relation that obtains across possible worlds rather than within a world. So while it seemed to be a counter-example to Boorse that we think that epidemics involve a large number of people in the society (within a world) with the same disorder (thus a statistically frequent disorder) Cooper's attempts to cash out a notion as a relation between worlds where a disorder is a state that an individual is 'unlucky' to have in the sense that they would not have that condition in nearby possible worlds. Cooper's unluckiness criterion thus might be thought to be an attempted defence of Boorse insofar as her unluckiness criterion attempts to develop the statistical notion in more plausible ways. Cooper's also has two other criterion, however, in part to block the move that some statistical infrequencies (or 'unlucky' states in her technical sense) might be positively valued, beneficial, or lucky in the usual sense. It might be the case that Mozart was technically 'unlucky' to have such musical ability in the sense that he lacked it in close possible worlds and yet we still wouldn't

consider his musical ability to indicate dysfunction. Cooper's thus can only be interpreted as offering a partial defence of Boorse.

Her idea of cashing out disorder in terms of 'unluckiness' also seems problematic in that it shifts teh problem to one of needing an objective criterion on closeness of worlds. We need some way of knowing which worlds are possible and how far away each world is from this world in order to use the framework to tell us whether a condition is unlucky in the sense that Cooper's takes to be relevant. I haven't heard a satisfactory account of closeness of worlds. Is there a close world in which I (a genetic female) am genetically male? Is that world closer than one in which I (suffering from a virus) do not suffer from the virus? What do the answers to this imply about a person who is genetically xxy and a person who has a virus? If genotypes are necessary for the identity of individuals (as Kripke seemed to think) then what sense can be made of genetic disorders?

4.3.3 Logical, or functionalist

There is another notion of function that theorists don't often consider when looking at function and dysfunction talk in science. Often there is the claim that this notion is clearly irrelevant - but I think that it is not so obviously irrelevant and worth considering in more detail.

This notion is the mathematical notion whereby operators like '=' and '+' are considered 'mathematical functions' that take in values (e.g., numerals) to deliver a product. The operator specifies the transformation that is made from input to output.

The notion of function in 'functionalist' theories of mind etc seems to be a development of this mathematical notion. While there are many versions of functionalism (e.g., Turing machine functionalism, empirical functionalism etc on how we fix the values or variables that the computation is performed over) the notion is basically that mental states (e.g., beliefs, desires and so forth) are fixed by the function that they play with respect to the computation that they play over contents (values). Functionalism has a problem

accounting for dysfunction in the sense that if the 'function of belief is to accurately represent the world' then we have a problem of how people can have false beliefs. E.g., much debate around how certain kinds of delusions can be beliefs when 'delusional beliefs' don't seem to play the functional role of beliefs.

People have attempted to allow mental states (such as belief) to be both functionally characterized with respect to the role that they are supposed to or should or would typically play in the system and then explain certain other phenomena such as delusion as dysfunctions or deviations from that role. One way is to appeal to evolution by natural selection. Another way is to appeal to the mathematical average. Another way would be to appeal to the systemic view or the intentions of an intelligent designer. Seems that we need an account of dysfunction here too.

Chapter 5

Dysfunction II

5.1 Chapter introduction

Despite Wakefield's maintaining that the evolutionary approach is the only scientific game in town and that there is philosophical and biological consensus that the evolutionary account is the correct account of biological functions and dysfunctions the evolutionary view is not without its critics. More recently Davies has argued for a view of biological functions that is a-historic and along the lines of that offered by Cummins. In defending his systemic capacity approach to functions against the evolutionary view Davies argues that neither the evolutionary nor the systemic capacity view has the resources to naturalize dysfunction. If Davies is correct then this would seem to seriously undermine attempts to account for the role of science as one of their discovering the nature of the dysfunction that people have. In what follows we will consider both evolutionary and systemic capacity views before going on to state Davies objection that neither of them can adequately account for dysfunction. We will then return to the issue of the role of scientific facts and norms in bio-medical disorder and the nature of the relevant scientific facts in a later chapter.

We will begin with an account of the evolutionary and systemic views in a fairly broad sense. We will then turn to Davies objection. We will consider

a fairly obvious response to the objection and show what is wrong with it. We will then consider that if science can't discover dysfunction science still has a role to play with respect to the discovery of causal processes. While norms play a greater role than may have been supposed and while dysfunction turns out to be a non-necessary assumption rather than a discovery there is still much to be gained by scientific work that assumes either a systemic or evolutionary dysfunction aspect to bio-medical disorder.

5.2 Evolutionary

Wright was perhaps the first theorist to maintain that function and dysfunction were historical in a naturalistic way. The problem of offering a naturalistic or scientifically respectable account of biological function and dysfunction has long been a concern for philosophers and for philosophically inclined biologists. In the 1960's a number of philosophers attempted to naturalize talk of function and dysfunction in biology by appealing to evolution by natural selection as the naturalistic process that fixes functions and dysfunctions. The thought is that if talk of 'function' and 'dysfunction' in biology can be successfully translated into talk of evolutionary functions and dysfunctions then biologists use of the terms are unproblematic from a scientific point of view.

5.2.1 Evolution by natural selection

The basic thought here is that evolution by natural selection requires only three things. The first is that there be variation in some trait. The second is that there be competition for resources such that some variations of the trait result in greater relative fitness than others. The third is that that variation in the trait be heritable such that the offspring are more likely to have the variant of their parents than the other variants in the population. The thought is then that if that obtains evolution by natural selection could result in the population coming to be fixated on certain variants of the trait. Having a heart that pumps blood, for example, could be such that individuals

that lacked that variant would be at such a disadvantage that it would surely make sense to say that they had a dysfunctioning heart in the evolutionary sense.

The thought is that there are types of organisms in an environment that have variation in the tokens they have of some kind of trait. If some variants of the trait result in greater relative fitness than other variants and if that trait is more likely to be inherited by future generations of that organism then that trait can come to proliferate over the other variants. In this way certain features can come to be fixated in organisms (e.g., the majority of people are born with hearts and brains and limbs etc).

Millikan made much of the finding that the evolutionary function or a trait could come apart from the statistical function. Sperm achieve their biological function of fertilizing ova only statistically infrequently, however this doesn't change their evolutionary function, for example. Three things are often thought to be required for evolution by natural selection to occur. Firstly, there needs to be a trait that has different forms or variants. Secondly, some forms or variants need to be 'better adapted' to their environment such that they result in that variant surviving better than others. Thirdly, there needs to be a mechanism of heredity such that the 'better adapted' traits survive better than others in the sense that the proportion of variants alters due to the offspring of one variant being more likely to resemble their parents than the parents of others. If these three things obtain then evolution by natural selection will occur - which is just to say that the relative frequency of

Fixing function

Jerome Wakefield has appealed to work done by these theorists in his arguments that the evolutionary notion of function and dysfunction is the relevant notion for psychiatry and general medicine. While Wakefield does attempt to offer a rough analysis of how evolution by natural selection is supposed to fix the biological function his analysis is very rough indeed and it is probably fair to say that it raises at least as many issues as it helps illuminate. The general approach is familiar to us, however, from Dretske, Millikan, and

Neanders work on biological function. Very roughly, we can say that

- M has the function of causing behaviour B* iff
- 1) M has been naturally selected in virtue of causing B*

The notion of biological function has inspired a great deal of controversy within philosophy. Some theorists maintain that appealing to the process of evolution by natural selection is not enough to fix biological functions and malfunctions. One would need to add a clause to the effect that eyes enable us to see under normal conditions, for example, but it is unclear how one is supposed to go about specifying normal conditions. One cant simply appeal to conditions that were statistically frequent in our evolutionary pasts, for example, as it would have been dark around half the time and yet we wouldn't say that a person had a malfunctioning eye if they couldn't see in the dark.

Evolutionary views of function are similar in being historic. The thought is that our common-sense notion of function appeals to historic processes for fixing function. If we consider Paley's argument from design as employing some pre-theoretic notion of function then we can see how he appeals to a historic notion of function in his argument for the existence of an intelligent designer. Paley begins his argument by appealing to something that he takes to be obvious on the basis of observation - that certain features of the biological world seem to be very well or perfectly suited to performing some function or purpose in the environment the organism finds itself in. Paley proceeds to maintain that the best explanation that there is for the existence of such obviously adaptive or functional features in the biological world is that the traits were designed by an intelligent agent with such a purpose or function in mind. While Paley's argument is obviously super-natural in that it appeals to the intentions of an intelligent designer for fixing biological functions the evolutionary account of function is similar to Paley's notion in that it takes features of the world which are thought to be obviously adaptive on the basis of observing how well they fit into their environment and then proceeds to explain how they came into existence by way of appealing to a historical process. The evolutionary view differs from

Paley, however, in appealing to a natural rather than supernatural process in the attempt to explain how adaptations came into existence. Wakefield's argument is basically that our concept of function has us appealing to historic processes to explain the existence of biological adaptation because these are the processes that fix the function of the feature. We then look to science for the best theory as to how adaptive features come about and see that science says that evolution by natural selection is the relevant process that explains their apparent adaptedness. It thus follows that functions are fixed by the historical process of evolution by natural selection. And so we have a (very schematic) characterization of biological function as follows:

- F is the function of trait T (or a particular variant of trait T) in organism O in environment E iff:
- 1) Past instances of T in O did F in E
- 2) It is in virtue of T doing F that O's in E with T proliferated relative to other traits or other variants of T

Evolutionary accounts of function and dysfunction are similar in maintaining that the relevant process for fixing biological functions and dysfunctions is evolution by natural selection. While there are forwards looking, dispositional accounts of how evolution by natural selection fixes functions and dysfunctions I'll begin by considering the backwards looking, historical account that Wakefield ends up endorsing[1].

One statement of the historical evolutionary account is that:

- **Def** The function of a trait T in organism O in environment E is to x if and only if: Past tokens of T in past tokens of O had greater fitness (relative to other variants of T in O) in E.

[1]It is unclear whether he is aware that there are forwards looking accounts since he seems to conflate functions with historical explanations.

5.2.2 Malfunction fixing

Traits that aren't so good vs malfunctioning tokens of types. Need to sort this out in order to respond to Davies. It is a little tricky to see how one gets malfunctions or dysfunctions out of the evolutionary view. One analysis would be that traits which are selected against (are becoming less prevalent) in the face of the success of other variants are malfunctional or dysfunctional. Another analysis would be that traits which are not selected for are malfunctional or dysfunctional (trouble here distinguishing normal variation from dysfunctional variation - especially with something on a continuous scale like height and arguably IQ).

(There seem to be 2 ways we could go with respect to evolutionary dysfunction. One way is to consider traits that are different and then talk of functional and dysfunctional traits. Another is to talk of variations of traits and talk of functional and dysfunctional variants. Another (perhaps more relevant for disease?? unclear... Is to talk about trait (or variant) types vs trait (or variant) tokens where tokens can malfunction or dysfunction with respect to the functions that are assigned (though not defining) of the type).

Modelling

In sorting out the role of evolutionary theorizing of psychiatric phenomena we first need to get clearer on the explanandum or range of psychiatric phenomena that we might want to explain. Theorists typically attempt to explain either the origins or maintenance of traits or variations on traits in populations. Models can aspire to explain current prevalence or models can be dynamic modelling how proportions of different traits or variants of one trait alter over time.

While the *Diagnostic and Statistical Manual of Mental Disorders* offers symptom cluster analyses of psychiatric disorders some theorists have recently begun to focus their research interests into particular symptoms such as delusion or hallucination. Here it seems that we could view particular symptoms as traits or as variants on traits that we want to explain. Alternatively one could

attempt to explain symptom clusters or disorders. While one of the main critiques of the DSM is that it offers a categorical analysis of phenomena that might best be captured dimensionally this doesn't seem to pose a particular problem for evolutionary modelling. In particular there can be evolutionary models of phenomena that are continuous, such as height, and thus one could attempt to model anxiety or depression as continuous phenomena.

If we consider a population then it seems that there are different things that could be going on over time. If we consider a range of traits (or the trait that we are interested in along with 'rival' alternatives) or variations on a trait, or points on a continuum then the population could (with respect to these) stay the same, shift such that one achieves fixation, remain static with the proportions of different ones, cycle in predictable ways, or alter in unsystematic ways.

While it is tempting to think that evolution by natural selection would be most appropriately applied to the case where one has achieved fixation (is thus intuitively adaptive or functional) this need not be the case. A population could fixate on a trait due to drift (especially small populations seem most susceptible to this). Alternatively, a population could fixate on two traits where one is an inevitable by-product of the other and where the former is not selected for even though the latter is. In this case the trait would turn out to be a spandrel.

A number of theorists have provided a number of different accounts of evolutionary functions. In this section we will attempt to state the evolutionary view in a way that is fairly abstract, that thus applies to all or at least most, while also raising some of the issues that different theorists have disagreed on. Evolutionary function views seem to be similar in that they appeal to historical processes for fixing the function of various traits. While it is tempting to say they are similar in their appeal to evolution by natural selection as being the historical process that fixes the function of various traits most evolutionary function theorists acknowledge a role for non-evolutionary historical processes such as drift.

A number of theorists have provided a number of different accounts of evolutionary functions. In this section we will attempt to state the evolutionary view in a way that is fairly abstract and thus applies to all while also highlighting some of the differences that different theorists have taken. It is tempting to say that what evolutionary views of function seem to have in common is the notion that evolution by natural selection has an important role to play in fixing the function of various traits. While this may seem obvious enough, this will turn out to be too strong, however, as evolutionary function theorists often draw our attention to such processes as drift, and such phenomena as spandrels and ex-aptations that problematize attempts to focus exclusively on evolution by natural selection as a function fixing process.

What evolutionary views of function seem to have in common is the notion that evolution by natural selection has an important role to play in fixing the function of various traits.

There are a number of different ways that theorists have attempted to make sense of evolutionary functions. What evolutionary approaches to function share, however, is an appeal to evolutionary processes for fixing function. It is commonly accepted that in order for evolution by natural selection to occur three things need to obtain: Firstly, there needs to be variation in some trait T. Secondly, there needs to be differential fitness such that some variants outperform other variants (in a particular environment). Thirdly, there needs to be a mechanism of heredity such that variants 'outperform' other variants (in a particular environment) by way of that variant proliferating over other variants (due to survival of those variants and / or reproduction of new instances). If those three things obtain then the proportion of variants in the population will alter over time with some variants becoming more prevalent in the population (directional selection).

The thought is then that the variants that are 'selected for' (that are becoming more prevalent in the population) are the 'functional' variants. So, for instance, if we have some trait such as hearts and we have variation in the trait such that some hearts pump blood and others don't then if the hearts

that pump blood survive better / out reproduce alternative variants then the pumping hearts will be more prevalent in the population and the function of the heart (the evolutionary function) is to pump blood. According to the evolutionary view the function of a trait (T) is thus fixed by whatever feature of the trait resulted in the trait becoming more prevalent in the population.

There are some caveats that need to be noted for the evolutionary view. The first is that traits are only functional relative to an environment where one important aspect of the environment is the other variants that exist in the environment. Environmental change can thus result in functions shifting and directional selection (alterations in the frequency of variants) can result in functions shifting. Griffith's appeal to the most recent period of selection for fixing functions (at the present time - if that is what we are interested in). Important to note 'is this variation functional' can alter according to the society that the individual is located in (depending on how we cast populations). Polycystic ovarian syndrome or body type. Spandrels - free riders aren't difference makers but free riders. Also important to note different kinds of ways that evolution by natural selection can impact on populations. Directional selection, stabilizing selection, extinction, equilibrium etc.

Evolutionary views of function are historic in the sense that whatever historical process resulted in certain traits being present now is the process that fixes functions and also dys- functions. If we consider Paley's argument for the existence of an intelligent designer then we can see this intuition in play. On the basis of observing certain features of the natural world Paley thought it just obvious that some of them were so well adapted to playing some role in a particular environment that the best explanation for the adaptedness or functionality was to appeal to an intelligent designer who designed them with their purpose or function in mind. While Paley used this argument to argue from the existence of adaptedness or function to the existence of an intelligent designer we can consider his argument as proceeding from obvious adaptedness or function to an appeal to some historical process for fixing the adaptations or functions. The thought is then that while Paley was led to adopt a super-natural explanation for his observation the best scientific

theory that we have tells us that such adaptations or functions can arise from a process of evolution by natural selection. If we add a naturalistic requirement to the historical requirement then what we seem to be left with is that evolution by natural selection is what fixes functions and dysfunctions in the world since the adaptedness or functionality of certain traits were in fact fixed by the historical process of evolution by natural selection.

Also issues around whether there are different notions of evolutionary function hence dysfunction in play or whether they are fighting over getting the analysis of evolutionary dys- function right. Seems like different notions could have different utility in a scientific theory.

One (fairly abstract and rough) characterization of the evolutionary view is as follows:

- The function F of trait T in organism O in environment E is to B iff:
- i past instances of T in O did B in F
- ii T was heritable
- iii past performance of B by T in O in E resulted in O reproducing more than those that lacked B[2].

Unpacking this argument we can see that we are interested in organisms O that have traits T with certain effects in certain environments. An often overlooked feature of the evolutionary view is that the effects of traits in environments is going to vary (for adaptedness) depending on features of the environment. One of different grains of analysis.

Once we have a model of the distribution of variants in the population either at a time or over time the issue then becomes one of fixing the functional or adaptive variants. Biologists have long spoken of variants that are adaptive or functional and much work was done by philosophers of biology and philosophically inclined biologists to show that function and dysfunction talk in the biological sciences was scientifically reputable. While there are senses of

[2]The above characterization is adapted from Davies.

function and dysfunction which are decidedly normative theorists argued that there was at least one sense of function and dysfunction that was employed in the biological sciences that was fairly unproblematically translatable into talk of natural properties and processes. The thought is (roughly) that to say that the function is the heart is to pump blood is to say that past tokens of hearts that pumped conferred such a selective advantage on the individuals that possessed them relative to other variants that pumping blood is the evolutionary function of the heart. Conversely, since the evolutionary function of the heart is (roughly) to pump blood hearts that fail to pump blood are evolutionarily dysfunctioning.

While we have thus far considered the evolutionary notion of function it seems that what we really need for medicine and for psychiatry (on the assumption that biological dysfunction is necessary for medical disorder) is an account of dysfunctional. It doesn't follow that something doesn't perform a function to the idea that it is dysfunctioning with respect to that function. We need a way of differentiating the functions from the non-functions (the absence of function or difference of function) from the dysfunction. One might attempt to adopt a position that is kind of anti-Paley. Here the idea is that instead of it being intuitively obvious on the basis of observation which features are adaptive or functional it is instead intuitively obvious on the basis of observation which features are maladaptive or dysfunction. One might think in particular that it is obvious that mental or physical disorders are maladaptive or dysfunctional if anything is.

One thing that is important to note is that an assumption is not a discovery. We have seen already that there are alternative hypotheses for why it is that a trait results in fixation or seems obviously adaptive and similarly there could be alternative hypotheses for why it is that a trait is obviously maladaptive. In particular if we are interested in current adaptive value then this would seem to be fixed by relative number of offspring?

The issue of ex-aptation is the issue of how far back in history the evolutionary view considers functions to be fixed. While some have considered the majority of mental functions to be fixed in the plastocene Griffith's appeals

to the most recent period of evolutionary adaptation for fixing the functions that are presently existing now. While there might well be controversy as to how long ago the most recent period of evolutionary adaptation extends back it does seem that there can be facts about this.

With respect to drift we see that a trait might not have come about by way of evolutionary processes at all - or alternatively drift may play a more significant role in how a trait became prevalent in the population than evolution by natural selection plays in the trait.

5.3 Systemic capacity

This alternative view of function arose as a development of the work of Cummins. The basic idea is that while some areas of biology or some questions in biology might well be inquiring into historical processes there seems to be a sense of biological function that is not essentially historical. In order to see this we just need to consider the obvious truth of the claim 'Harvey understood the function of the heart centuries before Darwin'. The idea here is that when Harvey came to understand that the heart functioned as a pump within the circulatory system we learned something about the function of the heart even though we didn't learn anything at all about the general or specific historical processes that have resulted in hearts. While for the evolutionary theorist the questions 'how did x come about' and 'what is the function of x' are to be given the same answer for the systemic theorist these questions come apart while the questions 'what is the function of x' and 'what role does x play in some greater system' are equivalent.

On the systemic capacity account functions are assigned to components in virtue of the role that they play in the production of an output in some greater system. If one wants to give a systemic account of some trait or variation on a trait then firstly one appeals to some system that produces the phenomena that one wants to explain. Once one has the relevant system then one proceeds to analyze the system into components and assign functions to the components in virtue of the role they play with respect to the

production of the phenomena that one wants to explain. Davies maintains that it is important to note that assignment of function to components is relative in two respects. Firstly, which components are relevant is going to partly depend on what phenomena the researcher is interested in offering an account of. Secondly, which features of the components are functions is going to partly depend on what phenomena the researcher is interested in offering an account of. Despite these two aspects of the systemic capacity view being partly determined by the interests of the researcher Davies maintains that there are also several features of systems that are not dependent on the interests of the researcher.

We have already seen that systemic capacity analysis involves appealing to two distinct levels. There is the level of the phenomena and the system that produces it and there is the lower level with the components and their functions. Davies maintains that systems must consist in two distinct levels and once we hit a level at which the outputs are basic where the 'system' cannot be analyzed into further components then we have reached the end of the systemic capacity chain of explanation. Aside from this bedrock we can often reiterate the systemic capacity framework down - explaining the workings of the circulatory system, the heart, certain kinds of tissue, certain kinds of cell, and so on.

Davies enumerates the systemic view as follows:

- i I is capable of doing F,
- ii A appropriately and adequately accounts for S's capacity to C,
- iii A accounts for S's capacity to C, in part, by appealing to the capacity of I to do F.
- iv A specifies the physical mechanisms in S that implement the systemic capacities itemized in A.
-

5.3.1 Too cheap / observer relative

The main criticism of systemic functions is that they come too cheaply. If we identify some relevant output of a system then we can determine the function of the parts by seeing what they can contribute to the output of the system. It seems that there isn't enough constraints on the outputs It seems that it comes too cheap.

One thing that is interesting to note about both the evolutionary and systemic notions of function is that they have a much broader notion of function and dysfunction in mind than many supporters and critics have acknowledged. In particular, the systemic view seems to provide us with resources not only for attaching functions to bodily organ parts with respect to individuals, it also seems to provide us with resources for attaching functions to individuals with respect to the greater society. Perhaps even different societies with respect to some greater society. With respect to the evolutionary notion there is much controversy over the unit of evolution by natural selection. Some candidates are genes, neurological structures, cognitive capacities, and the behaviours of groups.

While the majority of theorists attempt to show systemic capacity functions to be grounded in evolutionary functions or to show that the different accounts are involved with different explanatory projects Davies argues that evolutionary functions turn out to be a certain kind of systemic capacity functions. He maintains that by viewing a population as a system and viewing members of a population as constituents of the system we can offer a systemic capacity analysis of the phenomena that we want to explain in a way that captures all the verdicts of the evolutionary view. He also maintains that it is an advantage of the systemic capacity view that it can help us understand what is going on with evolutionary explanation or modelling of other phenomena that the evolutionary function view can't explain such as drift. Davies seems right to be putting pressure on the evolutionary view to move from the adaptation assumption to other phenomena that is of evolutionary 'interest' even if it isn't straightforwardly explained by evolution

by natural selection.

While it might be thought to be a feature of Davies view that it can be applied to phenomena that the evolutionary view (at least in its simple version) can't explain it might be thought to be a vice of the view that it is over-inclusive. While Davies has no trouble applying the view to artefacts that produce things such as assembly lines intuitions are divided as to whether we want a unified account of artefacts alongside biological phenomena. While Davies briefly considers Godfrey-Smith's concern that the systemic capacity and evolutionary views are both important because they pick out importantly different causal chains at different levels of analysis he moves on from the objection and doesn't consider it further. I don't see the problem in reserving the term 'proper function' for solely functions arising from evolution by natural selection or for solely functions arising from the historical analysis of biological phenomena. Hard to know where to draw the line on mental phenomena but hard to distinguish psychiatry from neurology at any rate.

5.3.2 Systemic dysfunctions

One problem with attempting to ground medicine and psychiatry in the systemic rather than the evolutionary notion of function is that the systemic view (as enumerated by Cummins, anyway) doesn't allow us to differentiate dysfunction from the absence of function. Cummins enumeration is that the function of some part mechanism x is fixed by the causal contribution makes towards the output of the system. So the function of a heart valve might be (roughly) to regulate blood flow as the casual contribution the heart valve makes to the hearts pumping of blood is to regulate blood flow. The trouble is that if the valve fails to regulate blood flow then this view doesn't provide us the resources to say that the valve is malfunctioning. This is because if the valve doesn't play that causal role then it simply fails to have that function rather than it dysfunctioning.

Other theorists have attempted to develop the systemic notion of function in such a way that it can account for dysfunctions. One way of going about

this would be to make use of a type and token distinction. On this view the function of the heart valve is to regulate blood flow because this is what mechanisms of the valve type do with respect to contributing towards the hearts pumping of blood. Because the functions are type-functions rather than token-functions a token valve that failed to regulate blood flow could be described as malfunctioning because it is not playing its type function.

The problem with this view is that we need some independent way of stating how tokens get to be members of a type. If we are attempting to explain how the type has the function that it does then we can't say that a token is a member of a type in virtue of exhibiting the type function because part of what we are trying to explain is how the types have their function. We don't seem to have grounds for saying that a token is a dysfunctioning member of a type rather than saying that insofar as the token doesn't play its usual contributory role it fails to be a member of a type. And thus it lacks a function rather than dysfunctioning.

Davies (2000a, 2000b, 2001) develops a systemic view of function and he simply acknowledges that it doesn't have the resources to handle dysfunction talk - but then he maintains that the evolutionary notion can't adequately account for dysfunction either so that is no reason to adopt the evolutionary notion over the systemic notion. While it might not be an adequacy constraint on function talk in general that it can account for dysfunction it does seem that insofar as medicine and psychiatry attempt to ground their subject matter in dysfunction an adequate account of medical functions must be able to account for dysfunctions, however. If Davies is right that neither the systemic or the evolutionary notion can allow for dysfunction then this will have very skeptical implications for medicine indeed. While Davies does talk about medicine a little he doesn't seem to realize the role that dysfunction talk is supposed to play with respect to grounding medicine and psychiatry in particular in the natural sciences. He thus doesn't realize how significant his finding that neither can account for dysfunction would be with respect to medicine and psychiatry.

The obvious way to provide an account of dysfunction is to see functions as

properties of types. On this account a type (e.g., hearts) have a functional property (e.g., functioning as a pump). Particular tokens or instances of hearts can thus be functional or dysfunctional hearts depending on whether they function as a pump or not. What is needed for this style of account is for there to be properties that are sufficient to make a particular instance a member of the kind but where the functional property itself is not needed in order for the instance to be a member of the kind. If the functional property was needed for kind membership then we wouldn't have dysfunctioning hearts because an alleged heart that didn't have that property would not be a heart after all.

In defending the systemic account of function as being primary (where evolutionary functions are thought to be a subset of systemic functions) Davies offers an argument that seems to create a problem for naturalistic accounts of function more generally. While we considered briefly above that a number of theorists think that attempts to naturalize function are doomed to fail because of problems with dysfunction being normative and biology being non-normative Davies maintains that while it is commonly thought to be a virtue of evolutionary accounts of function that they can offer a naturalistic account of dysfunction he maintains that evolutionary accounts fail to do so. It is thus no objection to the systemic view that it cannot either. Davies doesn't seem to explicitly consider the role that dysfunction has played in attempts to natural- ize bio-medical disorder. As such it is hard to know whether he would be happy or unhappy with this implication of his view. It is worth considering whether one can get dysfunctions out of the evolutionary notion of dysfunction (or an alternative naturalistic account of systemic). This will better help us understand the role and limits of sciences contribution to fixing what conditions are bio-medical disorders.

Davies has fairly recently raised a couple of objections to the evolutionary view that are worth considering. If Davies objections are well founded then there would seem to be significant problems with appealing to the evolutionary view that haven't been properly unpacked in the literature. Davies notes that one of the great appeals of the evolutionary function view is that

theorists have done much in order to show that it can provide an account of malfunction or dysfunction. The main objection to the systemic function view is that the systemic view does not have the resources to account for dysfunction. Davies maintains that despite this common wisdom the evolutionary view is not able to provide an account of dysfunction. He maintains that as such it is no objection to the systemic view that it cannot. If Davies is right that neither the systemic or the evolutionary view can offer accounts of dysfunction then this will create a significant problem for the two-stage view insofar as the appeal to evolutionary and / or systemic functions is supposed to provide an account of biological dysfunctions which is supposed to be what science discovers about psychiatric and mental disorder. Davies also maintains that the evolutionary view can be shown to be a particular kind (or variant on) systemic capacity functions. The main objection to this line has been that one can account for dysfunction but the other cannot. This part is less relevant for here. If we can't get dysfunctions then that seems very problematic. Davies does not seem to have considered the implication of this or how theorists have attempted to use the terms in medicine and psychiatry.

Cummins systemic notion of function offers an alternative to analysing function and dysfunction talk in the biological sciences. While the evolutionary notion might be important in some aspects of biology (in evolutionary biology, most notably) it is far from obvious that it is the relevant notion in physiology and in other biological sciences that don't make explicit reference to the history of the trait.

The notion here is that we begin with some feature of biological systems that we would like to explain. We might want to explain vision, for example, or the circulatory system. What we then do is discover how there are mechanistic components that contribute to the explanandum. In the case of vision we discover that there are parts to the eye (e.g., the cornea and the lens) and that they each seem to contribute differently to the explanandum - vision. The systemic notion maintains that the functions are fixed by the contribution that the component part makes with respect to the relevant output of the

system that was our initial explanandum. This is thought to account for functions and dysfunctions in physiology in particular where physiologists often make no reference to evolution by natural selection.

The obvious move to make is to maintain that functions attach to trait types rather than trait tokens. A token of the type can thus be a functional token or a dysfunctional token. Davies has argued that in order to do this we need some independent way of characterizing types. I think that this is problematic. Perhaps one could appeal to morphology at this point? I'm not sure. I think there are significant problems with attempting to account for functions simplicitor... Observer relativity (relativity to a standard) might simply be the way things are. Then the issue becomes which standard is most useful to us given our interests.

Cummins offered his systemic account of function as an analysis of what was going on in at least some areas of physiology. While evolutionary biologists may at times make use of an evolutionary notion of function it seemed clear to Cummins that there was a notion of function in play that didn't explicitly make reference to evolutionary considerations and he attempted to analyse this. Many theorists have found Cummins notion of systemic function to offer a plausible analysis of function talk in physiology in particular. One might thus think that this notion of function might be more relevant to medicine and to psychiatry.

On this notion of function functions are assigned to components at a level with respect to how the components work together to allow some greater process. There has been much controversy over whether Cummins has offered a genuine rival to the evolutionary account of function. One might consider something like an ecosystem, for instance, and then take the systemic approach by attributing functions to components of the ecosystem such as clouds and predators etc. Theorists have argued that there needs to be some non-arbitrary way of fixing the relevant systems. Systems can't be arbitrary mereological fusions, for instance. Thus one way of restricting the range of systems that the systemic notion employs is to use evolution by natural selection. Similarly, one might argue that the evolution by natural selection is

more fundamental than systemic analysis because the systemic analysis only works in virtue of evolution by natural selection operating over the systems.

One line of argument has been that the systemic notion is redundant if we properly understand the resources that the evolutionary view has available to it. Another line of argument would be to maintain that the evolutionary view is redundant if we understand the resources that the systemic view has available to it. Another line would be to maintain that they are both suited to answering different questions. There has been much debate over how they relate and which they are best suited to explaining and whether one of the notions is reducible to the other of these notions.

5.3.3 Homeostasis

Firstly, the notion of a 'set point' or 'set point range' is introduced. The notion is attributed to Cannon. The thought was that in studying cells he noticed that the internal temperature of the cell was fairly invariant to change despite the alterations in external temperature. He noticed that the internal temperature tended to not move much around a fixed point. The average? Was the set point. The degree of variation is the set point range. What happens when the internal temperature varies outside the set point range? Sickness and death. Whatever staves off death. Whatever preserves the characteristics of life. Important to note that thermoregulation isn't a characteristic of life, but perhaps there are subsidiarity functions that are required for those characteristics to be present. Death seems to be an objective measure (while there is trouble characterizing death 'around the edges' we have a fairly intuitive understanding of the notion in the majority of cases). 'Health' or 'well functioning' is harder. There seems to be another level of functions (of which thermoregulation is one). Other things thermoregulate of course and in explaining this a little more we turn to considering the philosophers favourite example of a thermostat.

We are now in the position to see how anatomy, physiology, levels of analysis, characteristics of life, the notion of a set point are inter-related from the per-

spective of anatomy and physiology. While anatomy and physiology come apart they seem to relate to and constrain each other in important ways. While philosophers often think that H2O could have a different chemical constitution and play the same qualitative role it is interesting to note that the question takes on a different problematic aspect when considered from the perspective of the properties of the atoms and how they confer properties on the molecules which in turn confer properties on the observable interactions. They are more tightly bound from the perspective of different levels in science than philosophers have often supposed with variations to lower levels not conferring much in the way of change to variations at higher levels. Philosophers who have taken the sciences seriously seem less inclined to multiple realizability intuitions. Might be that they are missing something of philosophical importance here or might be that philosophers are missing something about what scientists have to show us about the way the levels are related in science (much less multiple realizability for scientific kinds than philosophers have supposed).

Despite Wakefield's taking the evolutionary approach to be the only scientific game in town, the evolutionary view is not without its critics and we are far from a consensus on the correct analysis of function and dysfunction talk in biology, general medicine, or psychiatry. The systemic capacity view provides another way of understanding function talk in biology. The systemic capacity account is different from the evolutionary view in that it makes no essential reference to historical processes and instead attempts to ground functions in component capacities of systems. While the systemic capacity account originally offered by Cummins did not have the resources to account for dysfunction many theorists have thought that the approach could be adapted so as to do so. While evolutionary theorists often consider the main virtue of the evolutionary approach to be that it provides a naturalistic account of dysfunction this has recently come under fire by Davies who defends a modified version of the systemic capacity view. He argues that the evolutionary view is not really an independent theory and that evolutionary functions turn out to be a particular kind of systemic capacity function. He

also argues that neither the evolutionary nor the systemic capacity view have the resources to offer a naturalistic account of dysfunction. If this is correct then it seems that we are left with a significant problem. In particular, if Davies is correct that neither view has the resources to offer an account of biological dysfunction then this would seem to undermine the two-stage views assumption that the role of science in bio-medicine and psychiatry is to discover facts about biological dysfunction. In what follows we will consider both the evolutionary and systemic capacity views of function and then turn to problems that each view has in providing a convincing account of function. We will then consider how each fares with respect to providing an account of dysfunction and end with some thoughts on the role of science in discovering facts about bio-medical and psychiatric disorder.

Thus far most of the discussion has focused on the problem of fixing functions and very little has been said about dysfunction or malfunction. Something clearly needs to be said about dysfunction as there is a third option that any theory must be able to rule out: the problem of distinguishing the functional from the non-functional from the malfunctional. While some theorists might not consider it a criterion of adequacy on a theory of function that the theory have the resources to account for dysfunction (as opposed to dif-function or non-function) given the role that dysfunction is supposed to play in medicine and in psychiatry being able to account for dysfunction must be a condition of adequacy on any account of function that purports to be relevant for general medicine or psychiatry.

There seem to be two ways that we can approach the problem on the evolutionary view. The first is to consider traits where the idea is that traits are binary (all or none) and mutually exclusive. On this view where we have a case of stabilizing directional selection for one trait then we might consider we have the best case of selection against the other traits and thus the other traits are dysfunctional. A similar alternative would be to consider there to be different values within a variant. Similarly, where we have a case of stabilizing directional selection for one of the variant (or values of the variant) then we seem to have the strongest case of selection against the alternative

variants or values. Both of these seem to amount to a similar thing. The main issue that arises here is how we individuate or type traits or variants on traits. We have already considered above the considerable problems that arise when we try and assign function to traits or variants of traits. Similar issues seem to arise when we try and assign dysfunction to traits or variants of traits. But perhaps this whole approach to the issue is misguided. Maybe what we really need is a type and token distinction.

The usual way that evolutionary theorists talk about dysfunction is to distinguish between a type that has a function (or a variant that has a function) and particular tokens of that type or variant that lack the function. On this account the function of the type heart is to pump blood because pumping blood is what resulted in past hearts proliferating such that there are token hearts now. A heart can malfunction by not pumping.

Davies objects to the above characterization maintaining that the evolutionary view does not have the resources to account for malfunctioning instances of types. Davies argument for this is that evolutionary theorists individuate types according to their functions. Since the functions are thought to be necessary and sufficient for membership in the type it is thus impossible for an instance to both be a member of the type (possess the necessary and sufficient condition or function) and yet lack the function and hence dysfunction. Davies thus maintains that instead of a heart malfunctioning all the evolutionary view gives us the resources to say is that the instant that does not pump is not a heart after all and thus it doesn't have the function to pump and thus is isn't malfunctioning or dysfunctioning so much as lacking the function that we wanted to assign.

Davies argument relies on the evolutionary theorist individuating types according to the function that the theorist assigns to the type. Insofar as types possess their function as a matter of necessity he seems correct that a instance of a type cannot malfunction. Despite his maintaining that this is the way that every evolutionary theorist has individuated types it seems that there is another way that seems more licensed by the evolutionary view. He also admits that evolutionary theorists individuate types according to their

aetiology. This seems naturally at home with the evolutionary view and it is important to note that the best theory we have of species membership is aetiological rather than morphological (where morphological might be thought to be more in line with the systemic capacity view).

Davies argument that the evolutionary view does not have the resources to account for dysfunction relies on traits being types according to their function. The problem is basically that IF traits are typed according to their function THEN an instant that fails to exhibit the function fails to be a member of the type and hence we do not have the resources to say that the instant is a malfunctioning or dysfunctioning member of its type. Davies claim seems correct in the sense that if having some function F is both necessary and sufficient for F's being classified as a member of the functional kind K then if F were to lack the necessary and sufficient condition for being a member of kind K then it would simply stop being a member rather than being a dysfunctioning member. By analogy if we consider an instant of gold and we then apply a proton gun and remove one of the protons then the instant isn't a malfunctioning or dysfunctioning or abnormal instance of gold in virtue of having one less proton. Rather, the thing to say would be that the instant that was a member of the kind gold is no longer a member of the kind gold - rather it is a member of kind (whatever has one less proton than gold).

In response to Davies objection one needs simply note that it will not do to individuate kinds functionally rather some other criteria must be used for kind individuation. While Davies writes that all evolutionary accounts appeal to functional kinds there is an ambiguity with respect to what is meant by 'functional kind' here. In particular, by functional kind one could simply mean 'kind with a function' where the conditions for kind membership come apart from the function that is attributed to members of the kind.

Chapter 6

Internal critique of the harmful dysfunction analysis

6.1 Chapter introduction

In this chapter I will look at criticisms of Wakefield's account that I'm going to dub 'internal'. These sorts of criticisms are criticisms that we might have with the particular details of this account. It is in response to objections like these that Wakefield has develop his account as he has forced to become more precise about various things. This chapter runs the risk of listing a lot of objections. I'll do my best for them to be significant and not just do a list of every objection to the view that can be thought up. I think that the cumulative case that can be made against Wakefield can act as something of a primer for the next chapter where I wish to more thoroughly shake at the foundations of the view. I will go on to argue that the notion of malfunction that is relevant is essentially normative or evaluative but that this is the case for medicine as well so psychiatry isn't really worse off.

6.2 Problems with fixing evolutionary dysfunction

Woolfolk and Murphy maintain that there could be mental mechanisms that do not have an evolutionary function and hence they could not malfunction according to Wakefield's analysis. It seems that these mechanisms could result in pathological behaviour, however, and that we would be inclined to regard the individual exhibiting such pathological behaviour as mentally disordered. These arguments are part of their attempt to argue that malfunction (in Wakefield's evolutionary sense) is not necessary for mental disorder.

Murphy and Woolfolk attempt to put pressure on this account of function by maintaining that firstly, mechanisms with no function might produce mental disorder and secondly, mechanisms with a naturally selected function might produce mental disorder where they are not malfunctioning. Wakefield broadens his account of function in light of the first objection. He maintains that there may be mechanisms that were naturally selected for one function whose prevalence in the population now is best explained by their having come to serve another function. One example of this might be the mechanisms that subserve language. These mechanisms presumably evolved by natural selection for some other function but then acquired the function of subserving language.

6.2.1 Alien environments and faulty social learning

While one might be tempted to consider failure of function to be due to a breakdown in an internal mechanism Woolfolk and Murphy consider examples where we would be tempted to say that there is mental disorder and yet where the internal mechanism is not broken. One example of this would be if there was nothing wrong with the internal mechanisms but where there is an input problem which results in pathological behaviour. One way this could happen is when the person finds themselves in an environment that is very alien from the environment that the mechanisms were selected to operate in. Woolfolk and Murphy consider a smoke detector that is positioned too

close to the stove and hence is prompted to give false alarms. In response to this Wakefield maintains that smoke detectors are designed to tolerate some degree of false positives in order to avoid false negatives. He considers the startle response which has this function and maintains that we would not consider someone to be disordered if they lived in an environment in which there were many false alarms unless we had reason to believe there was inner mechanism malfunction. Wakefield maintains that we must be careful to distinguish between problems in living and mental disorder. He maintains that in the smoke detector case there is nothing wrong with the smoke detector as it is behaving in accordance with its design. While the frequency of false alarms might not be valued the behaviour is not due to an inner malfunction and hence the smoke detector is not malfunctioning.

Such a response seems to be in tension with his notion that systems can acquire secondary functions on the basis of present selection forces, however. While ancestral environments were lacking in gunshots and the sound of backfiring cars such features are normally present in certain parts of many large cities. People who have lived in such environments all their lives might be thought to have malfunctioning mechanisms in virtue of acquired functions malfunctioning. The function of the relevant mechanism would seem to be to alert one to danger. To regard misrepresentation of danger as being an inevitable part of the biological function of the mechanism as a consequence of the trade-off between speed and accuracy of responding might well be misguided. One could instead say that false alarms are indeed malfunctions because the function of the alarm is to represent danger and hence false alarms are malfunctions. In the case of the smoke detector it may be begging the question to regard the relevant function as detection of smoke rather than the detection of fire. The way that we specify the relevant function has implications as to whether there is malfunction or not. If the function of the smoke detector was to detect smoke then it wouldn't seem to be malfunctioning when it gives off many false alarms in an alien environment. If the function of the detector was to detect fire, on the other hand, then false alarms would constitute malfunctions. The problem here seems to be

the general problem raised by appealing to functions, the problem of how we determining what the relevant function is in order to assess whether there is malfunction.

In another example he allows there to be malfunction in the absence of a mechanisms being broken is when there are problems with interactions between two mechanisms that are individually functioning but together result in malfunction. The example he offers for this is when a persons brain mechanisms produce a neurotransmitter at the extreme high end of the normal range and produce the neurotransmitters inhibitor at the extreme low end of the normal range. The notion is that each is individually functioning appropriately but their interaction is outside the designed range and has harmful effects. The example of mental disorder that he offers is when someone has self esteem that is low but within the normal range and social anxiety that is high but within the normal range. Together these may result in disorder.

Wakefield maintains that there are indeed some very tricky cases where we arent quite sure whether we want to say there is mental disorder or not. He remarks: there is of course an enormous amount of fuzziness But this fuzziness corresponds to a fuzziness in the concept of mental disorder itself, for as it becomes less clear whether there is a genuine failure of function, it also becomes less clear whether there is a genuine disorder, exactly as the HD analysis would predict 264.

At this stage one might start to wonder what could count against the HD analysis of mental disorder. On the one hand after considering a few cases of the sort we have been considering it does seem intuitively plausible that whether we think someone is mentally disordered or not has something to do with their behaviour being considered abnormal, aberrant, or harmful in some way and also that there is something going wrong with the inner mechanism that are causing their behaviour. Wakefield maintains that evolution by natural selection is the way to cash out what is going wrong with the inner mechanisms though he also seems to make concessions to present day functions as well as evolutionary functions which might be thought to undermine the role he envisaged being played by evolutionary psychology. It might

be profitable at this point to take a step back from Wakefield's account and instead inquire into what Wakefield is intending to do with his account. In particular, I shall focus on the issue of the relationship between his a-priori conceptual analysis of the concept of mental disorder and the nature of mental disorder to be empirically determined by science. (Need to check what he is doing NOW with grief and DSM V).

The DSM has exclusion principles so sometimes even though someone may meet the criteria for a diagnosis the exclusion criteria mean that they do not have that mental disorder. Most of the diagnoses have the exclusion criteria that the symptoms not be due to a general medical condition, and some diagnoses have the exclusion criteria that the symptoms not be considered an understandable, culturally sanctioned, or normal response to environmental events. Depression, for example, has exclusion criteria so that a person meeting the behavioural symptoms in response to the death of a spouse, for example, would be considered to be having a normal response to that event and a response that is not considered to be mentally disordered. If the symptoms continue for 2 months after the event, however, then the individual is considered to be mentally disordered.

Woolfolk and Murphy maintain that in these cases clinical judgement as to whether an individual is mentally disordered or not seems to depend on normative criteria. The depression needs to be considered to be not a normal response. Terms such as expectable, proportionate, appropriate. And normal seem to be value laden 246. Wakefield maintains that appropriate etc is determined by evolutionary ecology.

Another objection that forced Wakefield to clarify his view is the notion that there is one simple behavioural dysfunction to one mechanism malfunction mapping. It could be the case that each diagnostic kind (or category) is due to a certain kind of malfunction in a certain kind of mechanism. This seems to be the upshot of seeing that behaviour is evidence that there is an inner malfunction. It could be the case that there is a malfunction in a mechanism that is designed to bring about the normal form of behaviour. For example, Wakefield sometimes talks about a sadness generating mechanism or a loss

response mechanism. That makes it sound as though depressive disorders result from this mechanism being abnormally activated in circumstances that do not warrant its activation. The trouble with this view is that it seems to commit us a-priori to very strong assumptions about the nature of the structure of mind, assumptions that seem very implausible given what we are learning about the mind. There is also something ad hoc about attributing a mechanism that is responsible for every normal form of behaviour. One shouldn't posit a mechanism when there is no independent justification for believing in it. For example, in the Cotard delusion people say they are dead. One could attribute a mechanism whose proper function is to produce the belief that one is alive. When the mechanism malfunctions it produces the belief that one is not alive. What reason to we have for believing in such a mechanism aside from that this is an easy explanation for the delusion. Without independent evidence to believe in such a mechanism, without the mechanism serving to unify a class of phenomena (not merely the belief that one is alive or not alive) positing such a mechanism is merely ad hoc and raises more questions such as why we have this mechanism, etc.

Woolfolk and Murphy maintain that it would also seem to follow from this that the totality of abnormal behaviour would give us a set of malfunctioning versions of the mechanisms that produce normal behaviours (unless there are additional mechanisms that cant malfunction). This seems to be a *reductio ad absurdum* of the view, however. On the basis of abnormal behaviour x we can infer evolution equipped us with a mechanism specifically designed to produce the normal counterpart of x. That is a very crude interpretation, however. There isnt a specifically designed sadness generator but there would seem to be a bunch of interacting systems or some other undenoted malfunctioning mechanism that produces sadness. 249.

More charitable interpretation is that on the basis of abnormal behaviour of type x we can infer the failure of some pertinent adaptive mechanism. He maintains that it is counter-intuitive to consider that things not due to inner malfunction are mental illnesses. If we found a sociological explanation then we would no longer judge them to be mentally disordered. He emphasises

that whether a person can benefit from treatment is different from whether they are mentally disordered, however. Problems in living may benefit from treatment, but they should not be mentally disordered. Some people maintain that all mental illnesses are problems in living which is to say there are no inner malfunctions. Even the subjectivists are often led to conclude that there aren't any mental illnesses. The best way to make sense of their claims is to grant that there is an a-priori assumption of our concept of mental disorder that there is some kind of inner malfunction in the person. If it turned out that there wasn't an inner malfunction of the person, rather there was a problem with their environment or there was a problem with their behaviour in the absence of malfunction then we would not judge the person to be mentally ill. People who maintain that the exemplars are best explained by sociology or by our judgements in the absence of our believing in inner malfunction often go on to claim that there are no mental illnesses.

6.2.2 Vestigial organs

Vestigial organs are organs that lack an evolutionary function and hence would be incapable of malfunction on Wakefield's account. The typical example of a vestigial organ is the appendix. For our purposes it doesn't really matter whether the appendix is a vestigial organ or not, so long as mental vestigial organs are possible and we conclude that intuitively vestigial organs can result in disorder. It seems plausible that an infected appendix is a medical disorder, for example, and yet if we grant that the appendix lacks a biological function then the appendix could not malfunction on Wakefield's account.

In response to this objection Wakefield maintains that just as functions exist at many levels (intracellular, cellular, intercellular, tissue, organ, and so on), so dysfunctions exist at many levels. If an organ is vestigial, that only shows that there is no function (or dysfunction) at the organ level. But in an inflamed appendix, the functions that are failing are at the sub- organ level. While a vestigial organ like the appendix doesn't have a function thus would not be capable of malfunction the cells that constitute the appendix do have

a function and hence would be capable of malfunction as when they are infected. While we have been considering organs and levels that seem most relevant to general medicine thus far it is now worth turning to how Wakefield envisages this account applying to mental disorders. He maintains that in the case of mental disorders there can be functional mechanisms and hence the possibility of malfunctioning mechanisms at both the neurological and cognitive levels. Thus while Wakefield initially seemed to be considering mental disorders to be the result of neurological malfunction he allows that they can result from malfunction on different levels of analysis. This broadens the notion of malfunction considerably. It seems that Wakefield considers it to be a-priori that mental illness is a result of mental malfunction, but he also considers it to be a-priori that mental illness can be a result of malfunction at different levels of analysis.

If too many levels have functions then malfunctions might come too cheap, however. The clinicians handbook The Diagnostic and Statistical Manual of Mental Disorders (DSM) concurs with Wakefield that dysfunction is necessary for mental disorder but instead of maintaining the dysfunction must be within the individual they consider behavioural dysfunction as well. The DSM asserts that in order to diagnose mental disorder it must currently be considered a manifestation of a behavioural, psychological, or biological dysfunction in the individual DSM xxxi. Wakefield differs from the DSM in his focus on evolution by natural selection fixing the relevant function and he also differs by maintaining that mental disorder is the result of malfunctions in mechanisms that are internal to the individual rather than mere dysfunctions of behaviour.

Wakefield objects to behavioural malfunction even when the behaviours are judged to be harmful on the grounds that it is not in line with our considered intuitions. To motivate our intuitions he describes two different people who meet DSM criteria for reading disorder. The DSM would categorise both people as being mentally disordered on the basis that their behaviour must be disordered in order for them to meet behavioural criteria for reading disorder. Wakefield maintains that if the best theory of one persons inability to read is

that it is the effect of malfunctioning inner mechanisms then we would indeed conclude that the person is mentally disordered. He also maintains that if the best theory of the other persons meeting DSM behavioural criteria was that nobody had ever tried to teach them to read then we would conclude that this person was not mentally disordered, however. Wakefield thus maintains that our intuitions of whether someone is mentally disordered or not is in line with whether we take the best theory of their behaviour to be that it is caused by inner malfunction or not. Wakefield seems correct with respect to our intuitions here though appealing to mechanisms whose function is to enable people to read may be problematic given the way he initially described biological functions as being fixed by the process of evolution by natural selection.

6.2.3 Adaptations, spandrels, and ex-aptations

Another problem would be if some mental disorders turned out to be adaptive in our evolutionary environments. There has been some suggestion that depression, anti-social traits, and histrionic traits might have been adaptive, for example. Whether or not this turns out to be the case it doesn't seem to be ruled out a-priori. Wakefield concludes that if we had reason to believe that those traits were adaptations rather than being the result of malfunction then we would revise our judgement that these traits constitute mental disorder. While Wakefield's account would seem to give us the resources to justify that conclusion Wakefield's account would also seem to give us the resources to justify the alternative conclusion that these disorders do count as malfunctions. One could consider that the relevant functions for disorder are present day functions not evolutionary functions and thus Wakefield's account would seem to support the drawing of either conclusion. Wakefield makes much of the notion of evolutionary function by conceding that present day functions might also be relevant and capable of function he is making a significant alteration to his account. While this may well be a more plausible way to go with respect to the role of the cognitive neurosciences helping to fix present function it does seem to undermine the role that he envisaged

being played by natural selection. Woolfolk and Murphy maintain that there could be mental mechanisms that do not have an evolutionary function and hence they cannot malfunction. They maintain that these mechanisms could result in pathological behaviour, however, and thus malfunction (in Wakefield's evolutionary sense) is not necessary for mental disorder. They offer two examples of mental mechanisms that lack an evolutionary function. The first is that some mechanisms could be like spandrels, and the second is that some mechanisms could be like vestigial organs.

The notion of a spandrel is that there could be something that is a product of evolution but that doesn't itself have an adaptive evolutionary function. If such spandrels exist then there would be mechanisms that cannot themselves malfunction yet they could be capable of producing pathological behaviour. Wakefield responds to this objection by asserting that The HD analysis predicts that, even when spandrels are valued, failed spandrels in themselves are not considered disorders. Rather, the failure of a spandrel implies a disorder when and only when it implies the failure of a naturally selected function. This is a bold conjecture 254

the failure of the very same spandrel will sometimes be considered a disorder and sometimes not, and this will depend entirely on whether or not the spandrels failure is taken to imply a failure of a naturally selected function. The prototypical example here is reading disorders. The ability to read is surely for us a spandrel a side effect of our various mechanisms that was not itself naturally selected but is an invented way of exploiting our selected mechanisms for our own purposes Some people fail to learn to read because they lack educational opportunity, or they are unmotivated, or they are immigrants who do not understand the language of instruction in their school, or for myriad other such reasons that do not appear to involve dysfunction. Other people seem incapable of learning to read even under optimal learning conditions, and 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).

Vestigial organs are thought not to have evolutionary functions either. The

most common example of a vestigial organ in humans is the appendix. In response to this objection Wakefield clarifies his notion of malfunction saying that: just as functions exist at many levels (intracellular, cellular, intercellular, tissue, organ, and so on), so dysfunctions exist at many levels. If an organ is vestigial, that only shows that there is no function (or dysfunction) at the organ level. But in an inflamed appendix, the functions that are failing are at the sub-organ level.

Thus while Wakefield initially seemed to be considering mental disorders to be the result of mental mechanism malfunction he allows that they can result from malfunction on different levels of analysis. This broadens the notion of malfunction considerably. It seems that Wakefield considers it to be a-priori that mental illness is a result of mental malfunction, but he also considers it to be a-priori that mental illness can be a result of malfunction at different levels of analysis.

It seems plausible that the psychological and neurological mechanisms that are involved in enabling us to read evolved for some other function initially and only recently acquired the additional function of subserving language. Such a process is called an ex-aptation and it seems plausible that the mechanisms subserving higher cognitive functions such as language are ex-aptations. Wakefield offers two suggested solutions to this problem. Firstly, he considers that while there is a contingent link between the original function and the acquired function there can only be failure of acquired function when there is failure of original function. This might turn out to be right, but such a response would seem to be making a significant empirical bet. A second response is that failures of acquired functions are appropriately regarded as malfunctions as well as failures of evolutionary functions. If present day functions are included in Wakefield's account of function, however, then he is broadening the notion of function such that evolution by natural selection isn't the only process that is relevant to fix functions.

Current functions wouldn't seem to be the subject matter of evolutionary psychology so much as the subject matter of the cognitive neuro-sciences more generally. This might be thought to be a virtue of Wakefield's recent

modification, though it seems to undermine the role that Wakefield envisaged evolution by natural selection playing in determining the functions and malfunctions that are relevant to our judgements of mental disorder.

The notion of a spandrel is that there could be something that is a product of evolution but that doesn't itself have an adaptive evolutionary function. If such spandrels exist then there would be mechanisms that cannot themselves malfunction yet they could be capable of producing pathological behaviour. Wakefield responds to this objection by asserting that The HD analysis predicts that, even when spandrels are valued, failed spandrels in themselves are not considered disorders. Rather, the failure of a spandrel implies a disorder when and only when it implies the failure of a naturally selected function. This is a bold conjecture 254

the failure of the very same spandrel will sometimes be considered a disorder and sometimes not, and this will depend entirely on whether or not the spandrels failure is taken to imply a failure of a naturally selected function. The prototypical example here is reading disorders. The ability to read is surely for us a spandrel a side effect of our various mechanisms that was not itself naturally selected but is an invented way of exploiting our selected mechanisms for our own purposes Some people fail to learn to read because they lack educational opportunity, or they are unmotivated, or they are immigrants who do not understand the language of instruction in their school, or for myriad other such reasons that do not appear to involve dysfunction. Other people seem incapable of learning to read even under optimal learning conditions, and 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).

(Now I'm having trouble because I would have thought reading was an exaptation rather than a spandrel. If reading was a spandrel then reading disorders couldn't be mental disorders)

Critics have rightly pointed out that M could have been selected for B in our evolutionary past, but be maintained in current populations in virtue of

causing C. One example of this would be that the mechanism that subserve language were selected for one function in our evolutionary past, and yet they seem to have an acquired function of subserving language now so that if language was impaired due to their failure this would be a genuine instance of malfunction. Wakefield responds to this objection by clarifying the role of evolutionary history by natural selection:

> an effect is a function only if it plays a continuing role in explaining the maintenance into the present generation (i.e., continued existence) of the mechanism in the species. A former function that ceased exerting selective pressure long ago is not currently a function because it has no role in explaining current species-typical structure (2003 dysfunction as factual p. 979).

Thus Wakefields revised view thus seems to be that:

- M has the function of causing behaviour B* iff
- 1) M is maintained in the population (by natural selection) in virtue of causing B*

The biological notion of function is thus thought to be fixed by objective facts about the mechanisms and facts about evolution by natural selection.

Another alternative is that despite our negatively valuing psychiatric disorders the population benefits from a certain proportion of individuals having the trait. Models of sociopathy. While there are problems in taking 'cheating and defecting' in a model of game theory as a model of sociopathy in the clinicians sense a-priori it seems possible that psychiatric disorders could actually turn out to be evolutionary adaptations after all.

While Wakefield maintains that this can't be the case his argument is defeasible the same way that if we discovered that water wasn't after all could be allowed on Wakefield's view. He takes himself to be making a 'bold empirical conjecture' but it is unclear why he does this. Doesn't help clinicians (who typically don't have a background in evolutionary modelling) understand whether or not an individual really is mentally disordered or not.

The above account is a fairly rough analysis of how evolutionary processes can be used to fix functions. The devil is in the details, however. In particular, it seems that rather than considering 'pumping blood' to be the relevant individuation of what evolution selected for alternative individuations are possible. In particular, one might plausibly think that merely pumping isn't enough, rather the amount of blood that is ejected from the heart per pump (the ejection fraction) is the appropriate individuation of the trait. One issue that arises from the fact that populations are dynamic over time is that slightly different traits could have been selected for at different periods of time. Adaptations are unlikely to arise from a single mutation event and it is much more likely that mutations result in a number of different variants and evolution works by gradually shaping the proportion of the variants over time. If this is so then it would seem likely that the particular variant that is being selected for at time T is going to be different from the particular variant that is being selected for at time T1. Griffiths maintains that if we are interested in current evolutionary functions then we can view these as being fixed in the most recent period of evolutionary adaptation. While this seems more plausible than the thought that mental or cognitive functions of humans were fixed in the Pleistocene it is important to note that our explanatory interests could be either in the current function of a trait or in the function of a trait at some past point in time or the evolving function of a trait over time.

One might be tempted to think that the clearest case of adaptedness or functionality that we have is when a trait is present in all or almost all members of the population which is to say when the trait has reached fixation in a population. If all or most individuals exhibit the trait then we might be tempted to think that the best explanation for this is that individuals that lacked the trait were at such a disadvantage that they failed to thrive or flourish and ultimately to replicate. There are a number of alternative explanations for fixation, however.

Firstly, it could be the case that rather than selective pressure acting on the trait that we are interested in the trait that we are interested in is an

inevitable by-product or a spandrel of some other trait where the selectional pressure was for the latter trait rather than for the trait that we were interested in explaining.

Secondly it could be that unsystematic causes of mortality or drift has resulted in a trait reaching fixation rather than the forces of evolution by natural selection. Alternative traits that would have gone onto achieve fixation could have been removed from the population by some chance event such as lightening strike. While this seems more plausible in small populations rather than large ones where it would be much less likely that a large number of individuals with the same trait would be systematically removed from the population it could turn out that fixation is the result of drift rather than evolutionary pressures.

What the above considerations show is that we need to be careful in assuming adaptedness or functionality on the basis of observable features. While Paley thought it was just obvious which features were adaptive we have seen that there are several other reasons why such an 'obviously adaptive' trait could achieve fixation in a population. We thus need to be careful in assuming that a trait is adaptive and in building that assumption into our models. The above cases of drift and spandrels were alternative explanations that were reached by way of scientific investigation. Scientific investigation showed us that they were plausible alternative explanations. It thus seems that if we treat the adaptation hypothesis as a hypothesis that can be supported or dis-confirmed we are better off than assuming that a trait is adaptive and attempting to model it with this assumption in play. We will consider the case of the dysfunction assumption and consider the caveats that may apply to that in a later section.

If there is some trait x that is the focus of directional selection (so that the trait is becoming more prevalent over time perhaps reaching fixation) and it is a by-product of x that y is also exhibited then surely observationally it could look as though y was subject to directional selection even though (ex hypothesis) y is not. This is the problem of spandrels.

If there is some trait x that used to have directional selection (was selected for) but then it came to be co-opted and eventually selected for some other effect then there is a problem of what the function is. Griffith's has a way to handle this in saying that functions are fixed in the most recent period of selective adaptation. It still seems that there are going to be blurry edges, however.

Cognitive capacities (if roughly accepted as the focus of psychiatry) or behaviour in a social context (think of the occupational or social criteria for psychiatry) are likely to be exaptations on prior capacities.

In the literature there has been quite a debate about ex-aptations. The idea here is that while certain traits may have been adaptive in environments different from ours (once upon an evolutionary time) the converse might also be the case where certain traits that weren't adaptive once upon an evolutionary time came to be adaptive. The case of ex-aptation is a case where a trait is initially selected for playing some function but where over time it comes to be selected for playing quite another. This issue arises particularly with dealing with mental capacities where it seems unlikely, for instance, that the mechanisms that subserve language were initially selected for subserving language and it seems more plausible that they were co- opted for language from their previous subservience to some prior function.

Griffith's maintains that we should consider current functions to be fixed by the most recent period of evolutionary adaptation. This debate is interesting in that we need to bear in mind that evolution by natural selection works over a number of generations and it is likely that evolutionary pressures are still in operation today. While it might be controversial whether evolutionary pressures are still operating on our mental capacities Griffith's offers us a nice solution to the problem. Insofar as it is unclear whether we are undergoing selective pressures it is unclear whether mental functions are changing or not.

While drift seems most plausible for very small populations and thus may not be directly relevant to the explanation of either the origins or maintenance of psychiatric disorder it is important to consider as it is a case where a variant

can be fixated and may even appear to be 'obviously adapted' and yet it simply was not the best suited variant to the environment. An alternative way in which a variant could appear to be selected for yet where that variant was not actually selected for is if the variant is an inevitable by-product of another variant that was positively selected for. This explanation might seem particularly plausible as it might be that a trade-off in having the fancy kind of mind / brain that we have is that it is susceptible to breaking down in (what scientists and practitioners both might hope are) systematic kinds of ways.

6.2.4 Directional selection, equilibrium, and drift

Adaptations are most clearly adaptive (in an environment or even across a range of environments) when there was directional selection for them such that they are either fast approaching fixation in the population or where the population has fairly much fixated on them. Language might be like this for homo sapiens. The significant majority of human beings acquire language across the majority of environments that they find themself in.

What is important, however, is that while some traits are subject to directional selection such that they are fairly robust other traits might exist in a stable equilibrium or even one in which there are stable cycles of variants. Drift could mimic the effects of selection - or the outcomes of selection could be more the result of drift than of evolution by natural selection.

Intuitively it might seem that the clearest case of evolutionary adaptation or functionality that we have is a case where there used to be a range of variants but where one been strongly positively selected for in the sense that the proportion of individuals exhibiting that variant relative to others has increased and eventually achieved fixation in all or almost all individuals in the population. Unfortunately, while this might seem to be the strongest case of evolutionary adaptedness or functionality alternative explanations still are possible. One possibility is that if we are dealing with a very small population then the population will be particularly susceptible to a variant

being wiped out for unsystematic reasons. A chancy event such as lightening strike could wipe a variant out of the population even though that variant might have gone on to fixate in the population had that event not occurred. Drift could result in a variant being fixated in the population even though intuitively more adaptive variants were present.

Even if we assume that the trait (or variant) came about most significantly by way of evolutionary processes rather than by drift we have seen that there are problems in inferring adaptation or function from the 'obvious adaptedness of a trait to its environment'. The clearest case of functionality or adaptation in an environment in a population is when we see a trait or variant in a trait increase until all or very nearly all members in the population exhibit that trait or variant. In such a case there seems to be a directional selective pressure resulting in the trait taking over the population. The effects of lacking the trait are so deleterious that individuals that lack it fail to thrive. One trait taking over a population is only one of any number of stable equilibrium that evolutionary processes could result in, however. The population could stabilize on two different variations or any number of different variations at different levels of prevalence. In these latter cases it it less clear which variants or traits are functional compared to mal-functional. Need not just be in the minority to be maladaptive or dysfunctional remember.

6.2.5 Units, or levels of selection

The relationship between cognitive and biological (particularly neurological) mechanisms is a matter of much controversy. While most people are materialists in the sense that they maintain that cognitive mechanisms supervene on neurological mechanisms (that one couldn't have a change in cognitive mechanisms without having a change in neurological mechanisms) there has been a great deal of controversy over a more precise account of a mapping.

While the most common view is that genes are the appropriate unit of evolution by natural selection Wakefield says surprisingly little (nothing in fact) about genes. Instead, he focuses on neurology and cognition as the appro-

priate unit for evolution by natural selection. If there are neurological and cognitive evolutionary functions and dysfunctions then it seems that neurological and cognitive mechanisms must be a unit for evolution by natural selection, however. The idea seems to be that genes produce neurological and cognitive mechanisms and those mechanisms interact with the environment such that differences in the neurological and cognitive mechanisms results in different inclusive fitness. The problem that we had with genes not making a difference in the world (and thus needing to talk about their effects in neurological and cognitive mechanisms) seems to recur with respect to taking neurological and cognitive mechanisms to be units of selection, however. In particular, it is only because neurological and cognitive mechanisms result in behavioural outputs that take on a range of values such that they have different inclusive fitnesses that would result in neurological or cognitive mechanisms being seen by natural selection.

It has been noted that genes are the unit of heritability in living creatures. Hearts and lungs and behavioural traits aren't directly passed on to successive generations, but the thought is that the genes that are responsible for their production are. In order to talk of hearts and lungs and behavioural traits having evolutionary functions there would seem to need to be a fairly tight relationship between genes and hearts and lungs and behavioural traits, however. Indeed, if there isn't a fairly tight relationship between genes and some kind of effects that interact with the environment then it is hard to see how the genes have differential effects such that some are more successful than others. Genes are opaque to the environment insofar as they don't result in effects that can result in different inclusive fitness.

It thus seems that evolution by natural selection needs to work on genes since genes are the units of heredity in living people and a unit of heredity is required in order for evolution by natural selection. It also seems that in order for genes to be differentially selected for genes need to be selected *in virtue* of effects that have differential fitness in the world, however. The most direct effects would be behavioural ones. Neurological or cognitive mechanisms would be a step back from that. It isn't uncommon for evolutionary

psychologists to talk of selection for behavioural traits (or the mechanisms that produce them). In particular with respect to affect program responses such as mating signals, attachment behaviours, etc. To say that we can't talk about evolutionary functioning and dysfunctioning behaviour directly runs contrary to the practices of evolutionary theorists.

Another problem that arises for the evolutionary view of function and dysfunction is how much it may or may not presuppose a modular structure to the mind. When we are dealing with fairly low level processes then modular it seems plausible. When we are dealing with comparatively high level processes (such as reading, rationality etc) then the modularity assumption seems less plausible, however. Neurological plasticity seems to be particularly true for higher cognitive functions and people seem to recover much better from losing cortical area than from losing a comparable amount of matter from the brain stem or thalamus, for example.

Mental mechanisms / cognitive mechanisms seem to be (largely) hypothetical constructs that are called in on for the purposes of placing something at the mid-point between genes and behaviour. Genes seem to express in behaviour in virtue of going by the way of developing genetic and behavioural structures. One thing that is nice about a careful enumeration of the evolutionary view is that we can only speak of functioning and dysfunctioning outputs relative to an environment / society. This seems reminiscent of Boorse's view whereby whether one was dysfunctioning or not is partly dependent on how things are with respect to other people. The notion here is that which variants have the best reproductive fitness value is going to be partly determined by which other variants (and how many of them) there are that one is competing against. So long as ones own variant is more adaptive (in an environment) than other variants one has a functioning one. There are problems with respect to where we draw the line between functioning and dysfunctioning variants.

6.2.6 Population and environment relativity

Since evolutionary processes work on populations of individuals it seems that evolutionary models or explanations of psychiatric disorder are similarly going to focus on populations. If we focus on a population at a single point in time then it seems that there are a couple of different ways that we could find the population to be with respect to the trait of interest compared to relevant alternatives[1]. One possibility is that all or almost all individuals in the population have the trait of interest, which is to say that the trait of interest is fixated in the population. Another possibility is that at that time there is either a discrete or continuous range of variants.

Since evolution by natural selection occurs over time we need to add a temporal dimension to our population. Now instead of fixed states of populations at a time with respect to the variant of interest we are dealing with trajectories of the proportion of individuals exhibiting that trait compared to relevant alternatives. One trajectory that the population could be on is that the proportion of individuals that have the trait of interest has increased over time. Conversely, another possibility is that the proportion of individuals that have the trait of interest has decreased over time. Another possibility is that the population is in an equilibrium such that there is either a discrete or continuous range of variants that are present and the proportion of those is constant. Alternatively there could be a more dynamic equilibrium where the proportion of individuals with a particular trait alters over time, but in predictable ways. One last possibility is that the population could alter over time in unsystematic ways.

One feature of the evolutionary view that is often understated is the potential for the view to handle both cross-cultural variation and the resources that the view has to being sensitive to environmental and social considerations more generally. While people often tend to think that if something is 'biological' it is invariant across cultures and inevitable in the development of the organism

[1]We will talk of 'traits' or 'variants of traits' despite the issues to do with individuating the phenomena that were discussed in the previous section.

the evolutionary view turns out to have the resources to be highly sensitive to contingencies of population make-up and socio-cultural environment.

One thing to note is that adaptations aren't adaptive simplicitor, rather they are adaptive for an organism relative to its environmental niche. It doesn't make much sense to ask whether it is better to have lungs or gills - it all depends on the environment that one finds oneself in. Similarly one variant is only 'most adaptive' or even 'least adaptive' relative to other variants that are in the population. One doesn't need to be best - one just needs to outperform alternatives. If you change the environment then you could well make a variant that wasn't terribly successful successful and vice versa.

One often understated feature of the evolutionary view is that effects are only functional relative to an environment. We can view them as being doubly relative. On the one hand they are relative to the other variants exhibited in the population. On the other hand they are relative to other features of the environment. Our environment is very changeable over fairly short periods of time. Consider the technological advances that we have made. Sitting in front of a computer screen all day or living in a high rise apartment or being required to fly all over the world. This environment is in many respects very different from our environment in the plasticine.

While evolutionary dysfunction views of mental disorder often treat evolutionary or bio- logical processes as offering explanations of what is invariant about psychiatric disorders across both time and populations it is important to note that evolutionary models are more sensitive to both other individuals in the population (the organisms social environment) and features of the non-biological environment than has commonly been supposed. Firstly, there are different ways that we can individuate populations. While one might be interested in modelling the prevalence of mental disorder in homo sapiens across the globe one might alternatively be interested in modelling the prevalence of mental disorder in general or of particular kinds of mental disorder in smaller populations. The population of interest could be picked out geographically by broad area (e.g., Europe) by country (e.g., France) by region (e.g., The state of Georgia), or by other demographics either around

the globe or in a more particular region (e.g., in Hispanic males aged between 20 and 30). One might find that prevalence varies depending on region or population and one might be interested in explaining the differences, or one might be interested in explaining the similarities.

Which variant is selected for varies as a function of both what other variants are present in the population and as a function of the environment that the organism finds itself in. While it might seem intuitively tempting to treat cases where almost all have a trait as being cases where selection has selected for that adaptive variant and cases where there is diversity as the population being in some kind of equilibrium neither of these follows. Significant problems in inferring whether a trait is adaptive or maladaptive or differently adaptive from the population dynamics. What more is required in order for a trait to be functional or dysfunctional? Alternative explanations would need to be ruled out.

6.2.7 Epistemic problems: Just so

One of the problems that has been raised with evolutionary explanations is that they seem too easy. We want to explain how some phenomena evolved or came to be and so we make up a story. This doesn't seem so very far away from making up God's intention. Making up models where we assume prior to the construction of the model that the phenomena should be modelled just so (as an adaptation or as a malfunction that is tolerated).

One of the problems that we might have with the dysfunction criterion is the risk of 'just so' stories. This is a problem with evolutionary explanation. This is a problem for evolutionary explanations more generally. Perhaps a feature of bad evolutionary explanations rather than explanations more generally. We start with a characterization and then we want to come up with a story. About how it is a break in a mechanism. Or sometimes not - about how it is a viable strategy. Make up Gods intentions. Make up models where they are selected for or selected against.

Epistemic problems in how we find out about whether a person is suffering

from evolutionary dysfunction or not. This ties into what work evolutionary dysfunction is doing for psychiatry. Might be metaphysically grounded in evolutionary functions (need to consider this) but we also need to consider epistemic issues of how we identify this. Murphy maintains that the dysfunction assumption does for psychiatry what the adaptationist assumption does for biology. He goes on to say that it is useful even though sometimes it is false and other times we don't know whether it is true or false.

The problem here is that we are asked to justify why we regard something to be a disorder. To say that it is because it is due to dysfunction seems to shift things. We then ask how we know it is a dysfunction. Insofar as it is an open question it is unclear how it is helping. If we are assuming that disorders are evolutionary functions then we can't call on their being evolutionary functions to do any explanatory work. An assumption is not a discovery. On the other hand, if it is open for us to discover whether they are functions or dysfunctions then it could turn out that the paradigmatic cases aren't dysfunctions. It could turn out that a number of different situations obtain.

A number of people have catalogued different things that a phenomena could turn out to be other than a function. It could be a spandrel (a by-product of something that was selected for). It could have arisen due to chance or accident or mere historical contingency.

6.2.8 Implications for evolutionary accounts

We need to plug in actual details to constrain the scope of 'how possibly's' - there might well be such a significant epistemic problem in figuring out what is going on that evolutionary theorizing about psychiatric disorder isn't a useful cognitive heuristic. In particular, viewing psychiatric disorders as evolutionary dysfunctions seems fraught - we have already seen that there are any number of plausible things going on (and all of these were described very abstractly indeed so we haven't considered the huge range of alternative hypotheses within each broad approach). Dysfunction is one. Only one.

One dimension of difference is temporality. Historical accounts appeal to evolutionary history for fixing function. Propensity accounts attempt to capture the forward looking teleological aspect of function talk by appealing to propensity or dispositions. Theorists could differ with respect to how long ago functions were fixed. Griffith's appeals to the most recent period of evolutionary adaptation. There are different ways that we can fix function depending on the point at which the functions were fixed. One important consideration is that things vary fairly gradually in the sense of varying over generations. Dennett on natural kinds of biological organism. Worry about when precisely speciation occurred.

Hopefully it is becoming apparent that the issue of functions and dysfunctions brings up concerns about how we are to characterize the types or kinds of things or the reference class that we are attributing functions and dysfunctions to. What should we take from the above considerations? One lesson is that we can't straight- forwardly infer that a trait or variant on a trait is adaptive on the basis that it 'just seems obvious' to us that it is. We have seen that 'obviously adaptive' variants can come about by way of spandrels or drift and that such traits are only loosely considered to be adaptive in the evolutionary sense.

One might think that the 'dysfunction assumption does for psychiatry what the adaptationist assumption does for evolutionary biology - which is to say that sometimes the assumption is false and sometimes we don't know whether it is true or false but that need not impugn scientific inquiry'. On the other hand one must also be careful not to mistake an assumption for a discovery. If scientists discover that schizophrenia, bi-polar, depression, and psychosis are evolutionary malfunctions or dysfunctions then this seems to be a significant discovery. We might well begin with the assumption that they are dysfunctional but whether they are or are not in the evolutionary sense seems to be a different matter.

How about the controversial cases where it is unclear to us whether it is an evolutionary dysfunction or function or quite what is going on? Can we then use evolution to tell us that a condition is a mental disorder rather than say

a problem in living or whatever? We would need to be fairly robustly sure that the evolutionary view gives us the intuitive verdict before we are willing to extend it to cases where our intuitions are borderline. We are a long way away from that yet.

6.3 Inner, mental mechanism

One of the virtues (appeals) of Wakefield's INNER malfunction assumption is that it promises to differentiate between disorder and problems in living. This distinction does indeed seem to be important and it would be nice if we could have an account of it but I have my reservations about Wakefield capturing it because the notion of function is indeed problematic.

Murphy is correct to observe that Wakefield does indeed seem to be capturing an intuition that we have that CAUSE of the behavioural symptoms is important. Some kinds of causes (play acting, attempt to get gain, drug induction etc) seem to be exclusion criteria for a person having a certain disorder even if they display the behavioural symptoms. Murphy is also correct to note that we don't need to assume that the relevant cause is a malfunction in order to capture the intuition that a certain kind of cause is important. Precisely what more we say about which causes are relevant and which are not will depend on how things turn out. The notion seems to be that those who are play acting (etc) are importantly different from the other cases. Maybe that they are not the typical cases (if all instances were play acting would we conclude that there is no such thing as mental illness or would we conclude that the nature of mental illness was that it was a play act? Depends whether we take it to be more revisable that mental illness is due to non-intentional causes or whether we take it to be more revisable that those people are mentally ill).

In maintaining that the dysfunctioning behaviour must be due to the malfunctioning mental mechanism Wakefield makes aetiology a constituent. Consider sunburn. An essential part of something being sunburn is that it be caused by the sun. A burn that is not caused by the sun is not a sunburn

but it may be a different kind of a burn.

Wakefield maintains that we need the aetiology to distinguish between behaviour that is the result of a disorder vs behaviour that might well be problematic but isn't disordered. The example he offers is of a persons inability to read. It is only the inability to read that arises from a dysfunctioning mechanism (e.g., and not an impoverished environment) that is regarded to be disordered. This the main work that Wakefield wants to do with this distinction.

The DSM doesn't draw such a distinction. Kraeplin (as we saw in chapter one) thought that the symptom picture would be different. Different kinds of burns might be identifiable from their morphology. Wakefield wants to utilize a distinction between personal and sub-personal description / explanation. This is a distinction that is drawn from cognitive science. The basic idea is that notions such as rationality, action, reasons, desires etc are person-level notions. When we are describing or explaining the actions of agents we can appeal to such notions. These are normative, evaluative, or agentive notions.

On the other hand, sub-personal explanation makes no reference to these person-level notions. Instead of reasons and norms we have causes and mechanisms. Wakefield maintains that the notion of 'harm' is a person-level notion that applies to persons (more on this later) whereas the notion of dysfunction is a sub-personal one, it applies to mechanisms.

Wakefield maintains that the inner mechanisms are what is functioning or dysfunctioning. His argument for this comes from our intuitions about whether disorder is present or absent seeming to depend on whether they are caused by a dysfunction or not. The thought is that no matter how messed up certain behaviours are those behaviours are only (intuitively) regarded as disordered when we think they are due to an inner dysfunction. Those same behaviours in the absence of inner dysfunction would still be problematic - of course. People might well benefit for help or treatment for those behaviours. The thought is that these people aren't mentally disordered however as inner dysfunction is required for mental disorder.

6.3.1 Mental vs non-mental

The first thing to note is that while Wakefield is interested in offering an account of mental disorder primarily he is attempting to offer an analysis of the bio-medical notion of disorder more generally. In keeping with the greater literature he doesn't have a great deal to say about what is distinctively mental about mental disorder.

6.4 Conceptual intuitions and science

The first two premisses of Wakefield's argument are supposed to be analytic. Wakefield maintains that we have some pre-theoretic grip on the notion of 'dysfunction' and that it is part of our concept of disorder that disorders are harmful dysfunctions (in some pre-theoretic sense of dysfunction). He maintains that pre-theoretically or a-priori there are many processes that could fix functions and dysfunctions. One is the intentions of a creator God, for example.

The second premiss is also meant to be a-priori or analytic. This premiss reveals that Wakefield thinks that it is a-priori or analytic that dysfunctions are natural kind terms. He draws an explicit analogy between his treatment of the term 'dysfunction' and the causal- historical analysis of the terms 'gold' and 'water' offered by a Kripke-Putnam style analysis.

The third premiss is meant to be a-posteriori or empirical. It asserts that scientists *have discovered* that the relevant process for fixing biological functions and dysfunctions is evolution by natural selection. The thought is that since scientists have discovered the relevant process for fixing biological functions and dysfunctions around here is evolution by natural selection that functions and dysfunctions are essentially fixed by evolution by natural selection.

It is important to distinguish two distinct views. According to the first it turns out that the nature of mental disorder is that it is a certain kind of normativist violation. According to the second it turns out that our judgements of whether a person is disordered are tracking whether they are violating

certain kinds of norms. There could be epistemological and metaphysical versions of both these views in the sense that the nature is fixed by our judgements on the one hand, and the reality on the other hand of norm violation. Or that our judgements of whether a person is disordered are only correct or justified when they do track norm violation.

It is important to distinguish between metaphysical normativism according to which the nature of disorder is that it is a normative violation and epistemological normativism according to which our judgements about disorder track whether it is a normative violation.

Often normativists maintain that our current judgements that individuals are disordered are nothing more than a reflection of prevailing norms of the society within which the judgement is made. If we think that people are correct to regard a person to be violating norms when they are in fact violating the prevailing norms and we think that psychiatry is concerned with treating such norm violations then it turns out that the view does not have the resources to allow us to criticize past psychiatric practices. Homosexuals were in fact violating the norms of their society, for instance, and as such the view doesn't seem to provide us with the resources to say that homosexuality should have been taken out of the DSM at the time when homosexuals were in fact violating social norms. Similarly, slaves who attempted to escape their masters did seem to be violating the norms of their society as were the political dissenters in Russia.

There are two different ways we can go here. Firstly, we can consider that instead of prevailing norms setting the relevant standard, there are some idealized norms and when individuals violate those idealized norms that is what is relevant (necessary) for our being appropriately justified in regarding an individual to be disordered. Secondly (as I shall consider in the next section) we can maintain that not just any kind of social and / or moral norm violation is relevant for determining who is and who is not disordered, rather the norm violation must be of a certain kind. Both of these moves is to make a distinction between our actual judgement that someone is violating the norms relevant for fixing disorder and facts that are at least potentially

independent of the judger that fix whether the judger is correct in judging either the presence or absence of norm violation.

The problem of relativism is something that comes up fairly standardly in ethical theory.

Ethical theories are typically concerned with hitting upon a theory of social and / or moral norms that transcend the norms actually embraced by any particular society or culture, and thus attempt to promote tolerance of difference while still allowing us to critique certain actual and possible moral or social practices as unwarranted. We typically do want to allow that some societies and / or cultures have social and / or moral norms that are unwarranted or illegitimate or criticizable in some way. Utilitarianism, for example, allows that a particular act that maximizes utility in one situation or culture may be quite different from the particular act that maximizes utility in another situation or culture. It seems to me that a similar move is available to the normative view of mental disorder. Instead of maintaining that mental disorders are in violation of a particular societies or cultures social and / or moral norms they are able to make a comparable move and maintain that mental disorders are in violation of the social and / or moral norms that would be held by a sufficiently enlightened or otherwise idealized society. Whether there will be one unique view remains to be seen.

The above lines make assumptions about the nature of science and norms. There are a number of different aspects or points that are being run together. For instance, it is assumed that science is objective whereas values / norms are subjective. Science is universal whereas values / norms exhibit cross cultural variation. We might tease some of this out, however. In order to get clearer on what is required for an account / what is going on with mental disorder.

Disorder: A State or a Process?

Wakefield seems wedded to the idea that a disease is a state rather than a process. He talks about malfunctioning mechanisms where the mechanisms are regarded as physical structures within individuals. When the malfunction

results in harm then the individual is disordered Alternatively, it might be that disease is better conceptualised as a process. Kraeplin was thought to be going against the grain here when he maintained that we should classify on the basis of etiology, symptoms, and course of symptoms over time.

Chapter 7

External critique of the harmful dysfunction analysis

Chapter introduction

The main critique of the two-stage view has come from those who maintain that function and dysfunction are normative notions that cannot be analysed non-normatively. They maintain that what is going on when we judge an individual to be disordered is that we start out by normatively assessing their behaviour and then we cast about for something to medicalize in order to justify our normative assessment. Murphy notes that science often deals in normative or idealized processes such as the normal development of an eye, a star, or an ecosystem. It isn't thought to be problematic for other sciences that they employ these notions of norms and hence is isn't problematic for the sciences of the mind / brain if they employ them either.

7.1 Different notions of function

Earlier we considered four broadly different approaches to analysing function talk - the teleological, the bio-statistical, the evolutionary, and the systemic. I don't claim that this list is exhaustive but it does illustrate that there are

a number of broadly different approaches that differ with respect to whether they consider certain phenomena to be functions or malfunctions. We might want to divide these up still further, and consider that there are different approaches that are roughly all evolutionary (for example) but different with respect to how they precisely spell out how functions and malfunctions are fixed and thus again they differ with respect to whether a phenomenon is functional or malfunctional. One way of understanding this is for them to be different theories of the same thing (functions and malfunctions) another way of understanding this is for them to be picking out different phenomenon (this ties into direct reference vs descriptivism which is why I worry about having the conceptual analysis chapter so late).

The different notions of function seem to deliver different verdicts (or at least there is much work to be done in showing that they do not). I think the burden of proof is on the theorist who maintains that there is a single notion here. I think the real issue is in which concept we should adopt given what it is that we want to do with the concept. Here normative issues are going to come to bear - this issue seems to be important because it seems analytic that people who are disordered would be better off if they weren't disordered and that we have some (defeasible) obligation to assist people who are disordered etc. Similarly, I'm not sure that it is so helpful to carve off the scientific notion/s from the ethical notion/s. I think that there might be some benefit to trying to capture an integrated notion otherwise it isn't so clear why we should care about the scientific notion or what we can do to help in the normative one. I think that Wakefield's attempted distinction between non-normative dysfunction and normative harm is problematic and that there are problems for two-stage views that maintain that one can be carved from the other. I think it might be more profitable to consider how they are related.

7.2 Function as a relation

Each of the broadly different approaches to function and dysfunction seem to share a common structure. Functions and dysfunctions are thought to be *relational* properties. Part of the relation is thought to be fixed by the state of the mind-independent or objective world[1]. The other part of the relation is something that seems to vary across the accounts. I'm not sure what to call this part of the relata. I'll call it an adoption of a *standard of evaluation* for want of a better term. The standard of evaluation can be captured (roughly) as follows:

- **Teleological** The intentions of an agent
- **Bio-statistical** Statistical average of a reference class
- **Evolutionary** What past tokens of a type did that resulted in the proliferation of the tokens of the type
- **Systemic** Whatever it is that we begin by wanting to explain (e.g., the circulatory system).

The standards that I have listed capture something of the broadly different approaches to fixing function and dysfunction. It can also be seen that we can capture variation within the broadly different approaches by making the standard of evaluation more precise, however. So, for example, on the teleological view we might identify the intentions of a particular agent as being relevant (that of a creator or designer, for example), on the bio-statistical view we might identify the reference class and the amount of variation from the average that is tolerated, and so on.

Once we have fixed on a relevant standard then we can translate function and dysfunction talk into talk of objective properties and processes in the following way: 'The function of the heart is to pump blood':

[1]I need a caveat here when it comes to intentions as intentions are paradigmatically mind-dependent. The thought, however, is that what an agent intends is of course dependent on the mind of that agent. What an agent intends is a fact about the mind-independent world, however, in the sense that whether or not an agent intends x is objective in the sense that it is something that we can simply be wrong about.

- **Teleological** Pumping blood is what God intended for hearts to do
- **Bio-statistical** Hearts typically pump blood
- **Evolutionary** Past hearts that pumped had greater fitness relative to non-pumping hearts
- **Systemic** The heart contributes to the circulatory system by pumping blood.

Then we can simply look to the world to see whether the claims are true (and whether the function of the heart - according to the relevant standard - is indeed to pump blood).

What it is really important to note about this view is that if we have the above mentioned facts in the *absence of a standard of evaluation* then we can't tell whether the function of the heart is to pump blood or not. The world under-determines the functions and dysfunctions and what it it that is needed in addition is a standard of evaluation.

The issue now becomes: What is the relevant standard of evaluation for bio-medicine and for psychiatry? While science is clearly needed in order for us to discover facts about the world we have seen that no amount of facts about the world will determine whether a trait is functional or dysfunctional until we have a standard of evaluation. We don't seem to need to commit to a standard of evaluation in order for science to discover facts about the world, however. We don't seem to need to commit to a standard of evaluation in order for science to discover causal processes relevant to producing behaviour that is of interest to us. We don't seem to need to commit to a standard of evaluation in order for science to discover how to intervene on those causal processes relevant to altering behaviour that is of interest to us. (I think this is a major thing. If we see that science can proceed with discovering causal processes and developing interventions for behaviours we are interested in then we simply don't need the dysfunction assumption except as some kind of normative notion).

I want to suggest that rather than functions and malfunctions being simple

properties, functions and malfunctions are best conceptualised as relational properties. The notion that function and malfunction is a relational property rather than a simple property is that it is illegitimate to ask what is the function of x simplicitor without identifying a relevant standard whereby one can measure the relevant effects or performance against that standard. As such, things can be more or less functional and more or less dysfunctional according to how close they come to or how far they deviate from the standard. This idea might seem counter-intuitive so I shall provide a couple of examples in order to make the point clear.

The statistical notion of function (and malfunction) sets the standard (or function) as that of statistical average. The functional effects are those effects that are statistically average or statistically normal and one can measure how deviant or malfunctional something is by measuring how many standard deviations something is from the average or the mean. Insofar as it makes sense to talk of statistical deviance or malfunctional the facts about how deviant or malfunctional something is can be read off from the facts about how far it lies from the average or the mean together with the identification of the relevant standard as one of statistical normality.

The evolutionary notion of function (and malfunction) sets the standard (or function) as the effects that resulted in relative fitness with respect to survival and reproduction. If one knows that a trait resulted in relative fitness with respect to the standard of survival and reproduction then one knows that the trait is an evolutionary function. If one knows that a trait resulted in diminished relative fitness with respect to the standard of survival and reproduction then one knows that the trait is malfunctioning or maladaptive. Once the standard of survival and reproduction has been set and once one knows whether the relevant effect contributed towards or away from the relevant standard then one knows whether the relevant effect is a function or a malfunction.

The teleological notion of function (and malfunction) sets the standard (or function) as the effects that contribute towards or away from some identified goal. Once we know what the goal is and we know whether the effects

contribute towards or away from the goal then we can simply read off whether the effect is functional or malfunctional. Similarly if we identify the relevant standard as one of propensity to survive and reproduce or if we identify the relevant standard as fulfilment of some Cummins output such as the pumping of blood.

If one operationalises rationality as following the laws of first order logic then it seems that people are irrational in that they commit the conjunction fallacy. If one operationalises rationality as adopting useful strategies for getting by in the majority of real world encounters, however, then it follows that people are rational to track relevance of information. Can we say that the people who commit the conjunction fallacy / track relevance are rational or irrational simplicitor? To argue that they are or aren't would be to argue that one notion of rationality is better than another or to argue that we should adopt one notion of rationality over the other.

It was once thought that xx (or female) was a malfunctioning sex in the sense that females were malfunctioning males. This was a particularly common view in medicine where anatomy texts took masculine anatomy and symptoms to be standard. We also find this in Freud's theory of human development where females were thought to have the additional problem of coming to see that they were in fact malfunctioning males. We don't regard females are malfunctioning males anymore, however. Instead we regard xx (female) and xy (male) to be dif-functions in the sense that they are two different (though equally functioning) ways of being. If someone is xxy or xxx, however, then we typically regard them to be malfunctioning. There is a current movement for these alternative genotypes to be regarded as dif-functions rather than dysfunctions, however. One biologist has argued that there are 5 sexes rather than 2 and it is hard to see how purely causal facts can settle these issues one way or another. The relevant issue is how purely causal historical processes determine whether something is a function or a dysfunction or a dif-function.

It might be objected that some standards are more relevant or more sensible or more worthwhile than others. That seems right, but here I would say that

the reason why some standards seem more relevant, sensible, or worthwhile is determined by our explanatory interests. Of course now it might be objected that there could be a whole range of explanatory interests that a variety of people have had, do have, and will have. It still seems that some of those interests are more relevant, sensible, or worthwhile than others. Here I can only re-iterate the problem of more relevant, sensible, or worthwhile to whom. The evaluation of our explanatory interests seems to be firmly within the scope of value theory; of figuring out what people do value and recommending what people should value. It would seem that our values determine what explanatory projects are worthwhile and our determining our explanatory project involves identifying a relevant standard. Once we plug the empirical facts in and see how much effects contribute towards or away from the relevant standard then we have fixed whether the effects are functional or malfunctional in the relevant sense. What further information does function and malfunction talk provide for us? I don't think that it provides any further information. Whether function and malfunction can be naturalised depends on how much values can be. This point (the point of fixing functions and dysfunctions) is so persistently important that we need some new terms to capture it. If dysfunction means to fail to function or to function badly, why not diffunction for functioning merely differently? The term superfunctioning is clumsy, but is all I can suggest for the equally important and relatively neglected (compared with dysfunction) notion of superior functioning, or functioning particularly well. footnote page 82. 2000 PPP

The naturalization project, then, is driven by the belief that somewhere deep down in the naturalization cascade there is a value-free foundation, an heuristic holy grail, on which biology, and in turn the theoretical cores of medicine and psychiatry can, in principle, be built up as mature scientific disciplines. 83

Most naturalizers recognise that values have to be brought into the naturalization cascade at some point Their event horizons for values, as I put it in the first section of this paper, differ: Boorses event horizon, in his ear-

lier work at any rate (1975), was between disease and illness; Wakefields is between dysfunction and disorder; Thorntons is between dysfunction and disease. But philosophical naturalizers all bring values in somewhere along the line. 83

it is common ground, among naturalizers of disease, as it is between naturalizers and their opponents, that some parts of the naturalization cascade are more overtly value laden than others: psychiatry is more overtly value-laden than more high-tech areas of medicine, medicine more than biology, and biology more than the natural sciences, such as physics.

7.3 Will the real notion of function and dysfunction please stand up?

Peter Godfrey Smith (?) has written that philosophy should never try and join together that which science has cast asunder. His point is that if there seems to be evidence from science that biologists sometimes use the notion of function differently then maybe there really are different notions of function kicking around. It certainly seems that there are different notions. As Murphy has pointed out it certainly seems to make sense to say that 'Harvey understood the function of the heart centuries before Darwin' and that his understanding that the heart worked to pump the blood around the circulatory circuit had little to do with the theory of evolution by natural selection.

Dennett has pointed out something similar with respect to the notion of a 'voice'. xxxxxxx. What is it that each notion of a voice has in common - what is a voice? It can be really very hard to say. Whatever they share in common must be so abstract... So elusive... And yet we seem to understand what is being said well enough.

I think that we need to reconsider a lot of what went on in Chapters one and two in order to get a better handle on what is going on (or a useful way to proceed). Chapter one illustrated that what we seem to care about it the role of science vs the role of norms / values for mental disorder. We would like

to know what individuals are disordered and what kinds of disorders there may be. We would like to know what sort of intervention might be best for putting things right.

7.4 Thick concept

Psychiatry's status as a specialist branch within medicine reflects an adoption of the medical model of mental disorder. While there is controversy over precisely what the medical model is committed too, and in particular, whether it is committed to biological reductionism, it is fairly uncontroversial that it is committed to a single notion of disease, disorder, dysfunction, illness, malady, dysfunction, etc that is shared by the different branches of medicine, including psychiatry. The notion is that just as there is cardiovascular disorder and neurological disorder, mental disorder is yet another kind of disorder in the bio-medical sense.

There has been a great deal of debate about the concept and nature of mental disorder, disease, illness, malady, abnormality, pathology etc. One way of getting to the heart of the debate is to distinguish between one-stage and two-stage views, or to distinguish between two different positions that one can take on what Fulford dubs the values in values out debate. According to the two-stage view (the values out view) the concept of mental disorder is a thick, or evaluative concept. While the notion of mental disorder is thick, it is thought to be a redeeming feature that the evaluative aspect can be carved off from the non-evaluative aspect, and that the science of mental disorder can progress by focusing in on the non-evaluative aspect. The non-evaluative aspect is typically characterised as being the notion of a behavioural and / or psychological and / or biological dysfunction within the individual.

The two-stage (values out) view can be contrasted with the one-stage (values in) view where, according to the one stage view, the concept of mental disorder is irreducibly evaluative in a way that threatens the prospects for a science of mental disorder. The notion is typically that while the notion of dysfunction that is employed in other branches of medicine is enough to

ground medicine firmly in the biological sciences the notion of disorder employed in psychiatry is metaphoric and / or somehow illegitimate in the sense of being a category error (see, for example, T. Szasz). The values in or one-stage view is typically advocated by some anti-psychiatrists who maintain that there is little more to mental disorder than that certain individuals violation of certain kinds of social and / or moral norms. As such, it is thought that mental disorder is distinctly different from other kinds of disorder and a science of the discipline would consist in investigating the social and moral norms and violation of those norms rather than investigating behavioural and / or psychological and / or biological facts about dysfunction.

7.5 Assumption vs discovery

Murphy maintains that the malfunction assumption does for psychiatry what the adaptationist assumption does for biology. Which is to say that sometimes the assumption is false and other times we don't know whether it it true or false but this does not impugn inquiry.

We need to consider what work malfunction is supposed to be doing, however. On a two-stage account and on Wakefield's HD analysis malfunction is supposed to capture what it is that grounds psychiatry in medicine and in the natural sciences of biology. It is supposed to be what prevents psychiatry being like law.

On Wakefield's account malfunction is necessary for mental disorder and thus it makes a significant difference whether or not the malfunction assumption is true or false of any given phenomenon. It could be the difference as to whether it is a disorder or a social problem (e.g., like not being able to read that is not due to a malfunctioning mechanism).

Malfunction was called in to justify (or to falsify) our judgements. It was supposed to prevent abuses of psychiatry. E.g., it was wrong for us to have considered political dissentors to be disordered precisely because their political dissent was not caused by malfunctioning mechanisms. Similarly we were

wrong (the story goes) to have considered homosexuality to be a disorder because again, it is not caused by a malfunctioning mechanism.

An assumption is not a discovery, however. If we start out assuming that mental disorders are the result of malfunction and we build that assumption into our model then it is not something that is (or that can be) discovered by science.

One thing we might conclude is so much the worse for psychiatry and medicine - they are not suitably grounded in the sciences after all. Another thing we might conclude is so much the worse for the dysfunction criterion as an account of the nature of the scientific contribution to the phenomenon. I'm particularly interested in this latter line.

Science is often characterized as making a number of assumptions. Cognitive psychology, for example, is often characterized as making assumptions about modularity of mental com- ponents / mechanisms and assumptions about mental processing being feedforward. Similarly, there seems to be a 'dysfunction assumption' in bio-medicine and in psychiatry.

The assumptions that are thought to be made are not something that is typically believed unquestioningly, however. Researchers working within the paradigm often undertake research that attempts to question the assumptions, show them to be false, or show what the science might look like in the absence of the assumption. This is the case for bio-medicine and for psychiatry when researchers attempt to model how mental disorders might be adaptive or differently functioning, for example.

One concern is that dysfunction or function seems to be something that is built into the model in the first place rather than being something that is fairly straightforwardly discovered by models. If we attempt to model a mental disorder as an adaptive strategy then we will get a different result from if we attempt to model a mental disorder as a maladaptive strategy. It is unclear that science discovers function or malfunction so much as producing results that are consistent with the constraints of the model where function or malfunction are considered one such constraint.

While medical research might need to appeal to malfunction or dysfunction in order to obtain research funding it is unclear to me whether science needs either the malfunction or dysfunction assumption in order to model causal processes. I don't see why they don't just stick to modelling causal processes except insofar as they want that health research funding.

7.6 Normativity

Wakefield intends for 'harm' to stand in as a place holder for the normative aspect of bio-medical disorder. We have already considered one stage normativist views and some of the objections to that style of account. I think that one-stage views could be developed to respond to the criticisms that have been levelled against them.

One objection to a one-stage normative view is that more must be said about the relevant kind of social and / or moral norm violation. We regard some norm violators as odd, we regard other norm violators as evil or morally bad, and we regard other norm violators as mentally ill. What is the difference in the norms that determine which of those applies to the person? It seems clear that not just any kind of social and / or moral norm violation is relevant for judgements of mental illness, so more must be said about the nature of the norm violation that is relevant for psychiatry.

Whether two-stage theorists are also required to get clearer on the relevant notion of norms that are in play or not depends on the robustness of the account of dysfunction that is offered. If it turns out that people who perform actions that are morally deviant or just plain odd aren't subject to evolutionary dysfunction whereas the people who are mentally disordered are then we may not need the norms that are in play to be different in order for theories of mental disorder to respect that intuitive distinction. If it turns out that the dysfunction (or other non-normative criterion) is unable to differentiate the morally bad from the odd from the psychiatric then the difference between them would have to be cashed out as a difference in either the kinds of norms that are being violated or in the degree to which they are violated

in the different cases.

Another objection to a one-stage normative view is that it doesn't give us the resources to critique past practices as wrong or misguided. The argument for that runs as follows: If there is little more to mental disorder than social and / or moral norm violation then it must be the case that whether an individual is disordered or not is a matter of whether they break the norms of their society. Social norms vary across societies, however. Some societies have social norms such that political protest or homosexuality is in violation of the norms of those society. The normative view of mental disorder would thus have it that political dissenters or homosexuals living in those societies would be mentally disordered insofar as they were in violation of the norms of their societies. This is a conclusion that we do not want to accept. We should therefore reject the premiss that mental disorders are solely a matter of norm violation.

This objection ignores much work that has been done from within ethical theory, however. We want to criticize the practices of people living within past societies (people who kept slaves for instance) and ethical theory has the internal resources to enable us to do so. One thing to do is to not identify the moral (or the psychiatric) by reference to social and or moral norms that are endorsed by one or more present society. Instead the thing to do is to identify the moral (or the psychiatric) by reference to some idealized ethical or moral standard. The notion would thus be that mental disorders aren't fixed by the person being in violation of the norms of any current society, but rather mental disorders are fixed by the person being in violation of some idealized norms. It would thus seem that there are prospects for developing a one-stage normative view such that it can respond to the objection leveled against it that it cannot allow us to critique past psychiatric practices that are (clearly, or clearly in relation to our moral or social normative viewpoint) not psychiatric problems (or social or moral violations) at all.

It seems that 'harm' could be (at least) partly objective in the sense that there could be a fact about whether a person was harmed or not that could be quite different from what you or I believe about it (so mind-independence

at least). For example, slaves seemed to be harmed even though there was a time where people thought they weren't. The GAF scale as an attempt to make the values in psychiatry explicit (the notion of what an 'optimally' (100) functioning person would be.

The problem with Wakefield building in malfunction a-priori is that firstly we identify the paradigmatic cases then we discover their inner causes and finally we attach the label malfunction to the cause we have discovered. If it is a-priori that mental disorder is caused by inner malfunction and scientists attempt to find the inner cause of mental disorder then when the scientists proclaim that the cause of depression is x Wakefield thinks that x is a malfunction by definition. Hence it isn't scientifically interesting whether someone is malfunctioning or not. Malfunctions aren't a scientific matter. We decide on a-priori grounds that someone is malfunctioning and science merely investigates the causes.

One might think so much the worse for building in the malfunction assumption a-priori and attempting to make it unrevisable. After his fairly devastating critique Murphy goes on to maintain that the malfunction assumption is a useful assumption to guide scientific inquiry. I fail to see why he backs down now, however. If the above process is how things are working then it would be the case that to say their behaviours are caused because they are malfunctioning would not add anything to the explanation. Why are they mentally ill? Because they behave that way because they have this inner state. Inner state in absence of behaviour a disorder? Dispositions? Intuitions are funny here when we go modal.

Is mental disorder like water or like watery stuff? We construct idealised models maybe not statistically normal as we can model statistically abnormal processes too. For example, we can describe the progression of cancer Cancer is often described as a malfunction but we can treat it as a function (the function of cells is arguably to replicate, after all) and on this kind of idealised model of cancer (statistically idealised) there could be we do seem to want to revert to normal here like Murphys notion of the stars. I don't have an issue with trying to identify common causal mechanisms but we

would seem to need independent to regard the system as malfunctioning We do seem to assume malfunction from behaviours In particular behaviours that result in harm DSM does seem to presuppose the function of a person (must interfere with functioning before particular dx can be given). While some of the arguments against objective malfunction seem to be in error (by considering Aristotelian function instead of Evolutionary function) Cumins function seems to raise issues of its own.

Assumptions of the Orthodox Program of Conceptual Analysis. Failure of a normal mind to function as it should is a necessary condition of mental illness - Normal functioning is a product of natural selection of mental components - Components of mind depend on functionally distinguishable components of brain - Normal psychology is common to all homo sapiens. The first assumption of mental illness as dysfunction isn't empirical but if this is right then there would be no mental illness if there aren't mental functions. This follows from common-sense and concept of disease in western medicine and in conceptual analysis. But if the mind isn't organised like that then there wont be mental illness. If schizophrenia wasn't caused by malfunction in identifiable mental organs there wouldn't be mental disorders. Wakefield thinks it is obvious from surface features (bollocks). Seems to be equivalent to saying that whether someone's psychology is functioning normally is a-priori. But surely this is an empirical matter. We should reject this. Isn't needed to vindicate the common sense assumption that aetiology (causal explanation) is important.

Three problems for conceptual analysis. Relation of science to common-sense about mental disorder - nature of the mental - nature of psychological function and malfunction. His view: objectivism does for psychiatry what adaptationism does for evolutionary biology. Cant explain everything and sometimes fail to apply but good heuristic for approaching a problem. We can assume they are as a starting point but there are cases where the assumption is false and cases where we don't know and that doesn't impugn dx. Murphy thinks science should search for the psychological processes that fit the folk concepts. Folk theory provides paradigms (only major psychoses correspond

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

'On the two-stage view, the criteria for assessing adequate performance are supplied by nature rather than by a human practice It is not the view that relative to human goals and interests, we can establish what psychological systems should be like and how they should be arranged to meet those goals and further those interests. Rather, it is the view that psycho- logical normality imposes non human, natural functional standards. Those standards exist independently of what people think they should be. He doesn't think that this notion of function creates a problem with respect to non-normative causal processes on the one hand and a normative notion of function and malfunction on the other, however. Some people will say that since even this view licenses statements about what some biological system ought to be like, it is in fact normative in a fairly weak sense All of medicine is normative in this sense the problem is whether any science is not, though, because all sciences license expectations about what ought to happen in a normal system: stars, for example follow a reliable progression through developmental stages, so we can predict what ought to happen to them'. P 85

In the literature on disorder there has been a sustained critique of the notion that function can be used to ground psychiatry in non-normative facts where those non-normative facts consist in purely causal processes. The main line of argument comes from those who think that function of parts are determined by teleological functions of the whole. This line of criticism is largely inspired by Aristotle's teleological notion of function and it seems to rely on Aristotelian notions of a good person. While I don't wish to engage with Aristotle's view here I do think that there is something to this line of criticism. We can see the problem as one of a complete description of the causal processes in our world failing to entail facts about function and malfunction. This is the familiar point that there seems to be a gap between purely causal processes on the one hand and normative facts about function and malfunction on the other.

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

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

One might maintain that these alternative genotypes are malfunctions because the individuals are unable to reproduce. If we consider things from the level of group selection rather than individual selection and individuals with other sexes were found to invest heavily in their kin, for example, then it isn't obvious that they are malfunctioning compared with dif-functioning,

however. It might be the case that there is a fact of the matter as to whether individual selection or group selection is the relevant process for fixing the functions and malfunctions but this doesn't seem obvious to me.

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

It seems that while one cant get normativity from purely causal processes one might be able to get normativity from a conjunction of our explanatory interests together with non- evaluative facts about statistical averages or causal processes. This conclusion is less disturbing for a science of psychiatry than the conclusion that mental disorder is determined by our moral or social evaluations as the anti-psychiatrists maintained, however.

Murphy maintains that it is far from obvious that the relevant notion of

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

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

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

One thing that concerns me about the malfunction assumption, however, is that it is supposed to be what grounds psychiatry as a non-evaluative science and that it seems to recommend a methodology for modelling mental disorders. The methodology seems to be that we model normal or functional biological or psychological processes and then we explain disorders by appealing to breakdowns in the model. Much work in the cognitive neuro-sciences and the bio-medical sciences has been done utilising this approach. We have explanations that characterise delusions as being the result of some kind of breakdown in belief formation and / or retention mechanisms; we have explanations of autism as a theory of mind deficit and so forth. The malfunction assumption cant make much sense of other projects that have been done, however. Instead of working with the malfunction assumption some theorists have worked with a function or adaptationist assumption where certain traits (such as histrionic or psychopathic) may be modelled as evolutionary adaptive strategies. Some theorists have attempted to characterise disorders such as depression, schizophrenia, and anxiety as evolutionary adaptive strategies that result in harm in present environments because environmental circumstances are far removed from those in savannah life.

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

Issue is... Can science tell us whether functions are mathematical or evolutionary? What would science have to find to settle between these issues? Which science do we look to? Social psychologists to see which makes the

best sense of our intuitions (which Wakefield seems to prioritize highly at times)? Why look to the evolutionary sciences. But then why look to the scientists to see what figures out function instead of looking to mathematicians to figure out the relevant reference class.

7.7 Grounding psychiatry: The naturalization cascade

In what follows I wish to focus on some of the recent controversy between theorists who maintain that psychiatry can be grounded in the natural sciences in a way that justifies psychiatry's status as a specialist field within medicine and those who maintain that psychiatry is importantly different from the natural sciences. In philosophical circles, in particular, the debate has come to be bound up in the debate of naturalizing function and dysfunction talk. While the anti- psychiatrists haven't had much to say about the kind of social and / or moral norm violation that is relevant for our judgements of psychiatric disorder I think that we can get clearer on how normativity features in psychiatry by way of getting clearer on how normativity features in other fields such as medicine, biology, and paradigmatically non-normative disciplines such as chemistry and physics.

The way in which I wish to approach this is to utilise this notion that Fulford has recently introduced of a naturalization cascade. His idea is that people have attempted to naturalize mental disorder by showing that the term is logically related to somatic disorder, that disorder is logically related to malfunction, that malfunction is logically related to function, and that function is logically related to purely physical properties and process talk. Instead of considering the logical relationships between terms, I wish to consider the nature of the relevant sciences, however. The logical relationships are supposed to show that psychiatry is grounded in medicine, that medicine is grounded in biology, and that biology is grounded in purely physical properties and processes. This ordering is top down in the sense that it starts with the high level science (psychiatry) and attempts to show that it is grounded in the

comparatively low level science of biology.

I wish to start bottom-up with low level sciences that deal in paradigmatically physical properties and processes, and then work my way up through biology, medicine, and psychiatry. In doing so I aim to show two things. Firstly, there is a normative aspect to all points in the cascade including the natural sciences. This aspect isn't often apparent to us, but can be shown by way of example nonetheless. There are indeed normative disputes in fundamental science such as chemistry and physics and so we shouldn't be surprised that there are normative disputes in the comparatively less fundamental sciences of medicine and psychiatry. Secondly, (and perhaps more controversially) I wish to maintain that the normative aspect becomes amplified at each point in the cascade. What may seem to be a fairly innocuous sort of normativity when our values coincide might seem problematically normative as our values diverge. Where there is more controversy over the norms it seems more apparent to us that the discipline has a normative aspect. The basic idea is that while there is indeed a way of grounding psychiatry in the natural sciences there isn't going to be a way of carving the non-normative aspects off from the normative aspects as the naturalization project had hoped. The normative aspect is tightly bound up in the non-normative aspect and this is precisely why it matters so much that we have an appropriate characterization of mental disorder. The project is thus to how the normative and non-normative aspects are bound to each other such that we can make some genuine progress on this issue.

At this point I should probably say a thing or two about normativity. One notion is that the normative and non normative distinction can be thought of as a prescriptive and descriptive distinction. Another notion is that some concepts are thick (have both normative and non- normative aspects) while other concepts are thin (in the sense of being solely prescriptive or solely descriptive.) Ill have more to say about the kind of normativity in later sections, but offer this merely as a way of initially characterising the phenomenon. The best way to get a handle on the issue that I shall be concerned with is the issue of whether there is a non-evaluative core that can success-

fully naturalize or ground psychiatry in the physical sciences.

If the presence of a malfunction is necessary (though perhaps not sufficient) for both mental and physical disorders and malfunctions (and functions) are determined by purely physical properties and processes then the normative aspects of medicine and psychiatry (both the prescriptive and thick nature of the central notion of disorder) might be thought to result from their status as applied fields. The problem is that theorists have denied that disorders can be grounded in functions that in turn can be grounded in physical properties and processes. The idea is that function talk sneaks normativity in the back door and that as such the naturalization project will fail.

7.7.1 Chemistry, Physics: Physical Properties and Processes

While there has been much controversy over the nature of the physical it seems that we do have one way of getting our head around the issue. The paradigmatic examples of physical sciences are sciences such as chemistry and physics. Chemistry and physics are also often taken to be paradigmatic examples of non-normative or descriptive sciences that deal with non-normatively thin concepts.

Now, it might be said that we value learning about some things whereas we find other things less valuable to learn about. We value learning about certain things and our values determine our explanatory interests (and what we are prepared to fund) with respect to deter- mining what we do and do not study. As such, all sciences are normative or value laden because our values determine that we study the subject matter. One might think that this is a fairly innocuous sort of normativity, however.

Another way in which chemistry and physics could be regarded as normative is that the nature of (at least some of) the subject matter is thick in the sense of partly being determined by our values. If our valuing learning about a certain phenomena leads to us making certain measurements which in turn determines what is observed then there might be a sense in which (at least

some of) the subject matter of physics is thick or partly evaluative in the sense of being partly determined by our values. The idea seems to be that had our values been different the phenomena would not have occurred the way that it did given our values. This normativity also seems fairly innocuous, however. Rocks exist and they would have the nature they have whether or not we study them. While the act of studying them might result in our altering the object of study that doesn't show rocks to be thick or partly evaluative.

Where things might become more problematically normative for the physical sciences in when we consider that there is something that scientists should be doing as scientists. Being a scientist is, of course, a social activity. We can talk about people being better or worse scientists and there is a notion of what it is that scientists should be doing qua scientists. There is similarly a notion of what science is like that is normative or prescriptive. If we say that the aim of science is explanation and prediction, for example, then the claim seems better understood as prescriptive rather than as a descriptive claim about what scientists are doing that could be supported or falsified by what it is in fact that scientists are doing. There thus seems to be normative constraints on science - we can ask which of the theoretical virtues should be maximised if we have to choose between them. Either for a given context or more generally speaking. Such a project seems normative, however, and scientists are thus bound by normative constraints on the way they conduct their experiments, the way they interpret the data, and so forth.

Which norms we accept with respect to the constraints on science are going to affect both how we go about studying and also what subject matter is a fit thing for science to study. One thing to note is that we can, of course, scientifically study a subject matter that is evaluative (e.g., rape). We can study social and / or moral norms scientifically too. We can find out how much they vary across cultures and so on and so forth.

7.7.2 Biology and the emergence of Function talk

In one way of looking at it function talk seems unproblematically descriptive and it is hard to see how the notion of function has generated the controversy that it has generated. Biologists may use the notion of function in a descriptive sense to tell us which traits have been the beneficiaries of evolution by natural selection or they might use the notion of function in a descriptive sense to tell us what role some mechanism or component plays with respect to some activity of a greater system. Both of those characterisations are rough, to be sure, but the notion of function doesn't seem to be particularly problematic in biology.

That being said there has been a lot of controversy over how we are best to understand function talk in biology. In particular there have been concerns over whether there is just one notion or whether there is more than one notion. The majority of the debate seems to have arisen because of the role that biological functions have been thought to play in naturalizing phenomena like representational content or mental disorder. Some theorists have maintained that the notion of function is unproblematically non-normative while others have maintained that the notion of function is normative. Another way would be to say that the notion of function is non-normative but when we talk about malfunction then that is normative.

The notion of malfunction in psychiatry is supposed to be one and the same as the notion of malfunction in general medicine. The notion of malfunction in general medicine is thought to be one and the same as the notion of malfunction in biology. The notion of malfunction in biology is thought to be translatable into non- teleological, non-evaluative physical processes. If this is right then sceptics of psychiatry are wrong to insist that there is no more to mental disorder than that certain people (or conditions) violate certain kinds of social and / or moral norms. I want to show that the relevant notion of function is importantly evaluative or normative. Part of developing as a science will thus involve our sorting out what kinds of values should drive the scientific enterprise.

If the relevant notion of malfunction is shared between psychiatry and general medicine then more must be said on what distinguishes mental disorder (psychiatric) from non-mental disorder (neurological). Murphy argues that there is no coherent conception of the mental that is in play and that the distinction between neurology and psychiatry is due to contingencies of history. Murphy maintains that the distinction is one of extra-scientific concerns and here I wish to argue that Murphy has a narrow conception of the scientific concerns and that that narrow conception of the scientific concerns is unable to fix the phenomenon that is of interest (either with respect to fixing the class of disorders or with respect to fixing the class of the mental).

FULFORD SUMMARY:

In Teleology Without Tears: Naturalism, Neo-Naturalism, and Evaluationism in the Analysis of Function Statements in Biology (and a Bet on the Twenty-First Century) Fulford dis- cusses three positions that one could take on biological functions: causes without teleology (naturalism), teleology without values (neo-naturalism), and teleology with values (evaluationism).

Fulford states: The philosophical project of naturalisation in biology, medicine, and psychiatry has been concerned mainly with five key terms: function, dysfunction, disease, illness, and disorder. The meanings of these terms, moreover, most authors recognise, are linked. The details vary, but broadly speaking they are taken to form a logical cascade. In this naturalisation cascade, as I will call it, disorder includes disease and illness, illness (the experience of illness) is defined by reference to disease, disease by reference to dysfunction, and dysfunction by reference to function The importance, therefore, of biological function statements to the naturalisation project is that they appear to provide a value-free scientific foundation on the basis of which the other terms in the naturalisation cascade can be built up. Most authors recognise that values must come in at some point in the cascade: if not with dysfunction, then with disease; if not with disease, then with illness; if not with illness, then with disorder. But provided biological function statements are value free, the naturalisation project, it is widely assumed, can at least get underway. Pg. 78

Wakefield is a naturalist with respect to function. On disorder, he is an evaluationist. Szasz is an evaluationist on mental illness and a naturalist on physical illness. Kendall is a naturalist on both physical and mental illness. Boorse (in the early versions) was a naturalist on disease and an evaluationist on illness. He has become a naturalist on illness.

Wakefield is a naturalist on these terms essentially because this secures for medicine a basis in science. Medicine is ostensibly value-laden: its key terms of art illness, disease, dysfunction, etc. are, to all appearances, value terms, and this reflects medicines paradigmatically human sphere of practice.

Causal processes cannot be sufficient to mark out functions because causal processes are at work in spheres in which we do not regard there to be functions and malfunctions.

A purely causal biology would contain no functions (only self organising systems). No reproduction (only replication) no death (only radical instabilities in self organising systems. But Fulford thinks that there could be a purely causal biology.

Functions are very much a part of the language of biology.

Explain can mean give the cause of and it can mean give the function of. Argument in other paper that functions cant be reduced to causes.

If causal language is incomplete then what do we need to complete it?

Even if causal language isn't sufficient for functions that doesn't entail that values are what is required in order to complete it. Thornton suggested that intentions could do this. Seems that desires are evaluative, however.

The naturalisation project, then, is driven by the belief that somewhere deep down in the naturalisation cascade there is a value-free foundation, a heuristic holy grail, on which biology, and in turn the theoretical cores of medicine and psychiatry can, in principle, be built up as mature scientific disciplines. Most naturalisers, though, as I noted, and particularly those with a philosophical background, recognise that values have to be brought into the naturalisation cascade at some point.

Some parts of the cascade are more overtly value laden than others. Psychiatry more than medicine medicine more than biology and biology more than the natural sciences such as physics. Much of the debate about mental illness is in effect a debate about how the more overtly value-laden nature of mental illness compared with physical illness should be interpreted: for psychiatrists (e.g., Kendall) and some philosophers (early Boorse) the more overtly value-laden nature of mental illness is epiphenomenal to its underdeveloped status as a science. For those opposed to psychiatry, on the other hand (e.g., Szasz) this feature of mental illness is a sign that it is not really illness at all like physical illness, but moral or life problems.

Values in values out debates.

Whatever view one takes about the possibility of a value-free science, it is common ground that some areas of science are more overtly value laden than others. Human sciences more than the bio and the bio more than the physical.

So there is nothing wrong with naturalisers bringing values into the naturalisation cascade. To the contrary, naturalisers HAVE to bring values in, at some point or another, in one way or another, and either to endorse their logical role or to show that it is epiphenomenal, if they are to account for the presence of values as a given feature of the naturalisation cascade. ANY account of the cascade, therefore, whether naturalistic, neo-naturalistic, evaluationist, or some other account altogether, must explain what values are doing there.

Naturalisers are wrong (or at least incautious) in being matter of fact about values.

Reason 1: LOGICAL INCOMPLETENESS

Values have a fine structure that people have neglected or failed to deal with adequately.

They are negatively valued and they are a particular kind of negative value. Biological / medical / psychiatric value is distinct from, for example, moral,

aesthetic, prudential, and a host of other kinds of value. (mad or bad is a familiar example in forensic psychiatry). Also peculiar closeness in the sense that the values in the cascade must be intrinsic to the cascade. It must be part of the meanings of the terms and not a contingent add-on.

Reason 2: LOGICAL INCONSISTENCY

no ought from an is then, means that to get an evaluation out of the meaning of a term you first have to put an evaluation in. Hence, need an evaluation at a lower point in the cascade to get an evaluation out of a higher term. Logically inconsistent to leave values out of some points in the cascade and introduce them in another part. They need to either be in or out.

The recursive causal processes that in modern evolutionary theory are believed to drive natural selection do indeed explain, without recourse to teleology, the emergence of biological forms. But similar processes explain similarly the emergence of meterological, stellar, and geological forms. In all these evolutions, there are among the properties of the systems in question one or more properties that, in the phraseology adopted by recent naturalizers, offer the best explanation of why the system of which it is a part is there. But we speak of functions only in the case of biological systems. Ergo, recursive causal processes, since they drive the development of non-biological evolutionary forms (of which we do speak of functions), cannot be sufficient to mark out functions (Fulford, 2000 p. 79).

Fulford (2000 p. 79) isn't sceptical about a purely causal biology being complete in the sense that a purely causal physics could be complete. We wouldn't have any functions, malfunctions, death, etc, however.

It is one thing to show the negative conclusion that the language of causes is incomplete. It is quite another to show the positive case that it is values, rather than some other Factor X, that is required to complete it (Fulford, 2000 p. 80-81) [he is contemplating Thorntons approach here].

The move from function to dysfunction is a key step in medicines project of naturalizing disease. Thornton argues that salience and naturalness, or

other similar linguistic resources derived from the space of reasons, can be used here as well as at the earlier stage of defining functions. But it is not clear (and he does not say) how this could be done without involving the evaluative element in the meanings of these terms. Suppose, for example, that the function of a biological system could be distinguished from its other properties by some descriptive aspect of the meaning of salience, by the fact that the relevant property is (to draw on the descriptive meanings noted a moment ago) jutting, projecting, or prominent. I don't say that it can, but suppose that it could. How, then, would we move on to distinguish, using only the same resources of descriptive meaning, good from bad, successful from failed, functioning? To be more or less jutting, to project in a different direction, to be to a greater or less extent prominent would all, in the absence of evaluative meaning, be merely, to function differently. How, then, are we to distinguish dysfunction from, to coin another neologism, dif-function (Fulford 2000 p. 82).

The failure of natural selection to naturalize function statements in terms of causes does not prove that such statements are (even partly) evaluative in meaning. Similarly, then, the failure of any particular approach to naturalizing function terms is not, in itself, proof that the whole project of naturalizing such terms, and indeed of naturalizing any of the other terms in the naturalization cascade (dysfunction, disease, illness, disorder, and so forth), is void (Fulford, 2000 p. 83).

Their event horizons for values, as I put it in the first section of this paper, differ: Boorse's event horizon, in his earlier work at any rate (1975), was between disease and illness; Wakefield's is between dysfunction and disorder; Thornton's is between dysfunction and disease. But philosophical naturalizer's bring values in somewhere along the line. Well, there is nothing wrong with that, you may say. After all, it is common ground, among naturalizer's of disease, as it is between naturalizer's and their opponents, that some parts of the naturalization cascade are more overtly value laden than others: psychiatry is more overtly value laden than more high-tech areas of medicine, medicine more than biology, and biology more than the natural sciences, such

as physics. This is common ground, too, in the closely related debate about the validity of mental illness. Much of that debate, as I have shown elsewhere (Fulford 1989, Chap. 1), is in effect a debate about how the more overtly value-laden nature of mental illness compared with physical illness should be interpreted: for psychiatrists, such as Kendell (1975), and indeed some philosophers (such as Boorse, 1976), the more overtly value-laden nature of mental illness is epiphenomenal to its underdeveloped status as a science. For those opposed to psychiatry, on the other hand, notably one of my co-contributers to this issue Thomas Szasz, this feature of mental illness is a sign that mental illnesses are not really illnesses at all, [from 83-84] like physical illnesses, but moral or life problems (1960). (Fulford, 2000 p. 83-84).

(Fulford, 2000 has an account of the structure of values. Negative. Psychiatric / medical / biological different from moral and aesthetic. Different people with different interests in different contexts and at different times).

7.8 Norms and harms

7.8.1 Dysfunctioning behaviour and problems in living

Wakefield critiques the DSM view by being too liberal in maintaining that the relevant dysfunction can be biological, psychological, or behavioural. In particular, he maintains that behavioural dysfunction is insufficient for the dysfunction criterion. His argument for this consists in examples that are meant to act as intuition pumps. The form of the objection is that we have the intuition that only behaviour that is caused by inner dysfunction (either biological or cognitive) is necessary for disorder. He provides the example of a person who is unable to read. They thus meet the DSM criteria for a reading disorder. If we find out that the person had received instruction that was comparable to other people and they had learned to read whereas this person had not then we would have the intuition that there was something wrong with this persons biological / cognitive mechanisms and they were dysfunctioning. If we learn that the person had not received adequate instruction, however, then even though this person might have the same be-

havioural presentation as the other person we would not have the intuition that the person was disordered, however. Wakefield uses this example to attempt to persuade us that behavioural symptoms are insufficient, and that the causes of the behavioural symptoms matter.

7.8.2 Subjective vs objective

Wakefield seems to think that biological functions and dysfunctions are objective features of the world in much the same way that species are. The thought is that they are a fit matter for scientific discovery. Harms, on the other hand are not thought to be objective features of the biological world. Instead, Wakefield maintains that harm is subjective - whether an individual is harmed or not, whether a society is harmed or not is a matter of 'subjective opinion'.

This distinction is too simple, however. It is possible to have a science of norms and it is surely the case that there are facts (objective, scientific facts) about what norms a given society adopts. It also seems that there will be facts about whether an individual is violating the norms of their society. While there might not be objective scientific facts about what a given culture or society should adopt (unless we relativize it to certain aims or goals) there would seem to be objective facts about both the norms that are endorsed by a given society, whether behaviour is in violation of those norms, and what norms a society should adopt relative to certain aims or interests.

Wakefield maintains that harm is a person-level, normative notion. That it is behaviours that are harmful or not. If we think about the pre-theoretical notion of harm it seems that there objective facts about it. Whether or not an individual is harmed or not. We might wonder why Wakefield thinks this is normative / evaluative.

Wakefield considers that normativity is a person level notion and that harms thus apply to the person (and / or society). It is thus behaviours (in a society) that are harmful or not. For all that Wakefield has said about the notion of harm it would seem that there are objective facts about whether an individual

and / or society is harmed, however. Once we have fixed the relevant norms in the society we should be able to read off from the conjunction of the behaviour and the normative standard whether the behaviour is in violation of the normative standard or not. If the behaviour is then we can say that the individual and or the society is harmed whereas if it is not then the person (and others) are not.

Wakefield seems to think that behavioural dysfunction is insufficient for disorder when it occurs in the absence of inner dysfunction. He offers an example that is supposed to pump our intuitions in this direction. Since I wish to question Wakefield's distinction here I will need to offer a different response to the example than Wakefield does. Lets consider the example: A person can't read. If we learn that the person has had instruction that was comparable to the instruction that others received and yet they were able to read and he was not then we are inclined to think that this person has an inner dysfunction that prevented him learning and this person is thus disordered. If instead of learning that he had sufficient instruction we learned that he had never received any instruction then our intuitions are quite different, however. We seem to have no grounds for considering that he has an inner dysfunction and we also seem to have no grounds for considering that this person is disordered.

Wakefield uses this example to conclude that behavioural dysfunction (inability to read) is thus insufficient for disorder - even if we agree that the normative aspect is also present and that this person is harmed in virtue of not being able to read. He concludes from this example that the relevant dysfunction must be internal to the person and must be the cause of their behaviour that is harmful.

7.8.3 The distinction between dysfunction and harm

The first thing to note is that the motivation for the harm component comes from the idea that it seems plausible that a person could have a malfunctioning mechanism and yet not be harmed. An example of this would be someone

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

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

There are a number of notions that Wakefield runs together that might best be considered apart initially at least. On the one hand we have an aspect that is objective in the sense of being discoverable by science. This objective aspect is thought to apply to inner states of people (according to Wakefield though not the DSM). Facts about dysfunction are supposed to be culturally invariant (universal). On the other hand we have an aspect that is subjective

in the sense of not being discoverable by science. The normative aspect is thought to apply to behaviour. Whether an individual is violating norms is meant to vary and hence not be universal.

Wakefield might be thought of as a dualist insofar as he presents these two separate aspects of disorder. He might be thought of as a reductionist insofar as he focuses in on the objective aspect of disorder (in the exclusion of the normative aspect) in his attempt to ground psychiatry in the natural sciences. He doesn't have much to say about the normative aspect of disorder at all other than to dub it 'harm'.

What do they mean to say by maintaining that the relevant notion of malfunction is objective? The notion seems to be that the relevant functions (and malfunctions) are to be discovered by the sciences of the mind/brain. To say that the relevant notion of function / malfunction is normative seems to be the claim that the sciences of the mind / brain cannot determine the relevant functions / malfunctions. Often it seems to be thought that the function of a part of a person cannot be determined until we know the function of a whole person. The function of a whole person is notoriously difficult to figure out but it would seem that the DSM has something to say about this. Social, occupational functioning etc. Hard to figure out The relevant facts here would seem to be social with respect to how well the individual is fitting into and flourishing in society Sociology, anthropology.

To say that employing the above notions of function is normative is to make science into a normative endeavour. As Murphy notes science often deals in idealised or normal processes such as the normal development of the eye or a star or an ecosystem or whatever. It isn't problematic for the other sciences that they employ these notions of norms and hence it isn't problematic for the sciences of the mind / brain if they employ them either.

If the malfunction can be behavioural (and does not require an inner malfunction) then there must be behavioural functions. Wakefield argues against harmful behavioural malfunction being sufficient for mental disorder on the grounds that it delivers verdicts that are counter- intuitive. We could con-

sider that the DSM criteria list behavioural malfunction such that if someone meets the behavioural criteria for mental disorder then they are thereby malfunctioning in their behaviour. Wakefield considers that even if someone were to meet behavioural criteria for reading disorder our intuitions as to whether they are mentally disordered or not depend on our intuitions as to whether there is inner malfunction or not. If someone meets criteria because they have inner malfunction then intuitively they do have a mental disorder whereas if they meet criteria because they have never been taught how to read then intuitively they do not have a mental disorder.

7.8.4 Problems with the subjectivity of harm

Another sort of worry we might have is over the subjectivity of harm. While it is a common view in the literature on mental disorder and on medical disorder that whether an individual is harmed or not is subjective this seems to be something that we might want to avoid once we consider it further. We considered before that we might well have the intuition that a person who has a brain tumour that will kill her is harmed by that brain tumour even if the individual, some members of the individuals society, or all members of the individuals society think that she is not harmed.

We might thus think that there are facts about whether an individual is harmed or not and that particular individuals or perhaps even whole societies can simply be mistaken in their views about whether the individual was harmed. We might consider that the attempted genocide of the mentally ill and the Jewish people was wrong and that part of why it was wrong was that those people were harmed, for example. Even if the dominant view of Nazi society was that these people weren't harmed (perhaps because they didn't constitute people) that doesn't seem to undermine our intuition that they were harmed in fact. Similarly with respect to people being enslaved, animals being kicked etc.

It also seems that we can (and indeed we have) a science of norms and values. We can survey people and find out about what values they profess to hold

and while there are problems with correlating verbal reports with action it seems that this is one way we might get at the values that people hold. It seems that there are objective facts about what values a particular person or a particular society actually holds. It would thus seem that there are also objective facts about whether an individual is acting in accordance with or violating the norms of their society. Even if we are skeptical about their being ethical facts that transcend all societies and cultures it seems that there are facts about current norms and whether someone is violating current norms though what societies norms *should be* and whether an individual is violating those idealized norms remains more problematic.

Norms and Science

While there is controversy over whether there are ethical facts and about the nature of those facts (if they supervene on non-normative facts, for example) it seems inadequate to say that normativity can't be the subject matter of science as Wakefield seems to assume. There are already sciences that investigate normativity, from attitude assessment in social psychology to the cross-cultural investigations of anthropologists and, of course, evolutionary psychologists.

Cross-cultural variability vs universality

Wakefield runs together subjectivity with cross-cultural variability and universality with objectivity. The idea here is that there is something disreputable about norms because they vary across cultures whereas science is about objective facts where universality is conflated with objectivity.

This is too swift as there can be objective facts about different cultures as we have al- ready considered. There are also objective facts (I'm sure Wakefield would agree) that aren't universal. Species aren't eternal (on most views), for example, and yet they are often taken to be paradigmatic of natural kinds that are a fit subject matter for the biological sciences.

He ties the notion of harm to the behaviour, however. Wakefield doesn't say a

great deal about the notion of harm. It is clear that for Wakefield the notion of harm is a notion that is supposed to cover the normative aspect of mental disorder - but he is more interested in showing psychiatry to be grounded in the natural sciences than in offering an account of that normative aspect. The notion seems to be that behaviour that is harmful in one society may well not be harmful in another. His notion also seems to be that the notion of harm can be distinguished from the notion of dysfunction, however, as in a society where nobody reads it might be the case that the person with the inner dysfunction who would not have been able to learn had reading been something he was instructed in isn't harmed by his not reading in a society where people don't read.

The examples that show there to be inner dysfunction in the absense of harm, and harm in the absence of inner dysfunction are supposed to show these to be relatively independent of each other. So much the better for the grounding project.

Wakefield seems to think that this is analytic but it is important to note that there is a great deal of controversy over what levels it is appropriate to speak of the process of evolution by natural selection acts on. We might be inclined to say that genes replicate and so the only thing that can be selected for are genes. Then it would seem that in order for there to be inner dysfunction there would have to be genetic dysfunction, however. There are problems with this account as the inclusive fitness of genetic replication seems far removed from our interests in persons in psychiatry.

Wakefield seems to be thinking that the dysfunction will be neurological. While genetic replication is artificial because genes only have the advantage they have in virtue of their effects in a particular environment it would seem that similar issues would come up for neurology. Whatever is going on with the neurology selection would have to work on the individual. Tricky... Tricky... But then why not talk about selection for those traits? The reading example that Wakefield offers is rather artificial. It is unlikely that whatever mechanisms subserve our capacity to read have that as their evolved function. While it is frequent that people have a theory of the causes of mental

disorder which involves something like 'autistic people seem to lack theory of mind, which is to say they have a malfunctioning theory of mind mechanism' we must be wary of such an inference from behaviour to malfunctioning inner mechanism in a way that doesn't have independent evidence to believe the inner mechanism is malfunctioning.

Neuroimaging can show us difference. Difference in systemic function perhaps. Unclear to see how it is relevant for evolutionary function, however.

Wakefield needs to make sense of theorists who don't grant that mental disorders are evolutionary dysfunctions - or theorists who maintain that it is still up for grabs. Murphy outlines different strategies that one could take if one was seeking an evolutionary explanation for mental disorder. While one strategy is to assume inner dysfunction another strategy would be to attempt to show (or to investigate whether) mental disorders are inevitable by-products of something that confers a selective advantage. One candidate could be creativity, for example. Another strategy would be to say that these people aren't malfunctioning - rather society has changed so significantly that what would have been adaptive in past (hunter gatherer environments) simply isn't adaptive now.

There are problems with precisely when evolutionary functions are supposed to be fixed. If we build in that functions are relational - partly dependent on the outputs, but also partly dependent on how they fare relative to other outputs and also dependent on how all of those fare relative to a specific environmental niche then it would seem that functions would be dynamic. Insofar as the variants change over time (e.g., a population moving to fixation or some variants being weeded out) the fitness value could alter. As the environment changes the fitness value relative to other variants could alter. It is unclear why a particular period of time (where we don't know a great deal about either the environment or the variants) fixed the mental functions once and for all. Or insofar as they do it is unclear why we should care about that notion for psychiatry.

We do seem to have this intuitive distinction between there being something

wrong with the individual being distinct from there being something not good about the environment for that individual. While the majority of theorists are sophisticated now to appreciate that instead of it being either or they engage in dynamic reciprocal relationships it might be that the notion of inner dysfunction and behavioural harm in an environment are similarly intertwined such that there isn't a great deal of utility in pulling them apart, however.

The notion of a poor fit - Wakefield could grant. What he doesn't grant is that the person counts as disordered without that inner dysfunction, however.

There are problems with fixing the functions and there are thus problems for fixing the dis-functions. Dysfunction is often an assumption that is built into theories rather than something that is discovered by them. We could describe the facts but disagree about what dysfunctions there are insofar as we think that different things fix the relevant functions. Wakefield thinks that it is fairly simple to conclude that the relevant functions are fixed by natural selection but this is simplistic firstly because biology makes use of other notions of function and secondly because there is a great deal of controversy over the notion of evolutionary dysfunction.

His account is useful, however. He says that while we might not know whether there is a dysfunction or not it is the fact that there is that determines that the person is disordered. In order to be justified in believing that a person is disordered we need to be justified in believing them to be mentally dysfunctioning, however. Everyone grants that there is something wrong with the behaviour of people with disorder. The joke has been, however, that 'everyone can tell the brains of the schizophrenics - they are the ones that look normal'. Even if we find a difference getting from there to a dysfunction is problematic.

But suppose we grant Wakefield his dysfunction view. I'll come back to further complications on the behavioural vs inner and the dysfunction vs harm. At this stage I want to consider that even if we agree one some notion of dysfunction - how is that relevant for psychiatry? If dysfunctions come too

cheap then it would seem that the bulk of the work is being done by the harm notion. It would seem that initially what concerns us is the behaviour. We want to do something about that. It would seem to be the case that the best thing to do about that would be to have some notion of what produced it and attempt to alter that. That seems to be the difference between prescribing a course of instruction and not prescribing a course of instruction - rather? What? Attempting to get a cognitive psychologist or something? Give up?

One way of attempting to define psychiatry is that psychiatry is concerned with a particular variety of treatment. Prescription of medications. This seems contingent, however. Initially it was distinctive in the sense of running an asylum. Then it was distinctive for the variety of 'talking cure' (psychoanalysis). Now it is distinctive in the sense of the medications. There doesn't seem to be anything that suggests that psychiatry should be delineated or defined with respect to the current treatments that it employs. But if the treatments don't make psychiatry distinctive then the subject matter must make psychiatry distinctive. mental disorders must be different from neurological disorders or from problems in living that result in people seeking out marriage guidance counsellors or careers or losing weight advice.

In offering his analysis of the concept of disorder Wakefield maintains that he is attempting to provide an account that is broadly in keeping with the accounts that have been offered by theorists such as Kripke and Putnam with respect to natural kind terms such as gold and water. Wakefield thus is committed to a causal theory of reference and the view that 'mental disorder' is a natural kind term. One virtue of this style of account is that it clearly distinguishes between the phenomena in the world (and its nature) and the concept that we have and our beliefs about the concept that we have. It seems that we are a great deal less interested in offering an account of the concept of disorder (or the beliefs about disorder) that people have. We are a great deal more interested in offering an account of what the nature of disorder is.

This also fits in nicely with the account that Paul Griffith's adopts towards developing a theory of emotions. The thought is that while one can attempt

to get people to say what they think emotions have in common and what is necessary and sufficient for someone to have guilt as opposed to remorse this seems to be a fairly different project from one in which we attempt to investigate what (if anything) emotions have in common by way of the empirical sciences.

One thing that Griffiths ends up concluding, however, is that emotions aren't a natural kind term after all, though there may be three natural kinds in the vicinity of our term 'emotion'. Natural kinds are a considerable issue... Theoretically useful in science. One might well ask 'how much do they have to have in common' or 'what kinds of features do they have to have in common' in order to count as a natural kind? Prinz also investigates scientific theories of emotions and he ends up offering an analysis of necessary and sufficient features that a state must have in order to be an emotion. He concludes that since emotions have those features in common that he enumerated emotions are natural kind terms.

I think that the dispute between Griffiths and Prinz is more apparent or verbal than real. Griffiths focuses on the differences between the three notions in the vicinity whereas Prinz focuses on the similarities between the three notions. They might agree on all the features but disagree whether the features are enough for 'natural kindhood'. Seems to be more a dispute about how many features (and what kinds of features) are enough for something to deserve the term 'natural kind'.

One might similarly wonder whether mental disorders are a natural kind. There are three worries that we might have in the vicinity. One is the boundary between disordered and not disordered. Are there indeterminate cases or is there a fact of the matter either way? do the disordered and / or the non-disordered have enough of the relevant kinds of features in common in order to deserve the label 'natural kind term'. It is unclear. The second is whether mental disorders have enough in common (or physical disorders) such that they form natural kinds. The third is whether particular disorders or syndromes form natural kinds.

7.8.5 Person-environment fit

In considering evolutionary models to be dynamic or temporal such that they capture population dynamics over time the possibility arises that psychiatric disorders may have been adaptive in the evolutionary sense at some point in the past but that due to recent alterations in our environments they are no longer so. This notion brings out an interesting point that variants are only adaptive in relation to other variants and also in relation to the environment that the variants are in. While the case of drift shows that variants aren't adaptive or maladaptive to their environment simplicitor - but rather are adaptive or maladaptive in relation to the other variants (relative fitness seems important here) another important feature of evolutionary explanations is that a trait is only more or less adapted in relation to the environment. It seems senseless to ask whether gills are better adapted than lungs until we know something about the environment in which the organism exists.

Homo Sapiens have radically reconstructed their environments. We build buildings, work in high rise office blocks in front of computers, live in inner city apartments, and negotiate public transport systems or navigate our own motor vehicles. We have complex social structures of banks and educational schools and government departments. What is required in order to be considered 'not significantly impaired in ones social, occupational, or educational functioning' are obviously different in at least some respects from the conditions our ancestors operated under in the plestocine. While a phobia of falling might significantly impair one who is expected to live and work in high rise apartments and fly around the world for business and / or family a phobia of falling might have positively benefited those who lived in environments where heights were usually cliffs where wind was particularly strong. Similarly, while social aloofness and communication with non-apparent things might impair ones negotiation with social services, potential employers and educators such behaviour might result in a very different outcome in a society where such behaviours result in a person being revered as a holy prophet or healer with special gifts.

While biological explanations are sometimes contrasted with social explanations and biological explanations are sometimes thought of as genetic and invariant or fairly inevitable in development whereas social explanations are thought to offer an account of cross cultural variation etc it seems that evolutionary views have more to offer with respect to cross-cultural variation than is commonly supposed. In particular, if we develop good models of psychiatric disorders as dysfunctions in current populations and accept the individuation of disorders at least in part on these grounds this still doesn't seem to rule out the possibility of advocating that we alter society our social attitudes towards those regarded as mentally ill (e.g., Szasz) rather than altering the individual (e.g., the medical model).

7.9 The normativity of harm

We have already seen that medical model theorists have attempted to define disorder largely in response to the one stage normative view. They have thus focused on offering an account of an objective biological aspect of disorder rather than focusing on attempting to offer an account of the normative aspect. While one-stage objectivists maintain that normativity doesn't play a role in disorder few theorists seem to adopt a one stage objective view. The majority of theorists thus accept that there is a normative aspect to mental disorder (and indeed to disorder more generally). The majority of the controversy has been over whether there is an objective aspect and so Wakefield is in keeping with the majority of the literature when he focuses in on attempting to naturalize an objective aspect of disorder rather than attempting to provide an account of the normative aspect.

Wakefield uses the term 'harm' as a place-holder for the normative aspect. He maintains that 'harm' is a person level notion as it is individuals or groups of individuals who are harmed. He agrees with the APA's definition of 'harm to the person and / or to society', but he doesn't offer much in the way of characterizing the normative aspect in any more depth. 'Harm' is to be understood as a placeholder for the normative aspect on Wakefield's view.

He does argue that harm is necessary for disorder, however, as we shall soon see.e doesn't attempt to naturalize 'harm' (because he views it as normative). There is a dictum that has guided much of ethical theorizing in philosophy and seems to be what is behind Wakefield's view. The Dictum is that you 'can't get an ought from an is'. The notion is that ethical facts are distinct from non-ethical facts (in the sense of being pursuable independently from them). Wakefield may have taken the lesson from this that one shouldn't attempt to naturalize the normative aspect. Whether it is possible to pursue the non-normative aspect in isolation from the normative aspect as Wakefield attempts to do may be more problematic, however.

Chapter 8

Conceptual analysis meets empirical discovery

The need for a framework

Over the years quite a literature has accumulated on the bio-medical notion of disorder (see Boorse, Wakefield, Szasz, Murphy, Neander, Fulford, Coopers, the American Psychiatric Association and its critics etc). While some theorists are interested in the medical notion for its own sake, the majority of interest in the medical notion seems to have been driven by attempts to characterize the psychiatric notion in particular. The thought is that in once we have a characterization of bio-medical disorder then we will be able to see whether mental disorders are biomedical disorders or not. The debate is often referred to as the 'medical model' vs. the 'anti-psychiatry critique' when those who maintain that mental disorders are not bio-medical disorders are themselves psychiatrists. We can see a similar skeptical position upheld by social constructionists in the social sciences, however.

Accounts of both bio-medical disorder and mental disorder disorder remain controversial. There are a spectrum of views about mental disorder from biological reductionist to eliminativist to cognitive psychological to social constructionist to value theoretic, to combinations of the above. Even from

within the most highly prevalent medical / biological approach there are controversies over whether bio-medical disorders are appropriately conceived of as genetic and / or neurological and / or cognitive, or whether they are better characterized as functional behavioural symptom clusters that share a certain aetiology and course.

There is also much debate over whether conditions such as addiction, sociopathy, attention deficit disorder, and paedophilia are appropriately characterized as disorders or not. There is a similar debate within medicine with regards conditions such as infertility and polycystic ovarian syndrome. This latter issue of whether particular conditions are appropriately characterized as disorders or not is often thought to be the main motivation as to why providing an appropriate account of disorder matters. While some theorists are quite explicit in maintaining that science will discover the answer to these questions others (though perhaps more implicitly) seem to think that these issues will be settled by conceptual analysis.

While it is only too easy to get lost in the details of defending a particular view against counter-examples or attempting to offer counter-examples against a particular view here I want to attempt to stand back and offer something more diagnostic[1]. Murphy has faulted the majority of work that has been done on the notion of disorder for giving too much weight to conceptual analysis and to our pre-theoretic intuitions. He maintains that the project needs to be reoriented from one in which 'conceptual analysis sets the subject of inquiry and the constraints under which it proceeds' to one in which a variety of considerations both conceptual and empirical mutually inform and where both sorts of considerations are revisable (Murphy, 2006; Murphy & Woolfolk, 2000b, 2000a).

While I am sympathetic to Murphy's point that we are better to think that there is an inter-play between conceptual analysis and empirical discovery I think that much more can and indeed needs be said about how these projects

[1]I will turn to a detailed analysis of the most prevalent 'harmful dysfunction' view in the next chapter.

relate. The notion that both a-priori and a-posteriori intuitions are revisable is controversial as one or the other is often taken to be necessary and sufficient. We thus need an account of how both sorts of considerations are supposed to be revisable. Murphy also places himself very much on the empirical end of the spectrum when he maintains that the majority of work that has been done on the notion of disorder is an 'impediment to scientific progress.' While he does acknowledge a role for conceptual analysis I am concerned that with that statement he may have gone a little too far in prioritizing a-posteriori or empirical considerations and he may not be appropriately appreciating the significant conceptual constraints.

8.1 Figuring out the role the concept plays

In order to answer a question of the form 'what is x?' we need to begin by getting clearer on the role that x plays. One very traditional way of doing conceptual analysis is to begin by getting an individual to reflect on the concept that one wishes to analyze in order to come up with a list of features that are thought to be central to the concept. While there are problems with this conception that I will get to in the next section, for now, let us grant the assumption that a feature list generated from introspection is a worthwhile place to *start* the inquiry at least, and take the following example as illustrative of a feature list that might be generated for the concept of mental disorder on the basis of introspection:

- (a) People with mania, depression, and psychosis are paradigmatic cases of people who have a mental disorder
- (b) There is something wrong with people who are mentally disordered
- (c) Science will discover what is wrong with people who are mentally disordered
- (d) It is better to not be mentally disordered than to be mentally disordered
- (e) People who are mentally disordered have the right to treatment for

their bio-medical disorder

- (f) People who are mentally disordered are dysfunctioning
- (g) People are not morally responsible for behaviours that result from their mental disorder
- (h) People who are mentally disordered are harmed by their mental disorder

The most significant problem with starting with feature lists is that we don't merely want to get at features that people take to be central to their concept, rather we want to get at the necessary and sufficient features for the concept. This way of putting things is problematic as one thought is that mental disorder might be a cluster concept rather than one amenable to treatment in terms of necessary and jointly sufficient conditions.

It has been pointed out, however, that one can always take a cluster concept analysis and transform it into a necessary and sufficient condition analysis. It must be necessary that a certain number of the features in the cluster analysis are instantiated, and instantiating a certain number of them must also be sufficient. We can call this property of instantiating a necessary and sufficient number of the features p and then say that p is both necessary and sufficient for mental disorder.

Some people resist this move by maintaining that we want to get at natural properties or projectable properties or some such. The arbitrary properties that we get by the above manoeuvrer are the wrong kind of properties. Saying precisely what is wrong with these properties is problematic, however. I'll have much more to say about the notion of 'natural properties' later. For now, I want to focus in on the idea of offering necessary and jointly sufficient conditions for the notion of mental disorder, however. It should be understood that the conditions could be either 'natural' or 'arbitrary' at this point.

The following problems are all problems to do with how we individuate concepts with respect to necessary and sufficient conditions. They are the 'prob-

lem of grain' in the sense of how fine (specific to individuals) or coarse grained (shared between individuals) we want our analysis to be.

8.1.1 Disagreement between individuals

The first problem with feature lists (that should seem initially apparent) is that different individuals may well list different features. For example, one theorist might say that 'people who are bio-medically disordered have a right to treatment for their bio-medical disorder' while another theorist might deny this. One might be tempted to say that insofar as different individuals list different features they have different concepts of bio-medical disorder. This way of putting things raises two problems, however. Firstly, it seems to make it something of a mystery as to how genuine disagreement is possible. On this view different theorists wouldn't be disagreeing about which features are appropriately included in the feature list so much as talking past each other. Secondly, language is a social activity and part of learning a language involves hooking into a social network of concepts. It seems clear that we are more interested in getting at the social concept that is common to different individuals than getting at the concept that is particular to different individuals.

8.1.2 Disagreement between groups

The second problem with feature lists is that even if individuals from within a social group were found to agree there may still be considerable disagreement in the feature lists that are provided by different social groups. Some of the survey work that has been done on knowledge suggests that members of a cultural group can have very persistent and strongly held intuitions about paradigmatic cases and attempted definitions that are very different from those of other cultural groups. One might be tempted to say that insofar as different cultural groups list different features they have different concepts of bio-medical disorder. The problem with this, once again, is that it it would make it something of a mystery as to how different cultural groups could disagree. It seems that we don't simply want to know what the American,

Australian, or Indian concept of disorder is - we want an account that is common to all.

8.1.3 Defining the experts

Sometimes it is suggested that we should defer to the relevant experts and thus prioritize the features that are listed by them. This presupposes that there will be considerable agreement among experts, however. Murphy maintains that there might well be different concepts of disorder insofar as we have the scientific conception, the legal conception, and the moral conception. He also maintains that they are related, however, and that in particular, the scientific conception should feed into and inform the other conceptions. While I will consider how these projects relate in due course, for now it is important to note that there is a problem with deference to the experts insofar as there is considerable disagreement among experts.

If we focus in on the scientific conception, for example, then we can see that there is considerable disagreement between scientific theorists e.g., clinical psychologists, psychiatrists, anthropologists, sociologists. Even within a field there can be considerable disagreement as anti-psychiatrists are psychiatrists too.

8.1.4 The problem of individuating concepts: Necessary vs contingent features

The above problems seem to illustrate the point that while we can ask people (and indeed reflect ourself) on our concept it might be that the feature lists that are generated don't capture the *content* of the concept after all, but instead capture what people *take the content to be* where they might well be mistaken. The thought here is that while brainstorming a list of features (or surveying others and getting the features they brainstorm) is one thing we don't merely want a list of peoples associations. We want to start with feature lists in order to get at the notion of disorder where that will help us see what disorder really is. The best way to see this criticism (initially at

least) is to think of it in terms of necessary and sufficient conditions.

To say that a feature is necessary is to say that the property, process, entity etc could not lose that feature and yet continue to exist *as that* property, process, entity etc. To say that a feature is contingent is to say that the property, process, entity etc could lose that feature and yet continue to exist *as that* property, process, entity etc. So, for example, in the case of gold a sample of gold could not lose the property of its atomic weight without ceasing to be a sample of gold. This is because atomic weight is regarded as a necessary feature of gold. A sample of gold could lose the property of being liquid by becoming solid and still continue to be gold, however because liquidity is thought to be a contingent property.

Now, what reason do we have to think that people will list the necessary and sufficient features (the content of the concept) rather than listing contingent features that are merely associated with the content? The problem is that people might not have access to the necessary and sufficient features. This could be either because the person isn't suitably reflective (isn't able to follow through the relevant consequences) or because the person isn't suitably observant (where features aren't accessible to introspection but instead require empirical observation of the world). It seems clear that while feature lists might be a useful place to start there is still more work to be done in the form of figuring out which of the features on the list are necessary and in the form of figuring out how to suitably idealize features on the list such that we can get to necessary features that might not be mentioned.

Getting to necessary and sufficient conditions will require idealization. There is a further way in which idealization seems necessary, however. It seems that we simply aren't all that interested in what notion of disorder people *actually* have, and we are a great deal more interested in what notion of disorder people *should* have. As such, issues of concept individuation aren't centrally important. Whether or not different individuals, cultural groups, or experts actually have different concepts or not is less important than the issue of which concept different individuals, cultural groups, and experts *should* have.

8.2 Kinds of features

Even though there are good reasons for taking the feature lists generated by one competent speaker with at least a grain of salt it does seem that we have to start somewhere. Starting with the features listed by competent speakers (including their intuitions about judgments of cases) seems to be a fairly intuitive place to start. If we don't start here, at least, then where can we start? There don't seem to be any other candidates on the table[2]. What I want to try and do now is to talk about different kinds of features that may appear on feature lists. By calling them different 'kinds' I don't mean to say or imply that they are categorically different from one another. The reason why I want to explicate them in this way is that I think that this is a useful way of assessing where different theorists are at with respect to the different projects they are engaged according to which of the following kinds of feature they prioritize. By way of preview, the features I wish to consider are:

1. Judgment of cases
2. Bridge features
3. A-Priori descriptive features[3]
4. A-Posteriori descriptive features

[2]Though it may be that people have suggested different approaches that I don't know about. Haslanger has some interesting stuff to say about finding out about the social role of the concept through historical analysis including noting discrepancies between what people have to say and what people do. She thus distinguishes between the explicit role that the concept plays in theory and the implicit role that the concept plays in structuring our social practices. I'll have more to say about this project (which might be the project of 'history of ideas' that Focault etc are engaged in) later.

[3]By calling these features 'A-Priori' I really don't mean to suggest at all that the Judgment of Cases and the Bridge features are not A-Priori. By calling these features 'descriptive' I also don't mean to suggest that either the Judgment of Cases or the Bridge features can't be explicated in a way that is descriptive. By 'descriptive' I also don't mean to suggest that they are contingent rather than necessary - this is why I've called the a-posteriori features descriptive as well. I'm trying to remain neutral on necessary vs. contingent features at this stage as I'll talk about it at length later. If anyone can think up a better name for these features I'd be grateful.

8.2.1 Judgment of cases

The first kind consists in our intuitive judgements about particular cases or kinds of cases. Certain symptoms or conditions seem to be paradigmatic of bio-medical disorder, such as fever, HIV, cancer, and broken legs. Similarly there seem to be paradigmatic symptoms or conditions for psychiatric disorder, such as psychosis, schizophrenia, and mania. While we have (often strongly held) intuitions about paradigmatic cases we also have (often strongly held) intuitions about cases that are clearly not bio-medical disorders or mental disorders. There are many more cases that we are unsure about, however. Anti-psychiatrists seem to have the most divergent intuitions when they maintain that there aren't any mental disorders. I will go on to maintain that this claim isn't really driven by their having different intuitions about paradigmatic cases, however. Similarly, in the philosophy of mind eliminativism about mental states isn't really driven by a radically different set of intuitions about what sorts of phenomena are paradigmatically mental. I'll return to the issue of eliminativism.

8.2.2 Bridge features

The second kind consists in our intuitive judgement that the paradigmatic cases have something in common in virtue of which they really are members of the same category. When we consider that the paradigmatic cases really are instances of some category we seem to think that they have some relevant feature in common in virtue of which they really are instances of that category. This second step is preparatory for the development of a theory as to what the relevant feature is. It puts constraints on the sorts of features that are allowable candidates for determining category membership. One way to see the relevance and importance of bridge features is to consider two- dimensional semantic analyses of concepts. A fairly standard two-dimensional analysis of 'water' is to distinguish between a-priori features on the one hand, and a-posteriori features on the other. A-priori features include that 'water' is the drinkable, potable liquid in our environment that falls from the lakes and fills the skies, etc. A-posteriori features include that scientists have in-

vestigated that liquid and found that it has the chemical composition H_2O.

One bridge feature that we could adopt is to consider 'water' by way of the sortal 'natural kind of chemical'. On this analysis the essential properties of water are picked out by the natural kind of chemical in the vicinity - in this case H_2O. Another bridge feature that we could adopt, however, is to consider 'water' by way of the sortal 'the functional role that it plays in our daily lives'. On this analysis the essential properties of water are picked out by the functional role in the vicinity - in this case the drinkable, potable liquid that falls from the lakes and fills the skies, etc. It might be tempting to consider that the relevant sortal for 'water' is the natural kind of chemical sortal and the role that the notion plays in our social lives isn't relevant for picking out the necessary and sufficient features. This is controversial, however. There has been criticism of the identification of water with H_2O on the grounds that the impurities are important to us such that if you asked a person for a glass of water and they gave you a glass of H_2O then your request would not have been met. Similarly, if one was handed a glass of black tarry stuff that had chemical composition H_2O then while the chemical kind sortal would have it that your request had been met the social role sortal would conclude (rather intuitively) that your request had not been met.

We might be able to explain these intuitions by appealing to pragmatic features of language. Even so, it seems that one virtue of the two dimensional semantics approach is that it is able to offer an account of both sets of intuitions. Once one has analyzed the concept into its social role and into its chemical kind then what more is there to be said about the concept? The issue of which sortal is the 'correct' sortal for a concept seems to amount to the issue of which sortal provides the necessary and sufficient conditions for the concept. In the case of 'water' a lot of theorists have the intuition that science is the place to look for necessary and sufficient conditions. In the case of mental states things are more problematic as some theorists maintain the functional role provides the necessary and sufficient conditions while others maintain that we should take the realizers of the role to provide the necessary and sufficient conditions. In the case of money people often have social

role intuitions and the realizers of the role are thought to be irrelevant with respect to the necessary and sufficient conditions for something counting as money. In the case of mental disorder there is variation in peoples intuitions, as we shall see.

While two dimensional semantics provides two different ways that we might go I think that with regards to mental disorder in particular we are better to think of there being more ways than merely two. While a-posteriori features and a-priori features are going to capture all of the different ways in order to capture the dispute over the concept of disorder we are going to have to consider further distinctions between different kinds of a-priori and a-posteriori features. It also might not be terribly clear in some instances whether the features are best regarded as a-priori or a-posteriori.

While we considered that one conceptual analytic project (the 'super-rationalist' project) was attempting to explicate the implications of the first sort of feature and perhaps revising some of them if they turned out to conflict with other features of the first sort this is not the usual way of explicating the conceptual analytic project. The usual way of characterizing it is to consider that the first sort of features and the second sorts of features trade off against one another in a process of reflective equilibrium. This usual way of characterizing the conceptual analytic project thus seems to consider there to be two different kinds of features.

If the first and second kind of features are found to come apart then there are two different ways that we might go. One way is to prioritize the first kind of feature over the second kind (the 'super-rationalist' project), the other way is to prioritize the second kind over the first kind (which I'll call the 'rationalist' project). In the case of knowledge the justified true belief (JTB) account of knowledge was shown to come apart from our judgment of cases with respect to Gettier cases. The majority of theorists take the rationalist line in maintaining that this shows our analysis of knowledge to be deficient. In the case of moral theory some super-rationalist utilitarians are prepared to bite the bullet and say that insofar as our intuitions about the right thing to do diverge from utilitarian theory we should revise our intuitions about

the right thing to do.

8.2.3 A-Priori descriptive features

These are fairly hard to characterize. Examples of these kinds of features in the philosophy of mind would include such platitudes as 'a person will act so as to meet their desires on the assumption that their beliefs are true'. In the case of mental disorder such features might include 'there is something wrong with people who are disordered' and 'people who are disordered would be better off if they weren't'. In the case of knowledge the justification, truth, and belief conditions might similarly be thought to be examples of these kinds of features.

One thing that is clear is that these kinds of features might well be non-obvious. If we were to ask people about their concept of belief or of knowledge or of disorder then they might well not list the sorts of features that I have listed here. I suppose that one could say that these kinds of features should appear obvious once they are mentioned to people, however. Another thing that is clear is that it might turn out that the kinds of platitudes that we list here have implications (logical entailments) that people simply haven't thought through. This might be one way that we get to the sorts of conditions I have listed here. One thing that is interesting is that insofar as people haven't thought through the implications of these kinds of features it might be the case that thinking through the implications leads us to contradiction. One project would thus to be revise the least intuitive of these features in order to render them consistent. I'll call this project the 'super-rationalist' project and I'll have more to say about revisability (and how we should go about doing it) later.

While these features are often thought to be definitionally true or true a-priori this way of putting things is problematic. It is problematic partly because some of these features might be merely associated rather than being definitionally true. There is controversy over whether a state can be a state of belief if it does not play the functional role of belief, for example. This

controversy seems to amount to whether the functional role of belief provides the necessary and sufficient conditions for something being a belief or whether the functional role of belief is merely associated with belief and is neither necessary nor sufficient. Another problem is that if an analysis of our concept reveals that we are committed to contradiction then this would force us to eliminativism on a-priori grounds. A contradictory concept cannot be instantiated. We can of course attempt to revise our conception in the face of contradiction and if we are trying to get at which concept we *should* adopt then this might be a way out. I'll return to issues of revisability later.

8.2.4 A-Posteriori descriptive features

A-posteriori features are discovered a-posteriori most often by scientists. Standard examples of a-posteriori features are H_2O in the case of water. Scientists discover a (perhaps imperfect) correlation between whatever is picked out by ostension and / or by description and some underlying feature. While these a-posteriori features are often taken to be necessary (and it is thought that scientists don't discover a correlation they discover an identity) it seems that even if there is a perfect correlation there is scope for resisting an identification by maintaining that the concept is a social role concept (for example) rather than a natural kind concept. With respect to disorder even if theorists share the sortal 'scientific kind' there could be controversy over what kind of scientific kind (e.g., genetic, neurological, cognitive, behavioural, social etc) and even if theorists share the same kind of kind they could disagree about the particular neurological correlates that are relevant (for example). So there can be different theories of the neurological basis of schizophrenia as some theorists try and identify schizophrenia with enlarged ventricles and others try and identify schizophrenia with deficiencies in the dopamine system etc.

8.2.5 Social role

Sally Haslanger (find reference) distinguishes between the role that a concept explicitly plays in theory from the way that a concept may (implicitly) guide

our social practices. So it might be that a teacher takes the roll half an hour later on Monday and so the working definition might be that you are late if you arrive for class after 9am on Tuesday to Friday or if you arrive after 9.30am on Monday. The role that a concept plays in structuring out social practices might well be non-obvious and it might well be that an account of the role that the concept actually has played in guiding our behaviour and structuring our institutions is a non-obvious a-posteriori feature of the concept.

Theorists who maintain that empirical correlates should be prioritized over our intuitions about other kinds of features (if they are found to diverge) I'll call 'super-empirical'. Theorists who maintain that there is some kind of process of reciprocal illumination between the sortals and the empirical correlates (so we can revise what sortals are relevant to latch onto the best correlates) I'll call 'empirical'. (This really does need work. I will have to say something about the project of rigidifying on the correlates 'around here' as in Lewis' account of pain. Or the idea of restricted identities as in I can't remember who's account of pain in dolphins and pain in humans etc. These might both come under the empirical project as I've characterized it. Not sure about this yet.)

8.2.6 Summary

So, on the way I've told the story thus far firstly, we start with our intuitive judgements of cases. Secondly, we apply a sortal which tells us where to look in order to find the necessary and sufficient conditions. This second kind of feature might direct us to attempting to characterize the necessary and sufficient conditions a-priori (as in knowledge, or as in the morally right acts) or it might direct us to attempting to characterize the necessary and sufficient conditions a-posteriori (as in gold, water etc). The last step is thus to either figure out the necessary and sufficient conditions a-priori, or to discover the necessary and sufficient conditions a-posteriori. I've also considered that this project is constrained by what it is that we want to use the concept to do and there can be both an a-priori analysis of the role that a concept plays

in structuring theory and an a-posteriori analysis of the role that a concept plays in structuring our social institutions.

Another question that we could ask, however, is what features are driving our intuitive judgements. If we characterize these features a-priori then this seems part and parcel of the a-priori project. We are justified in judging a certain way if we think that certain conditions obtain. An empirical counterpart to this project would be to attempt to do a feature analysis on the kinds of features that actually are driving our judgements. This is relevant with respect to how much the DSM feature lists provide the rationale for diagnosis or how much clinicians judgements are driven by other features that aren't listed in the DSM.

One of the major insight of two dimensional semantics is the thought that there are two sortals (not sure that is the best way to put this) that can come apart and hence two quite different things (though equally legitimate things) that we can be tracking. One way we can go in the case of water is to identify water with the functional role (the a-priori description) rather than with the natural kind of substance (to be discovered a-posteriori). Two dimensional semantics gives us the resources to analyze both aspects of meaning without prioritizing any one of those. I always thought that the 2D stuff committed to natural kind of chemical stuff taking priority over the social role. JC said that he didn't think that was so, however. He thought that it just provides an account of both ways we could go without prioritizing either one. Perhaps that is what I'm really wanting to do in this chapter. Provide an account of the different ways we can go on disorder without my committing to prioritizing one way over the other. Offering an account of the different ways (in 2D semantics and here) seems to be useful for cutting through verbal dispute, however. I think that in a way that is my major concern here. There (seems to me at least) to be so much verbal dispute over disorder and we need to cut through that in order to get to the substantive.

8.3 Diagnostic tools

I now want to introduce a couple of tools that will hopefully help us make some sense of the disagreement between different theorists[4]. If we move away from the idea of essences / necessary and sufficient features and consider more of a cluster concept idea (where there are more or less central features that render phenomena closer or further away from paradigmatic cases) then the following tools may be of use.

8.3.1 Conditionalizing features

Let us suppose that there is an individual S such that S lists the features I enumerated earlier when they reflect on their concept of bio-medical disorder. We could say that according to S:

- An individual x is disordered iff $\exists(x)$ $(x_a, x_b, x_c, x_d, x_e, x_f, x_g, x_h)$

The above has been arrived at by taking the features that S listed, joining them with conjunction, and then taking them to be definitional of the concept. One problem that we might have with this strategy is that it might make eliminativism too easy, however. If it were found that there wasn't an x such that features a-h obtained then one would seem to have found that there aren't any bio-medical disorders. We would have eliminativism about disorder on empirical grounds. Alternatively, if there was a contradiction in the features then we wouldn't even have to look to the world to know that nothing in the world met the definition. We would have eliminativism about disorder on rational grounds. While some theorists have been eliminativists about mental disorders in particular, I don't know of a theorist who has been an eliminativist about the bio-medical notion of disorder more generally.

An alternative strategy - one that I wish to advocate - is that instead of taking the feature list to be *definitional* one takes it to be *conditional*, as follows:

[4]I am grateful to Wolfgang Swartz for introducing me to the notion of the Carnap Conditional and his characterization of the Canberra plan.

- If there is an individual x such that $(x_a, x_b, x_c, x_d, x_e, x_f, x_g, x_h)$ then x is mentally disordered.

Why do things this way? It is a way of remaining true to the features that S listed (if there is something that meets all of them then that is disorder) but it is also a way of taking seriously the suggestion that we can revise features that we take to be relevant. The conditional doesn't logically entail or say anything at all about what we should do were we to find that nothing met all of the features (either on empirical or on rationalist grounds). This allows us to revise one or more of the features.

8.3.2 Weighting features

One way theorists would seem to be able to disagree is by different theorists maintaining that different features are part of the conditional definition of disorder. Lets consider theorists S, T, U, and V with the following feature lists.

- According to S if there is an individual x such that (x_a, x_b, x_c, x_d) then x is mentally disordered.
- According to T if there is an individual x such that (x_a, x_b, x_c, x_e) then x is mentally disordered.
- According to U if there is an individual x such that (x_a, x_b, x_f, x_g) then x is mentally disordered.
- According to V if there is an individual x such that (x_a, x_f, x_h, x_i) then x is mentally disordered.

One might wonder whether the features listed by S and T are similar enough such that S and T have the same concept and genuinely disagree about it. One might similarly wonder whether the features listed by U have enough in common with the features listed by S and T such that U has the same concept and hence there is genuine disagreement. The features listed by V don't seem to have much in common with the features listed by S and T so one might wonder whether V has a different concept. There is as much in

common between S and T, and U as there is between U and V, however, so same though related concept vs. different concept gets tricky. It does seem that we intuitively want to say that these different theorists are genuinely disagreeing, however.

One thing that we can do is to assign relative weights to features. Let 1 stand for most revisable (least central) and let 4 stand for least revisable (most central) and let 0 stand for not relevant. We can thus characterize the views of the above theorists as follows:

- According to S if there is an individual x such that $(4x_a, 3x_b, 2x_c, 1x_d, 0x_e, 0x_f, 0x_g, 0x_h, 0x_i)$ then x is mentally disordered.
- According to T if there is an individual x such that $(4x_a, 3x_b, 1x_c, 0x_d, 2x_e, 0x_f, 0x_g, 0x_h, 0x_i)$ then x is mentally disordered.
- According to U if there is an individual x such that $(2x_a, 4x_b, 0x_c, 0x_d, 0x_e, 3x_f, 1x_g, 0x_h, 0x_i)$ then x is mentally disordered.
- According to V if there is an individual x such that $(4x_a, 0x_b, 0x_c, 0x_d, 0x_e, 1x_f, 0x_g, 2x_h, 3x_i)$ then x is mentally disordered.

While we can provide relevant weights for the different features, we can also consider that there are different *kinds* of features as I considered before. While different theorists could disagree insofar as they weight features differently it seems to me that different theorists could be more radically disagreeing (or perhaps even engaging in different projects) insofar as they weight the different kinds of features differently. So... Four projects: Super-rationalist, rationalist, empirical, super-empirical. The super-rationalists or the super-rationalist-rationalists seem to have the most trouble engaging with the empirical and super-empirical theorists. They often do seem to be talking past each other. Sometimes the claim is 'category error' (when the a-priori should be prioritized) and othertimes the claim is that we don't care about the concept we care about the nature and so we need to look to the world. Both of those extremes seem to me to be wrong headed. We need the sortal and that is a-priori. But then... Are we trying to get at what we have to say

about our beliefs or at the nature of the phenomenon.

8.3.3 Revising features according to conditional weight

We can now see that two theorists could agree as to what features should appear on the feature list and yet disagree as to the relative weighting of the features. How does this matter? If it turns out that there is no such thing in the world that meets the conditional definition (either because of rationalist constraints on consistency or because of empirical constraints from the world) then the theorists could disagree as to whether they should accept the eliminativist conclusion or whether they should revise their concept (or revise the feature list that is attempting to capture the nature of the concept). Sometimes we might have to choose between two different conditions as we can't retain both. In this case theorists could maintain that we keep one and give up the other because they prioritize one over the other.

8.4 Diagnosing the dispute

Can be disagreement within an approach and can be disagreement between approaches. There is a temptation to see theorists who prioritize different kinds of intuitions as talking past one another. Not trying to do the same thing, or not having the same concept. If we carve off the two extreme ends of the pole then things seem very weird, however. Yes, we are interested in a science of disorder. But the reason why disorder is so important to us is because we think that people are better off without disorder and we want to treat / help them. People often say there isn't a straightforward implication... But that is because there does seem to be a bit of a divide between the subject of science and the subject of intervention. Important to be clear about this, however, and not run things together.

I'll consider the ethics vs science division of labour in the next chapter. I'll also look at the attempt to ground the objective aspect of disorder by way of the dysfunction criterion (where dysfunction might be given the above kind of analysis as disorder has been and as mental could be).

I now want to run through the different kind of features once more in the case of mental disorder in order to locate some of the dispute. While some of the dispute is empirical some of the dispute is conceptual while other aspects of it are purely verbal. I think it is important to dispense with the verbal dispute in particular in order to make progress on these issues.

8.4.1 A-Priori features of disorder

In the case of mental disorder Wakefield maintains that when our intuitions about who is and who is not disordered diverge from our intuition that 'something is wrong with people who are disordered' then we should super-rationally revise our intuitions about who is and who is not disordered[5]. Murphy maintains, however, that when our intuitions about who is and who is not disordered diverge from our intuition that 'something is wrong with people who are disordered' then we should rationally revise our intuitions that there is 'something wrong with people who are disordered'.

The Harmful Dysfunction view is (on some readings at least) an account that starts with the non-revisable a-priori feature that 'something is wrong with people who are disordered' and progresses from there. When Wakefield's view makes predictions that comes apart from our judgement of cases he recommends that we revise our intuitions about our judgement of cases in order to retain the intuition that 'something is wrong'. I will have much more to say about the two stage view in the next chapter.

While this might initially seem surprising one take on the anti-psychiatry critique is that they deny that mental disorders are disorders on conceptual grounds. They maintain that (a-priori) the mind is not the brain and that (a-priori) disorder is a bodily state. They thus conclude that mental disorders can't be disorders and it would be a category error to think otherwise. We can see that they also have an a-priori view of the nature of disorder and

[5]Or at least Murphy would be sympathetic to this way of characterizing Wakefield's view. Wakefield, however, maintains that 'something being wrong' can be identified with natural properties and so characterizing his view is complicated. I will deal with his view at length in the next chapter, however.

they prioritize this intuition above our intuitions of our judgments of cases.

Issues of moral responsibility / diminished personhood. This seems conceptual too (not terribly sure what to say about this).

8.4.2 Judgment of cases of disorder

People also have intuitions about paradigmatic cases of disorder. HIV, cancer, infestation by parasites etc. In the case of mental disorder also, people have intuitions about paradigmatic cases. Schizophrenia, bi-polar, and depression are offered by Murphy as paradigmatic examples. We can also regard certain symptoms such as mania, delusion, and psychosis to be paradigmatic (as I'll return to the issue of natural kinds later). On some accounts paradigmatic examples of mental disorder are regarded as paradigmatic instances of medical disorders (on the bio-medical model, for example). On other accounts they are not (on anti-psychiatry views, for instance).

When a-priori features are rejected as being candidates for capturing the content on the grounds that the a-priori features do not match our intuitive judgment of cases then our judgment of cases is being prioritized over our a-priori intuitions. Conversely, when theorists (like Wakefield) maintain that we should revise our judgment of cases in favor of the a-priori intuition that 'something is wrong' then he prioritizes a-priori intuitions over our intuitive judgment of cases.

Likeness arguments. Different intuitions about paradigmatic cases.

8.4.3 Bridge features for disorder

There seem to be two different literatures on disorder that don't seem to come into contact. One of them views disorder as being a functional role notion where disorders are to be differentiated on the basis of a certain aetiology, list of manifest symptoms, and the evolution of those symptoms over time. The other views disorder as being a natural kind notion where disorders are to be identified with the causal mechanism that generates the behavioural

symptoms.

This way of putting things seems to suggest that a two dimensional semantics analysis could be of use. Similarly to how we can offer an analysis of subjective temperature (whether something feels hot or cold) and objective temperature (determined by thermometer reading) without prioritizing one of those with respect to what temperature really is we might be able to offer an account of the functional role and the causal mechanisms that produce disorder without prioritizing one of those notions with respect to determining what disorder really is.

8.4.4 A-Posteriori features

A great deal of the controversy has been over what sorts of causal mechanisms are allowed to count, however. Anti-psychiatrists, for example, sometimes maintain that 'there aren't any mental disorders' and then go on to offer a social role analysis of the causal mechanisms that produce the behavioural symptoms. In order to understand this claim we need to understand them as denying that social roles are the right kind of thing to count as disorders. Similarly, debate between biological reductionists (who maintain that mental disorders are neurological and / or genetic) and theorists who maintain that mental disorders are cognitive and / or social seems to hinge not just on empirical issues to do with causal mechanisms but also on conceptual issues to do with what kinds of mechanisms are allowed. The turf wars. Depth vs spread.

The debate between different theorists might be analysed into the different relative weightings that are assigned to the different properties and disagreements on facts about the world. For example, Murphy gives most relative weight to the intuition that central paradigmatic cases are mentally ill. Part of the motivation for this comes from his acceptance of the Causal Historical theory of reference fixing where we start with the central exemplars and form our concept in order to identity others that are similar in some respect. The intuition that Murphy gives the highest weight to is that certain paradig-

matic cases of psychosis, depression, and mania fix the reference of our term mentally ill and the business of science is to empirically investigate what (if anything) these people have in common. (interesting he doesn't think there is a tenable neurology / psychiatry which is to say mental / non mental distinction.. he doesn't think MENTAL disorders have anything in common. He does accept the HD account (though revisable in principle). He does accept that there are kinds of disorder??? Revisable however.

Wakefield, on the other hand, seems to prioritise his intuition that people with mental disorder are in fact disordered or malfunctioning. If we were to find out that paradigmatic cases of people with depression, psychosis, and mania were in fact not malfunctioning then that would amount to a discovery that they are not mentally disordered. He also prioritises INNER malfunction over behavioural malfunction so as to capture the intuitive distinction between problems in living and the like We could retain the inner causal mechanism assumption (As Murphy does even though he acknowledges that it may be false and is revisable in principle) above the malfunction assumption. Not revisable for Wakefield Seems to make malfunction a-priori.

Anti-psychiatry people can also be regarded as prioritising inner causal mechanisms and natural kind assumption then denying that there is anything in the world that plays the role. We can avoid this, however, if the best candidates we have allow us to revise these assumptions. It is a way of mapping their position, however. Go modal and check intuitions. It was supposed to be surprising in the case of water and even gold. Pain. Counter-intuitive yet it fell out of the theory. Insofar as it seems intuitive that there can be a planet with water - but can there really? How about mice with schizophrenia? Role vs Rigid Designators.

Conceptual analysis:

- enumeration of the intuitive judgements we make about whether an individual is or is not mentally disordered (survey) - enumeration of the intuitions we have about the content of our concept - what do the folk / specialists say is necessary vs contingent (survey) - elucidation of our concept for consistency

(elucidation, philosophers?)

Empirical investigation:

- sample initially fixed by the intuitive judgements we make about whether an individual is or is not mentally disordered (depression, psychosis, mania) - investigate what (if anything) these people have in common - note that the intuitions that we have about the concept comes into play here with respect to whether we are looking for behavioural, mind / brain, or socio- logical features

- empirical findings can feed into the concept (change our intuitive judgements)

- thus altering the judgements that we make about the sample / prototypic cases

our judgements are revisable in principle but they do in fact play an important role in our identifying the individuals who are the subject of investigation. Once we have identified the individuals then we attempt to find what they have in common, however. Sometimes we find that there are individuals who didn't initially fall under our concept / who we didn't previously identify. If they share the relevant properties in common we may decide to revise our concept such that we do identify them as being members of the same class. Sometimes we find that there are different groups We might thus come up with new concepts that may replace the old or not Sometimes we struggle to find what (if anything) they have in common.

Competing intuitions: might compete on internal consistency grounds or on external consistency grounds (there might not be anything in the world with those features).

When we revise our concept or when we say there is no such thing is a matter of which is most counter-intuitive to us. Really don't want to get caught up in verbal disputes over whether something exists or not when people agree as to the state of the world but disagree over whether we should revise our concept or throw it away In some ways It is a conceptual decision as to

what we do with our concept But really Who cares what we do with our concept? Don't we really care about what mental illness actually is? To say this does of course prioritise the intuition that certain individuals are mentally disordered. Not necessarily all of them There is room for dispute over how many of them But most of them, perhaps Others might disagree with this intuition. But we have a way of understanding where the dispute lies: in the weighting of intuitions.

There are of course other considerations. Whether people get treatment, whether health insurance should pay etc even the eliminativists need to say something about why these people act differently Usually appeal to social factors

So addiction and sociopathy how many features do they share with other, more paradigmatic mental disorders?

LIKENESS ARGUMENT (need to weight features)

I want to begin by talking a little about the kinds of things that we might be doing when we are doing conceptual analysis on the concept of mental disorder. One thing that we might be doing when we attempt to analyse a concept is to find out what people take the content of the concept to be[6]. We might survey the folk, for example, to see whether they regard certain cases to be instances of mental disorder or not and / or we might ask the folk what features or properties they take to be most central to their concept (e.g., whether they take inner malfunction to be necessary). This process might enable us to get at the folk concept of mental disorder. Instead of trying to get at the folk concept one might attempt to try and get at the specialist concept by surveying specialists. While some people have conducted surveys to attempt to get at the concept (or our beliefs about it) this doesn't seem to

[6]There is an issue with respect to how much the content of our concepts is transparent to us such that we are able to report the content. It might be that surveys assess our beliefs about the content rather than the content if, for example, content is broad. I shall have more to say about the issue of broad content later. At present, however, I don't think the distinction matters particularly for the way I characterize the survey project. Where I say 'content the reader may substitute 'beliefs about the content if this is preferred.

be what the majority of people take themselves to be doing in their attempts to define mental disorder.

Often it is said that rather than getting at what people say the content is conceptual analysis is about elucidating our concept such that it is consistent. If we find that we endorse contradictory beliefs about the concept then it would seem that we should revise some of our intuitions in order to have a set of consistent beliefs about the concept. Similarly, if we find that our beliefs about the central features of the concept are inconsistent with our judgements as to whether someone is mentally disordered or not then it would seem that we should revise some of our intuitions either about the concepts or about the cases or both in order to retain consistency. On this project it would seem that conceptual analysis would have more to do with elucidating our concept such that it is consistent rather than a matter of discovering what in fact we take our concept to be as in the survey project. While it may be useful to know what the folk and / or the specialists take the concept to be on this project surveys are merely viewed as a starting point for a systematisation of our intuitions. Often people proceed with this second project of conceptual elucidation and clarification without conducting a survey of the folk or the relevant specialists first. One can attempt to clarify for oneself though of course whether others are likely to be persuaded by ones analysis depends greatly on how much they share ones intuitions. When our intuitions conflict then we are motivated to revise them in such a way as to achieve consistency and the notion is that some features or cases are more intuitive than others and we should make the least counter-intuitive revisions in order to achieve consistency.

Related to the above project of elucidating our concept in a way that it is both internally consistent to our intuitions about the concept and consistent to our judgement of cases there is the project of elucidating our concept in such a way that we are better able to identify individuals who are mentally disordered. The American Psychiatric Association seems to have this in mind as it offered a definition of mental disorder initially in response to political pressure for them to justify why they regarded some conditions (e.g., homo-

sexuality) to be mental disorders. The APA maintained that the definition of mental disorder provided in the DSM was used to help determine whether a condition should feature in the DSM. The definition of mental disorder provided in the DSM is also meant to help clinicians identify individuals who are mentally disordered. Spitzer and Endicott (the main movers behind the DSM definition) attempted to operationally define the concept of disorder in such a way that clinicians had a criterion that they could use to identify whether individuals were or were not mentally disordered. Wakefield has responded to their attempted operational definition (as he responds to the DSM definition) by maintaining that when it successfully captures our intuitions that is because it falls into line with his HD account and when it does not successfully capture our intuitions that is because it diverges from his HD account. I shan't consider this debate in much more detail. The crucial thing to note, however, is that the DSM seems to view the project of defining mental disorder to be bound up with our being able to better identify both conditions and individuals who are and who are not mentally disordered.

It is worth drawing a distinction that is not often drawn in the literature on mental disorder between our concept of mental disorder on the one hand, and the nature of mental disorder on the other. This distinction is often drawn in the philosophical literature, especially with respect to natural kind terms, and the notion is that features that are central to our concept and / or the features that enable us to identify the referent can be quite different from the features that are essential to the instances being members of the kind . For example, our concept of water includes such features as its liquidity and transparency and potability. While these features are central to our concept of water they are not thought to define whether or not a sample is a sample of water. What is essential to a samples being water is that it is composed of H_2O. While being composed of H_2O never used to be a reported feature of our concept of water it turns out that being composed of H2O is essential for the samples being a sample of water whereas the features that were central to concept (liquidity, transparency, potability) are merely accidental features of the category. One might similarly expect that the nature of mental disorder

(to be discovered by science) may comes apart from the features that people report to be central to the concept and / or that people use to identify people who are mentally disordered.

The debate between different theorists might be analysed into the different relative weightings that are assigned to the different properties and disagreements on facts about the world. Murphy gives most relative weight to the intuition that central paradigmatic cases are men- tally ill. Part of the motivation for this comes from his acceptance of the Causal Historical theory of reference fixing where we start with the central exemplars and form our concept in order to identity others that are similar in some respect. The intuition that Murphy gives the highest weight to is that certain paradigmatic cases of psychosis, depression, and mania fix the reference of our term mentally ill and the business of science is to empirically investigate what (if anything) these people have in common. (interesting he doesn't think there is a tenable neurology / psychiatry which is to say mental / non mental distinction.. he doesn't think *mental* disorders have anything in common. He does accept the HD account (though revisable in principle). He does accept that there are kinds of disorder? Revisable however.

Wakefield, on the other hand, seems to prioritise his intuition that people with mental disorder are in fact disordered or malfunctioning. If we were to find out that paradigmatic cases of people with depression, psychosis, and mania were in fact not malfunctioning then that would amount to a discovery that they are not mentally disordered. He also prioritises *inner* malfunction over behavioural malfunction so as to capture the intuitive distinction between problems in living and the like We could retain the inner causal mechanism assumption (As Murphy does even though he acknowledges that it may be false and is revisable in principle) above the malfunction assumption. Not revisable for Wakefield Seems to make malfunction a-priori. Anti-psychiatry people can also be regarded as prioritising inner causal mechanisms and natural kind assumption then denying that there is anything in the world that plays the role. We can avoid this, however, if the best candidates we have allow us to revise these assumptions. It is a way of mapping their position,

however.

Go modal now and check intuitions Role or Rigid Designators?

The point is that there is a back and forth process at work, however. More in particular there is a back and forth process with respect to our clarifying our concept and revising our intuitions and there is also a back and forth process between the features that are identified in our conceptual analysis and the features that are identified in kinds that are in the vicinity by the relevant sciences. At this stage the only thing that people with disorder x seem to share is the behavioural symptom criteria it would be terrific if our kinds of disorder and the notion of disorder in general picked out some distinction in nature but it seems as though we are still searching for the relevant distinction.

Or as easy as Kraeplin thought it would be.

Issue around whether we rigidify over the inner causes / malfunctions or whether we take the cluster concept / functional role to be it. We don't know as yet what will happen with mental disorders. We simply don't know whether mental disorders will turn out to be the results of inner malfunctions or not. Murphy is right to say that we don't want to build it in a-priori. The inner malfunction assumption seems to be revisable in principle. It might well be the case that it is fairly central but it seems too strong to say that it is not revisable or that it is a-priori. If logic can be revisable in principle (though fairly resistant to be sure) then more so our notion that mental disorders are the result of malfunction!

My last objection to Wakefield is something that I shall just touch on briefly though it is something that I want to develop. The objection is that he commits himself to too much a-priori when he maintains that malfunction is necessary for mental disorder (and in particular when he identifies causal-historical processes as being necessary for mental disorder). A lot of theorists have attempted to construct counter-examples where we intuitively regard the person to have a mental disorder and yet where it is stipulated that they do not have an inner malfunction. Wakefield then responds to these

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

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

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

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

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

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

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

While survival and reproduction are fairly obvious standards for fixing functions if we are interested in evolutionary biology it does seem that our explanatory interests play a crucial role in allowing us to get normativity from the notion of function and malfunction. With respect to medicine there is widespread agreement that disorders that threaten a persons survival are disorders and the reasonableness of this view seems to be inherited from the reasonableness of survival as something that we are interested in promoting. Despite widespread agreement that some conditions are disorders there are controversial cases in medicine, however. There are some conditions where it

is unclear whether they are disorders or mere problems in living or whether surgery is a medical requirement or merely elective. The further away one gets from issues of survival and the more expensive the treatment the more controversy there is as to the status of the individual or the condition as appropriately being regarded as disordered. Psychiatry doesn't seem to be concerned with survival of persons in quite the way that biology and medicine are. So in my talk today my conclusion is largely negative. The claim is that the malfunction assumption cant do the work that is required of it. The biological notion of function cant ground psychiatry in facts about purely causal processes because the biological notion of function requires us to identify survival and reproduction as the relevant features for fixing functions. Given the explanatory interests of evolutionary biology this is a reasonable thing to do and I have no problem with the scientific status of biology. It is unclear, however, that survival and reproduction are the relevant features for fixing functions and malfunctions for psychiatry. I think that much more work needs to be done on the harm component with respect to understanding what disorders have in common.

Coopers has stated that mental disorders might be a little like the notion of weeds. That is to say that our values might well be crucial for determining the class of things that we are interested in. While there is no objective science of weeds because weeds don't have non- evaluative properties in common that differentiate them from non-weeds there can still be a scientific classification of plants, however. It could similarly be the case that individuals with certain kinds of mental disorders share certain causal processes in common though our values are an important part of how we distinguish the mentally disordered from the non-mentally disordered. While the anti-psychiatrists maintain that the relevant values are social or moral it is not the case that any social or moral norm violation is indicative of mental disorder. More work needs to be done on what kinds of norms are relevant. The notion of malfunction isn't very explanatory if it is merely an assumption that we have built into our model where we could re-describe the casual processes that we have discovered as dif-functions instead of dysfunctions.

8.5 The problem of conflicting intuitions

There seems to be three main intuitions that we have about the concept of disorder and that come into play in Wakefield's argument. The first intuition (or set of intuitions) are intuitions that we have about who and who is not mentally disordered. These intuitions seem comparable to intuitions that we have about what substances are or are not watery substances. In the same way that samples of watery substance fix the reference of water samples of individuals who we regarded to be mentally disordered fix the reference of mental disorder. In the case of water we have the watery stuff that fills the lakes, oceans, and rivers, and that falls from the sky. In the case of mental disorder we have the people who display floridly psychotic symptoms such as delusion and hallucination and grossly inappropriate affect. Similarly to how we would be very resistant indeed to accept the conclusion those samples aren't samples of water after all we would be very resistant indeed to accept the conclusion those people aren't mentally disordered after all.

The second intuition is the natural kind assumption. In the case of water we assume that the samples have something in common that determines whether or not they are of the same category as the instances in the sample. Similarly, in the case of mental disorder we assume that the people have something in common that determines whether or not they are mentally disordered. The assumption here is that there is something in common that the samples share. The samples are not a nominal category where the only property they have in common is that we regard them to be instances of the same category. Rather, different individuals with mental disorder share some relevant property in common that determines that they are in fact mentally disordered. The natural kind assumption might be thought to go further than this in stipulating that the relevant common properties are not to be found among the properties that we used to initially identify instances as being of the same kind or not; rather the relevant common properties are underlying properties that are responsible for generating the superficial properties.

The third intuition is that the relevant property that the individuals have

in common (that determines whether or not they are mentally disordered) is that they have an inner malfunction. Wakefield provides a number of cases to support this intuition and to attempt to show us that this is an intuition that (all things considered) other people share. He offers the example of someone who meets criteria for a reading disorder. The notion is that intuitively we regard this person to be mentally disordered on the basis of their having behavioural (or superficial) properties that typically lead us to regard someone as mentally disordered. He then maintains that it makes a difference to our intuitions whether they have the behavioural symptoms (or exhibit the superficial properties) in virtue of inner malfunction or in virtue of environmental circumstances (such as nobody ever having tried to teach them to read). He maintains that in this case we would regard the person with inner malfunction to be mentally disordered whereas we would not regard the person who had never been taught to be mentally disordered. This is supposed to show us that our intuitions about cases are in line with our intuitions about whether there is an inner malfunction or not.

Now what I want to do is to push this example still further. What would our intuitions be if we found out that some section of the people who we typically regard as being mentally disordered turned out not to have inner malfunction? In the above case Wakefield is asking us to assume that the majority of people who meet the behavioural criteria for reading disorder do have inner malfunction and it is just the odd case that doesn't. If we extend the thought experiment, however, such that every single person who meets behavioural criteria for reading disorder does not have inner malfunction then do we conclude that reading disorder is not a disorder after all, or do we conclude that mental disorder does not require there to be inner malfunction? I'm not too sure on my intuitions in this extended example, so Ill extend the ex- ample still further. If we found that every single person who was considered to be paradigmatic of mental disorder did not have an inner malfunction then would we conclude that there was no such thing as mental disorder or would we conclude that of course they are mentally disordered and therefore inner malfunction is not necessary for mental disorder after all?

Wakefield's intuition seems to be that we prioritise the inner malfunction assumption over the paradigmatic cases assumption. He is explicit about this when he maintains that since his HD analysis of the concept of mental disorder is correct it turns out that the DSM is over-inclusive with respect to regarding people who meet behavioural criteria to be mentally disordered. If one prioritises the DSMs (the experts) judgement of cases over the inner malfunction assumption, however, then one would be led to conclude that Wakefield's inner malfunction assumption was incorrect as an analysis of our concept of mental disorder.

[This essentialist definition uses the prototype properties not as universal criteria for the construct but only to indirectly refer to its essence. Thus the definition allows things very different from the prototype set, such as ice, steam, or H_2O atoms floating in space, to be water (Wakefield, 2004, p 79).

On this criterion it makes it sound as though ALL the behavioural features could vary (they might not meet criteria) and yet they would have mental disorder in virtue of the essence. Twin earth Harm might be saving him here.

The natural kind assumption could conflict with both the judgement of cases intuitions and with the inner malfunction intuition. The natural kind assumption is that the individuals share some underlying property in common that is responsible for generating the behavioural symptoms. It seems to be the natural kind assumption that drives Wakefield to maintain that the relevant properties that determine whether an individual is or is not mentally disordered are internal to the person and are to be determined by science. It is not a matter of a-priori conceptual analysis what property the individuals share, rather that is a matter for scientific investigation. The trouble with this assumption is that it could turn out (for all we know) that mental disorder is not a natural kind and that as such the paradigmatic cases don't have any internal property in common that is responsible for generating their symptoms. While we can build a natural kind assumption into our concept and also a sortal assumption such as substance or process we cannot know a-priori whether we are dealing with one assumption or one process. We cant rule out that at the end of scientific inquiry there will be several different

processes that are relevant for fixing functions. There are more problems with the notion of function that I shall deal with shortly, but at this stage I just want to make the point that we cant tell a-priori whether we are dealing with a natural kind or not. Could be that there are two natural kinds in the vicinity (like what happened with jadeite and nephrite). Could be that there are many more samples that (what happened with whale). Could be that there is nothing that they have in common We cant determine this a-priori.

Wakefield goes wrong in prioritising our intuition about malfunction over the other intuitions that we have such as our intuition about our judgement of cases and our intuition about the natural kind assumption. While it could turn out that Wakefield is correct about mental disorders being malfunctions he is incorrect that this is essential to our concept of mental disorder. Three intuitions seem to be important here and much of the controversy between Wakefield and his critics can be understood as controversy over whether our judgement about cases or our judgement about malfunction should take priority when those things come apart. Murphy has maintained that the project of conceptual analysis needs to move from stipulating conditions from the armchair and towards a view where our intuitions are revisable and I agree with him in this. Instead of analysing the concept of mental disorder into necessary and jointly sufficient conditions I think it might be better captured by a Ramsey sentence which lists the features that we take to be most central to the concept. Thus far we have:

- Judgement of paradigm cases - Malfunction assumption - Natural kind assumption - Harm assumption (conditions we should treat)

These intuitions seem to be revisable in principle. Whether we revise them or not depends in part (or should jolly well) in how the world turns out. It might be objected that if we were to throw away one (or more) of these conditions then that would involve our changing our concept. One could look at it that way (depending on ones view of concepts) but the crucial thing is that the majority of people who are engaged in the debate seem to be more interested in defining mental disorder than attempting to describe what we believe about our current concept (as in the survey project) at any rate. If

the project is to systematise our intuitions so that they are consistent and we do in fact want there to be a role for scientific discoveries then we must allow for the possibility that scientific discoveries will lead us to revise some of the assumptions. If one stipulates that mental disorder is malfunction (as does Wakefield) and one allows that whether there is malfunction or not is to be determined by science (as does Wakefield) then there seems to be a tension in that it is possible (for all we know) that science will inform us that every single individual who we have regarded to be mentally disordered was not mentally disordered after all.

Carnap conditional and how there can be trade-offs Need to see how the science turns out. Or it could be that Wakefield would say that they must have something in common its just that science hasn't found the relevant process yet But what if there simply isn't a relevant process that they have in common? Murphy maintains that the malfunction assumption does for psychiatry what the adaptationist assumption does for evolutionary biology. He maintains sometimes the assumption is false and sometimes we do not know whether it is true or false but that does not impugn dx

Wakefield has responded to this line of critique by maintaining that identifying isn't relevant to his project (though it is of course relevant to DSM project and it is of course relevant to our notion that conceptual analyses should be elucidatory in a useful way). He also responds to this line of critique by clarifying the notion of function that is relevant. I now wish to turn to some more concerns that I have with his account of function.

For now it is enough to note that Wakefield's argument that science has discovered that evolution by natural selection is the relevant process for fixing functions and dysfunctions is not something that has been settled as he seems to think it has been. While the systemic function view doesn't deny that evolution by natural selection has occurred, it thinks that causal processes more generally care capable of fixing functions. It might turn out that the systemic view can be shown to be grounded in evolution. It seems problematic, though, and the evolutionary view seems to attribute different functions to the same explanandum than the systemic view would in some cases. Further

argument is required for Wakefield to show that the evolutionary notion is the relevant notion or that when these two notions posit different functions the evolutionary notion trumps the systemic one. It might be that both views can be developed in such a way that they merge or something... It is unclear how much it is an empirical discovery what process fixes functions and dysfunctions, however. While Wakefield thinks you can fairly unproblematically treat 'the process that fixes the functions' as a rigid designator in the same way that 'the molecular substrate of water' can be fixed Davies maintains that this issue is one that requires conceptual work and conceptual clarification rather than scientific discovery. It is clear that much more work needs to be done to establish Wakefield's premiss.

Chapter 9

Classification and natural kinds

In this chapter I want to outline the features of natural kinds that are supposed to make natural kinds interesting to us. Such features include generalisability, predictive leverage, intrinsic essence, non-intrinsic essence, etc. In order to avoid getting caught in controversy over how we should apply the term natural kind I wish to focus on different kinds of kinds (with respect to whether the feature is present or absent and with respect to how much the feature is present or absent). This section will provide a framework for a subsequent discussion of what could be going on with respect to current psychiatric disorders and with respect to some of the considerations behind changing psychiatric classification (so that more features are present or so they are present to a greater degree).

We tend to have this intuition that some classifications are realist in the sense that they at least aim to describe different kinds that are really out there in the world - like the classification systems of chemistry and biology. Their success can be judged according to how successful they are at carving nature at its objective joints. Other classifications are nominalist, however, in the sense that they are arbitrary or can only be regarded as better or worse with respect to how useful they are to us given our interests - such as classifications of different kinds of books in bookstores. In this chapter I will consider a range of classifications in order to try and draw out a number of

dimensions that they seem to vary on.

MIND INDEPENDENCE - I put things this way because we can of course aim to have a natural categorization of mental states. Mental states are of course mind dependent, but the thought is that a natural classification of mental states wouldn't depend on the mental states or interests of the classifiers.

I'll consider the notion that a successful classification 'carves nature at its joints' in virtue of capturing 'natural kinds' that are to be found in nature and that do not depend on the mental state of the observer[1]. While this conception might work well for some purposes it is less clearly suited to others. By way of illustrating this I'll consider classification for chemical elements, biological species, anatomy, medicine, gardening and cooking, computer science, books, and parliament in order to loosen the grip of both the 'natural kind' conception and the notion that there is a way to classify nature that is independent of our interests. I'll then turn to psychiatric classification in particular and introduce some of the assumptions behind the present system of classification. Part of the development of a science consists in the development of a classification system of entities that form the subject matter of investigation[2] One very traditional way of viewing classification in science is that it aims to 'carve nature at its joints' by way of grouping together the

[1]This way of putting things is problematic both when it comes to natural sciences that take mental states as object (psychology, for example) and when it comes to physics where it might be that the mental states of the observer are responsible for the collapse of the wave function. This later view seems surprising and controversial precisely because it results in physics being mind dependent, however. With respect to psychology while mental states require the object of investigation to have mentality this mentality seems quite independent of the mentality of the scientist who is investigating the phenomenon, however.

[2]This picture is, of course, complicated by sciences that deal in processes (such as kinds of brain state) and properties (such as the speed of light and the relation between this and other physical properties). I'll have more to say about complicated cases in due course, but firstly I wish to consider the comparatively simple cases of the classification of entities that form the subject matter of investigation. It is important to note that even in the complex cases a significant part of the development of science consists in our getting the characterization of the subject matter correct (whether it be the characterization of processes or properties) and a great deal of time and effort is expended on doing so.

'natural kinds' of entities. The thought is that there are 'natural kinds' or real groupings that are to be found in nature and thus the aim of classification in science is to classify on the basis of those real groupings. On this way of looking at things there is such a thing as getting a classification system right. A right classification groups entities according to the way they are actually grouped in nature. Dennett talks about how the notion of a 'natural kind' has its origins in Plato. Plato maintained that each member of a natural kind shared an intrinsic and unchanging essence in virtue of which it was a member of that natural kind. Carving nature at its joints was thus conceptualized as a matter of developing categories or classifications that described the kinds that were to be found in nature. Paradigmatic examples of natural kinds are thought to be chemical kinds such as gold and water and biological kinds such as lions and tigers. The notion of natural kinds has been the subject of much controversy especially in light of classification in biology so I'll begin the story with a discussion of classification of chemical kinds before I get to the more problematic notion of biological kinds. and then biological kinds before I get to more problematic cases and ultimately to a discussion of classification of psychiatric ailments.

One way to attempt to get at essences (indeed, the place that one must start from) is to consider the properties or features that one can readily observe. Gold is a yellowy malleable metal, for example, and tigers have stripes and sharp teeth and four legs. Part of the problem of developing a 'natural' classification system is to sort the features that are necessary from the features that are merely contingent, however. The thought is that a necessary feature is one that the instance needs to have in order to be an instance of the category. Contingent features, on the other hand, are not essential to the instances being a member of that category. Gold can be either liquid or solid, for example, as liquidity or solidity are contingent features of gold. Gold can't have atomic number 47, however, because having atomic number ?? is thought to be necessary to the instances being correctly classified as a sample of gold.

The development of the periodic table of elements for chemistry and the

notion that atomic weight was the necessary feature of chemical kind membership constituted a significant advance. While we might start classifying on the basis of readily observable features atomic weight is not a readily observable feature, however. The periodic table classifies not on the basis of readily observable features but rather on the basis of underlying features. The periodic table classifies according to atomic weight. The atomic weight is thought to be essential to each particular instance being a member of the kind. Things are complicated slightly in that it is possible to change one kind of substance into another kind of substance (lead into gold, for example) by shooting a number of protons out of the nucleus. That being said altering its essence (its atomic weight) would be the only way that one could do so. Similarly, one can bring new kinds of substance into being by shooting a number of protons into the nucleus thereby creating a new kind of substance with a different (essential) atomic weight. Once again altering its essence (its atomic weight) would be the only way that one can do so. What these examples show us, however, is that while it might be tempting to see essences as unchanging the situation is more that the essence is unchanging insofar as the essence is essential to the identity of the substance as that particular kind of substance.

When it comes to biological kinds the situation is more problematic, however. The development of modern genetics seemed to promise that the essential properties of species would be uncovered for biology similarly to how the development of modern chemistry provided the essential properties of chemical kinds. The problem of sorting the essential features from the accidental features seems to have recurred at the level of genetics instead of being resolved as it was in modern chemistry, however. Indeed, while there are genetic differences between species there is also considerable variation between individuals that are intuitively of the same species. There don't seem to be genetic essences in the way that there are chemical essences.

While number of protons is continuous in that something can have one or two or three or four or five it also seems clearly categorical in the sense that there is an objective fact about whether a substance has one or two or three

or four or five. Even substances that are blends of more than one substance can be broken down into the proportions of the substances that they are composed of and there is thus a fact about both what the blend is composed of and what composes each constituent in the blend. In the case of biology the situation is much more complicated, however. There doesn't seem to be a genetic essence that is common to each individual member of a species that individuals who are not a member of the species lack.

Darwin's revolutionary idea was that it was possible for a species to evolve into another species over time. Not just revolutionary because of the idea of change (there is change in the chemical case too, as we considered). It is revolutionary because of the idea that there might be individuals where it is genuinely indeterminate whether they are a member of the same species or not. Instead of what I'll call 'morphological genetics' the idea is that genetics is useful for getting at essences insofar as we can get information about the history of an organism by looking at the genetics. What is essential to species membership is thought to be its place on the evolutionary tree where the evolutionary tree is a description of the history of life and interbreeding. Species are thought to be populations of individuals that share a gene pool by virtue of interbreeding. Such a notion is (of course) problematic. But that is thought to be the basic idea. What do we say in the face of scientists taking this to be the essential feature of biological kinds?

One thing that has been said is that scientists have discovered that biological species are not natural kinds after all. On this view what is essential to a category being a natural kind is that the instances share an intrinsic unchanging essence. If species don't share an essence and if that essence can change over time and yet the individual is still considered an instance of the species then scientists have discovered that species aren't natural kinds after all.

Another thing that has been said is that scientists have discovered that what is essential to biological species is that the instances share historical properties such as that of interbreeding. Biological kinds are natural kinds after all - it is just that natural kinds can have essential properties that are different

from what we had supposed (i.e., they can be historical properties rather than intrinsic properties).

Still another thing that has been said is that we can see the historical property as being intrinsic rather than relational if we view species as individuals. On this view a species is an individual with complex intrinsic properties of interbreeding and it is just that individuals are different from what we had supposed (i.e., they can be composed of more basic entities that we typically reserve the term 'individual' for). This move seems to make species not a kind, however. Not the case that there are intrinsic essences shared by different individuals. More that an individual (the species) needs to retain certain properties in order to remain the individual (species) that it is. ?? Does it . This is a funny view and I'm not sure what to make of it. What to say in the face of this?

One thing to say is that even if theorists disagree on whether biological kinds are natural kinds or not it is possible in principle that they could all agree with a correct description of the evolutionary tree which located every biological entity that had ever existed.

Dennett talks about how once we have a description of the biological tree one might want to know where to draw the line on it with respect to where one species ends and another begins. We can start with a paradigmatic instance of a particular species (homo habilis, for example) and ask precisely where we should draw the line as to which individual was the first homo habilis and which individual was the last living instance of homo habilis. This seems a little like asking 'how many grains of sand make a heap', however. While we might agree as to how many grains of sand there are in any particular instance, and while we might agree that some instances clearly constitute a heap while others do not there are going to be cases that are both controversial (where different theorists have different opinions as to where we should draw the line) and also cases where any particular theorist will be unclear as to whether that amount of sand does or does not constitute a heap. It seems senseless to state that there is a further fact of the matter that will resolve this once and for all.

But of course it might be that case that there are different places that are *theoretically interesting* places to draw the line. I'll have much more to say about this notion of 'theoretical interest' when I attempt to describe classifications that seem more or less real as opposed to nominal (arbitrary or in 'name only') Atomic weight again... Number of atoms circling the nucleus... Seems different...

This idea was revolutionary. As Dennett notes (quote):

...consider what your attitude would be towards a theory that purported to show how the number 7 had once been an even number, long, long ago, and had gradually acquired its oddness through an arrangement whereby it exchanged some properties with the ancestors of the number 10 (which had once been a prime number).

The idea of ETERNAL essences. Less clearly applicable to the natural world, however. Firstly, the natural world is contingent. It needn't have existed. Even if it did exist it seems clear that it needn't have existed in precisely the form it does now. There needn't have been any gold, for example. There needn't have been any light, either. There needn't have been any biological species.

Essential to the kind existing that there is an individual (or a substance or a property) with the essential properties for that kind.

Darwin gave us a radically different conception of species and thus a radically different conception of how to classify species. Instead of attempting to classify on the basis of an essence (morphological or genetic) the idea was that the history of species was what was essential. Common descent. Interbreeding. But of these features which is necessary? Where does one species end and another begin?

The tree... Where precisely do we draw species boundaries? Important to note that scientists could agree completely on the tree structure (not that they do) and yet disagree as to where to draw the line to individuate species. What does it matter? What do they use the species concept for? It might be

that one group finds it theoretically interesting to draw it in one place while another group finds it theoretically interesting to draw it in another. Who is right and who is wrong? What (in the world) could settle this issue? Is the dispute merely verbal or is there substantive disagreement? Important issues for classification.

I now want to turn to some other classifications that we have. The idea here is to move from the one that seems to most respect the traditional view (the chemical) to one that seems to respect some parts of it but not others (the biological) to ones that are more problematic until we come to classifications that are uncontroversially nominal. In order to get clearer on these issues we need to look at what classification systems are used for. part of the subject matter of the science. The thought is that there is a 'right way' to carve up the world and sometimes this is expressed as finding the natural kinds.

The periodic table of elements was a significant development for chemistry. It turned out that there were a number of atoms that had different atomic weights and if you knew the atomic weight of an atom there was a great deal that you could say about it with respect to how it would behave in a variety of circumstances especially in relation to other atoms.

Dennett talks about how the idea of carving nature at its joints (and the assumption that nature had joints there to be discovered) had its origins in Plato's theory of forms. The idea was that each kind had an internal, immutable, and unchanging essence. Science was thus conceived of as being in the business of discovering what kinds there were by way of discovering essences.

Chemistry provides a paradigmatic example of the success of this kind of view. Each of the elements in the periodic table have atomic weight as essence. While chemistry is contingent in the sense that gold need never have existed on our world the idea is that since gold does in fact exist in this world gold necessarily has the essence that it does.

Aristotelian classification of biology was based on salient features of morphological similarity. One proceeded by observing and describing biologi-

cal entities and salient differences were supposed to enable one to discover whether two animals were of the same or different kind. With the development of modern genetics one can do a similar thing with genetic similarity and disparity instead of morphological similarity and disparity. Against this backdrop we can see that Darwin's theory was revolutionary for a number of ways. Instead of each species having its own internal immutable and unchanging essence it was thought that different species had a single common ancestor species and that species could evolve into other species over time. If Darwin was right then it seemed that biology wouldn't have essences in the way that chemistry did.

There is much controversy as to how we should see species in biology. Some theorists maintain that species aren't natural kinds because natural kinds have an internal unchanging essence. Other theorists maintain that species are paradigmatic examples of natural kinds and so if it turns out that species don't share an internal unchanging essence then an internal un- changing essence is not needed for natural kind-hood. While there are substantive disagreements one must be wary of verbal disputes. It would be possible for both kinds of theorists to agree on all the biological facts and the disagreement might simply be over whether we are or are not to apply the term 'natural kind' to species or not. They might agree on all the natural properties (problematic notion) that species have, just disagree as to whether species should have the name 'natural kind' or not. On this view no one of the features is essential, but all of the features are relevant. (Searle?) has pointed out that in this case we can describe it such that there are necessary and sufficient features for category membership, however. There must be some number of the features that it is both necessary and sufficient to have to count as a member of the category and hence we can say that that is the essence.

Some people feel repulsed somehow - that this kind of feature is arbitrary. The feeling seems to be that it is not a natural feature but some contrived arbitrary and stipulated feature that doesn't really deserve to be called a feature at all. One can define properties into existence - mereological fusions

of objects - but these don't deserve to be called 'natural kinds'. This issue is complex and I'll have a great deal more to say about it later. For now, it is interesting to note that every classification system aside from the fundamental theory of primitive physical properties and relations will be (or currently seems to be) multiply realizable at one or more lower levels of analysis. We need to find a middle ground between unchanging essences all the way down and nominal categories.

One fix in the case of biological natural kinds is to say that essences can be external instead of internal. On this view what is essential to a biological species is its history - interbreeding and so on. There are problems in the details but I don't think they matter for what I want to say here. One objection to this line (aside from the technical objection) is that history is external and hence contingent. One response to this objection is to consider species as individuals and thus history counts as internal. Dennett on books. Classifying different kinds of fiction. Perhaps... Classifying different kinds of computer problems. Might distinguish 'software' from 'hardware' problems. Or might distinguish them in a user salient way (e.g., dark screen on powerup, failure to print) each of these could be caused in a variety of ways - that crosscut software vs hardware. This is an example that I'll make much use of in later sections. Clearly nominal...

What seems salient on one level (e.g., morphological similarity) might not turn out to be the most useful way to carve up the world into kinds. Why might this be? Differences that are salient to us might not turn out to be useful differences for the purposes of prediction and explanation. Science seems to be about explaining the world and a significant part of how we find out about the world (and a significant constraint on when certain explanations seem to be good or misguided) is how much they enable us to control the world in various ways. I'll return to this issue later when I look at natural kinds, causal mechanisms, and multiple realizability. For example, we might be interested in dark screens on boot ups. The screen being dark on boot up is very salient to the user. This is what the user wants explained or fixed (more to the point). Whether the screen is dark because of software

or hardware failure might well not be salient to the user. Indeed, if the user knew that they would be further ahead with respect to fixing the problem / explaining what the problem was. In order to fix the problem one must apply a number of interventions to the system and see how the system responds. If the computer beeps - then this could indicate that the RAM isn't installed correctly (hardware fault). If turning the power button on helps then the problem is that the user didn't do what was needed (there was nothing wrong with the system after all. The environment was outside the range that the system was designed to operate in). If ... Adjusting the contrast settings... No problem. If ... Need a software problem that results in failure to boot up.

Dennett on Classification

'...by Darwin's time the work of the great taxonomists (who began by adopting and correcting Aristotle's ancient classifications) had created a detailed hierarchy of two kingdoms (plants and animals), divided into phyla, which divided into classes, which divided into orders, which divided into families, which divided into genera (the plural of "genus"), which divided into species. [from p. 35 to p. 36] Species could also be subdivided, of course, into subspecies or varieties - cocker spaniels and basset hounds are different varieties of a single species: dogs, or Canis familiaris.

How many different kinds of organisms were there? Since no two organisms are exactly alike - not even identical twins - there were as many different kinds of organisms as there were organisms, but it seemed obvious that the differences could be graded, sorted into minor and major, or accidental and essential. Thus Aristotle had taught, and this was one bit of philosophy that had permeated the thinking of just about everybody, from cardinals to chemists to costermongers. All things - not just living things - had two kinds of properties: essential properties, without which they wouldn't be the particular kind of thing they were, and accidental properties, which were free to vary within the kind. A lump of gold could change shape ad lib and still be gold; what made it gold were its essential properties, not its accidents. With each kind went an essence. Essences were definitive, and as such they

were timeless, unchanging, and all-or-nothing. A thing couldn't be rather silver or quasi-gold or a semi-mammal.

Aristotle had developed his theory of essences as an improvement on Plato's theory of Ideas, according to which every earthly thing is a sort of imperfect copy or reflection of an ideal exemplar or Form that existed timelessly in the Platonic realm of Ideas, reigned over by God. This Platonic heaven of abstractions was not visible, of course, but was accessible to Mind through deductive thought. What geometers thought about, and proved theorems about, for instance, were the Forms of the circle and the triangle. Since there were also Forms for the eagle and the elephant, a deductive science of nature was also worth a try. But just as no earthly circle, no matter how carefully drawn with a compass, or thrown on a potter's wheel, could actually be one of the perfect circles of Euclidean geometry, so no actual eagle could perfectly manifest the essence of eaglehood, though every eagle strove to do so. Everything that existed had a divine specification, which captured its essence. The taxonomy of living things Darwin inherited was thus itself a direct descendant, via Aristotle, of Plato's essentialism. In fact, the word "species" was at one point a standard translation of Plato's Greek word for Form or Idea, edios.

We post-Darwinians are so used to thinking in historical terms about the development of life forms that it takes a special effort to remind ourselves that in Darwin's day species of organisms were deemed to be as timeless as the perfect triangles and circles of Euclidean geometry. Their individual members came and went, but the species itself remained unchanged and unchangeable. This was part of a philosophical heritage, but it was not an idle or ill- motivated dogma. The triumphs of modern science, from Copernicus and Kepler, Descartes and Newton, had all involved the applications of precise mathematics to the material world, and this apparently requires [from p. 36. to pg. 37] abstracting away from the grubby accidental properties of things to find their secret mathematical essences. It makes no difference what color or shape a thing is when it comes to the thing's obeying Newton's inverse-square law of gravitational attraction. All that matters is its mass.

Similarly, alchemy had been succeeded by chemistry once chemists settled on their fundamental creed: There were a finite number of basic, immutable elements, such as carbon, oxygen, hydrogen, and iron. These might be mixed and united in endless combinations over time, but the fundamental building blocks were identifiable by their changeless essential properties.

The doctrine of essences looked like a powerful organizer of the world's phenomena in many areas, but was it true of every classification scheme one could devise? Were there essential differences between hills and mountains, snow and sleet, mansions and palaces, violins and violas? John Locke and others had developed elaborate doctrines distinguishing real essences from merely nominal essences; the latter were simply parasitic on the names or words we chose to use. You could set up any classification scheme you wanted; for instance, a kennel club could vote on a defining list of necessary conditions for a dog to be a genuine Our kind Spaniel, but this would be a mere nominal essence, not a real essence. Real essences were discoverable by scientific investigation into the internal nature of things, where essence and accident could be distinguished according to principles. It was hard to say just what the principled principles were, but with chemistry and physics so handsomely falling into line, it seemed to stand to reason that there had to be defining marks of the real essences of living things as well.

From the perspective of this deliciously crisp and systematic vision of the hierarchy of living things, there were a considerable number of awkward and puzzling facts. These apparent exceptions were almost as troubling to naturalists as the discovery of a triangle whose angles didn't quite add up to 180 degrees would have been to a geometer. Although many of the taxonomic boundaries were sharp and apparently exceptionless, there were all manner of hard- to-classify intermediate creatures, who seemed to have portions of more than one essence. There were also the curious higher-order patterns of shared and unshared features: why should it be backbones rather than feathers that birds and fish shared, and why shouldn't creature with eyes or carnivore be as important a classifier as warm-blooded creature? Although the broad outlines and most of the specific rulings of taxonomy were undis-

puted (and remain so today, of course), there were heated controversies about the problem cases. Were all these lizards members of the same species, or of several different species? Which principle of classification should "count"? In Plato's famous image, which system "carved nature at the joints"?

Before Darwin, these controversies were fundamentally ill-formed, and could not yield a stable, well-motivated answer because there was no back- [from p. 37 to p. 38] ground theory of why one classification scheme would count as getting the joints right - the way things really were. Today bookstores face the same sort of ill-formed problem: how should the following categories be cross-organized: best-sellers, science fiction, horror, garden, biography, novels, collections, sports, illustrated books? If horror is a genus of fiction, then true tales of horror present a problem. Must all novels be fiction? Then the bookseller cannot honour Truman Capote's own description of In Cold Blood (1965) as a non-fiction novel, but the book doesn't sit comfortably amid either the biographies or the history books. In what section of the bookstore should the book you are reading be shelved? Obviously there is no one Right Way to categorize books - nominal essences are all we will ever find in that domain. But many naturalists were convinced on general principles that there were real essences to be found among the categories of their Natural System of living things. As Darwin put it, "They believe that it reveals the plan of the Creator; but unless it be specified whether order in time or space, or what else is meant by the plan of the Creator, it seems to me that nothing is thus added to our knowledge" (Origin, p. 413).

Problems in science are sometimes made easier by adding complications. The development of the science of geology and the discovery of fossils of manifestly extinct species gave the taxonomists further curiosities to confound them, but these curiosities were also the very pieces of the puzzle that enabled Darwin, working alongside hundreds of other scientists, to discover the key to its solution: species were not eternal and immutable; they had evolved over time. Unlike carbon atoms, which, for all one knew, had been around forever in exactly the form they now exhibited, species had births in time, could change over time, and could give birth to new species in turn. This

idea itself was not new; many versions of it had been seriously discussed, going back to the ancient Greeks. But there was a powerful Platonic bias against it: essences were unchanging, and a thing couldn't change its essence, and new essences couldn't be born - except of course by God's command in episodes of Special Creation. Reptiles could no more turn into birds than copper could turn into gold.

It isn't easy today to sympathize with this conviction, but the effort can be helped along by a fantasy: consider what your attitude would be towards a theory that purported to show how the number 7 had once been an even number, long, long ago, and had gradually acquired its oddness through an arrangement whereby it exchanged some properties with the ancestors of the number 10 (which had once been a prime number). Utter nonsense, of course. Inconceivable. Darwin knew that a parallel attitude was deeply engrained among his contemporaries, and that he would have to labour mightily to overcome it. Indeed, he more or less conceded that the elder authorities of his day would tend to be as immutable as the species they believed [from p. 38-p. 39] in, so in the conclusion of his book he went so far as to beseech the support of his younger readers: "Whoever is led to believe that species are mutable will do good service by conscientiously expressing his conviction; for only thus can the load of prejudice by which this subject if overwhelmed be removed" (Origin p. 482).

Even today Darwin's overthrow of essentialism has not been completely assimilated. For instance, there is much discussion in philosophy these days about "natural kinds," an ancient term the philosopher W. V. O. Quine (1969) quite cautiously resurrected for limited use in distinguishing good scientific categories from bad ones. But in the writings of other philosophers, "natural kind" is often sheep's clothing for the wolf of the real essence. The essentialist urge is still with us, and not always for bad reasons. Science does aspire to carve nature at its joints, and it often seems that we need essences, or something like essences, to do the job. On this one point, the two great kingdoms of philosophical thought, the Platonic and the Aristotelian, agree. But the Darwinian mutation, which at first seemed to be just a new way of

thinking about kinds in biology, can spread to other phenomena and other disciplines, as we shall see. There are persistent problems both inside and outside biology that readily dissolve once we adopt the Darwinian perspective on what makes a thing the sort of thing it is, but the tradition-bound resistance to this idea persists. He then talks about trees and where to colour the red.

9.1 Natural kinds

The main purpose of a scientific classification is often thought to be to provide a number of different categories that accurately capture kinds that are to be found in nature[3]. The traditional view of natural kinds is that instances are appropriately regarded as being members of the same natural kind when they share a certain essence in common. This is to be contrasted with nominal classifications that are in 'name only' where while different instances might fall under the same concept they don't have an essence in common aside from their falling under the same concept. There are a number of features that natural kind essences were traditionally thought to have and I'll consider each of these in turn.

9.1.1 Dimensions of variation

Necessary and sufficient for kind membership

Essences are thought to be necessary and sufficient for kind membership in the sense that whether a token instance is a member of a particular natural kind is solely determined by whether the token instance has the feature or property that is necessary and sufficient for membership in that natural kind. It could be the case that the particular token instance can persist through changes in the necessary and sufficient features for natural kind membership such that an instance could be a member of one kind at one point in time

[3]I'll use the term 'category' to refer to the types posited by the classification system and the term 'kind' to refer to the types to be found in the world.

and a member of another kind at another point in time. I shall consider this further in the section on chemistry. This seems to show that the identity conditions for tokens can be different from the identity conditions for kind membership, however.

Objective rather than mind-dependent

Essences are thought to be mind-independent or objective in the sense that whether the essence is present in a particular token is determined by facts about that particular token that are quite apart from beliefs that we have about the presence or absence of that feature. It is important to note that this way of characterizing mind independence allows for there to be mind-independent facts about psychological kinds that determines whether or not a token state is really a state of belief (for example). While beliefs are clearly mind-independent in the sense that beliefs are mental states the essential feature of belief is thought to be mind-independent in the sense that a token state either has them or lacks them despite what you or I or everyone believes about whether the feature is present or absent.

Intrinsic rather than relational

Essences were traditionally thought to be intrinsic (internal) rather than relational (extrinsic) though this view has been challenged by advances in the biological sciences in particular. There has been much debate over whether biology has shown us that essences can be relational, or whether biology has shown us that given that biological essences are historical, biological kinds aren't natural kinds after all. I shall consider this in more depth in the section on biology. For now, it is enough to note that I'm less interested in dispute around how we choose to apply or withhold the term 'natural kind' and more interested in the essential properties for the kinds that are posited by biology, and other classification systems. Whether we call these properties or features 'natural' or not isn't as interesting as the nature of those features.

Projectable rather than gerrymandered

The thought here is that the kinds of features that are candidates for essential properties are features of a certain kind. Sometimes this is put as their being 'simple' or 'natural' features. Disjunctive or gerrymandered features won't do. Sometimes this is put in terms of predictability where the idea is that natural features are projectable whereas gerrymandered or disjunctive features are not. Characterizing 'natural' features is a considerable problem, however. I'll have much more to say about this in what is to come.

Discovered A-Posteriori rather than A-Priori

The thought here is that essences are properties of objects that exist independently of our beliefs about their presence or absence and as such the features or properties must be discovered a- posteriori rather than a-priori. So, while one could attempt to give a necessary and sufficient condition analysis of 'guilt' a-priori this wouldn't be an account of the necessary and sufficient properties of the state that make it a state of guilt. In order to find out what is essential to the state of guilt we need to investigate guilt and discover the features that it has. I'll deal with this issue at length in chapter three.

9.2 Classification systems

9.2.1 Chemistry

The modern periodic table of elements seems to be a paradigmatic example of the success of the traditional view of natural kinds for capturing the nature of essences and also a paradigmatic example of a successful classification that carves nature at its real (essentialist) joints. It is thus worth starting with chemistry in order to see how the traditional view of natural kinds and scientific classification fares.

While the periodic table of elements describes a number of features of elements (atomic number, atomic mass etc) it divides up chemical substances according to their atomic number (the number of protons) and the number

of protons provides the essences (the necessary and sufficient conditions) for category membership.

Atomic number is objective in the sense that the atomic number of the elements is determined by the number of atoms circling in the nucleus rather than our beliefs about it. The essences are literally intrinsic or internal in the sense that the atomic number refers to the number of protons inside the nucleus. The essences are thought to be natural both in the sense of being discovered by science (no amount of a-priori reflection will tell us how many atoms circle the nucleus) and also with respect to the predictive utility that we have once we know the atomic number of a substance. The essence enables us to predict a number of the superficial or observable properties of substance.

This being said, there are a number of features of chemical substances that are of interest to chemists. If we are interested nuclear properties (including stability) then the atomic mass is of more interest to us where the atomic mass has to do with the number of neutrons in the nucleus[4]. While the periodic table of elements does state the atomic mass of the elements the atomic mass is arrived at by averaging the atomic mass of the instances of the element. There are three isotopes of hydrogen, for example. Hydrogen-1 -1 (sometimes called protium) has one proton and no neutrons in its nucleus. Hydrogen-2 (also called deuterium) has one proton and one neutron in its nucleus. Hydrogen-3 (also called tritium) is a radioactive isotope with one proton and two neutrons per nucleus. Each of these has a different atomic mass but the atomic mass given in the periodic table of elements is arrived at by considering the proportions of the different things in the world and averaging them. This number thus fails to distinguish between the radioactive versions and the non-radioactive versions.

One thing that we could say is that the periodic table of elements provides the natural kinds for chemistry where the natural kinds are determined on

[4]Atomic mass is arrived at by knowing the atomic weight. The atomic weight is determined by the number of protons together with the number of neutrons and the number of electrons. The weight of electrons is thought to be negligible.

the basis of the number of protons. There are different degrees of fineness of grain, however, and the isotopes (Hydrogen- 1, Hydrogen-2, Hydrogen-2) are different kinds of hydrogen as their names suggest. This way of putting things misses the predictive utility that scientists gain from regarding a substance to be a radioactive or not radioactive substance, however. One might say that radioactive substance is not a natural kind of substance because radioactivity is multiply realized with respect to atomic mass.

It is important to note that the atomic mass is similarly objective, mind-independent, simple etc as the number of protons. It is also important to note that there are objective facts about both such that we simply could be wrong. But if we have a classification based on the number of protons and another classification based on the atomic mass (and both of those are correct with respect to carving up different substances according to the number they actually have) then which of these classifications captures the mind-independent chemical kinds? It is important to note that two different scientists could completely agree on a description of reality with respect to the number of protons and the number of neutrons that different substances have. They could completely agree that the elemental classification classifies elements according to reality and that the atomic mass classification classifies substances according to their atomic mass.

Another thing that it is important to note about chemistry is that essences aren't eternal in the way that the traditional view of natural kinds had them to be. The thought here is that the basic substances were eternal in the sense that they could neither be created nor destroyed. We know now that that basically holds, but that the exception to that is in the context of a nuclear reaction. If the essential features are determined by atomic number or atomic mass then if we can alter the number of atoms or neutrons inside the nucleus we would have succeeded in transmuting a sample of one element into a sample of another element. Indeed, this has been done (though it is probably a great deal cheaper to simply purchase a new sample). I'll return to this point with respect to biological kinds.

9.2.2 Biology

While paradigmatic instances of natural kinds include chemical kinds such as water and gold they also include biological kinds such as lions and tigers. The story gets complicated here, however. Aristotle attempted to classify biological kinds on the basis of morphological features. Some creatures have feathers and some have scales and some have hearts and kidneys and others do not. Part of the trouble with biological kinds is the considerable variation that is found in nature, however. While most dogs have four legs some dogs can lose one or more of their legs through the course of their lifetime without ceasing to be dogs and some dogs can be born without four legs and yet still be dogs. Idealization thus seems to be necessary.

Plato had this notion that there was a realm of forms where entities existed in their ideal form and where things in the world were more or less perfect copies according to how much they resembled those ideal forms. This is problematic, however, as similar number of features, similar degree of features and so on. Cluster notions matter of degree seems possible for there to be funny borderline cases. The postulation of the ideal realm of forms is problematic, however. Instead of regarding the paradigmatic cases to be the ideal cases one could look to the actual world to fix the paradigmatic cases. This seems problematic with respect to family resemblance which is meant to be antimonous? to natural kinds, however. Disjunctive features seem problematic. Always possible to turn a cluster analysis into a necessary and sufficient analysis, though. 'natural' seems to be doing a lot of work... need a good account of that. problems, though. gruesome features.

The development of modern genetics seemed to promise hope that the essential features of biological kinds would be discovered as the development of modern chemistry offered the essences of chemical kinds. Non-obvious, underlying, to be discovered by science. If we shift from surface morphology to underlying morphology (genetic similarity) then the problem only seems to recur, however. Genetic disparity seems to be a matter of degree. There doesn't seem to be an underlying genetic essence that is necessary and suffi-

cient for kind membership in biology as there seems to be in chemistry. One suggestion is that what is essential for biological kind membership is lineage. This is to posit a extrinsic, relational essence, however. On this view what biological kind a particular instance is a member of is a matter of how that individual is related to other individuals. If one were to alter the external relations then the instance would no longer be a member of that kind. Swamp man, for example. One view of species that would have species membership determined by internal relations is the view that species are individuals, however. On this view species membership of particular individuals and how they relate to other individuals would be an internal feature of them.

I don't want to commit myself to whether this view is the best view of species at the end of the day. What this view does usefully seem to illustrate, however, is that whether a property is internal or relational is a matter of point of view. What is an internal relation from one point of view can be an external relation from another point of view. The bonds between hydrogen and oxygen is an external relation from the point of view of seeing hydrogen and oxygen as basic atoms, but the bonds are an internal or intrinsic relation from the point of view of seeing water as basic. Some theorists maintain that natural kinds must have internal essences therefore biological kinds are not natural kinds. Other theorists maintain that biological kinds are paradigmatic natural kinds and thus it is more appropriate for us to conclude that natural kinds may have relational essences after all. This seems to me to be partly verbal and I'm anxious not to get caught up in that debate. It seems more important (for my purposes at least) to get clearer on what makes a lion or a tiger a member of its kind than to debate about whether the kind deserves the name 'natural' or not.

9.2.3 Anatomy / neuroanatomy

Structure and Function.

Hearts, kidneys, etc constitute different kinds from the perspective of anatomy and physiology. There are structural ways and functional ways. If we go with

structure then idealization will be required (as in the biological case). One could go with historical properties, and indeed this is common. This idea is to distinguish according to functions, however.

A token counts as an instance of a type if it has the functional property of the type. We want to allow that individual tokens might not perform their functions, however. In this case it seems that the function attaches to the types rather than the tokens. But if the function attaches to the types rather than the tokens such that the tokens might perform the function or not perform the function then we need some independent way of specifying how tokens get to be members of the type other than that they have the functional property. One way of specifying the functional property is on the basis of history. Natural functions. A token counts as a member of a type if it resulted from natural selection working on past tokens and the function is whatever resulted in their surviving. We need to figure what it was about the token that enabled it to survive, however. I'll look at functional analysis in more depth in a later chapter.

There is a great deal of controversy about psychological types and neuroscientific types. The thought is that psychological types must map onto neurological types somehow. The field of cognitive neuropsychology attempts to provide a bridge between cognitive types and neurological types. Bit of a multiple realizability bump from psychological kinds and cognitive kinds, though. One thing that is important to note is that cognitive kinds seem to be functional whereas (on a first pass at least) neurological kinds seem to be structural. There are many ways that we can carve up the brain into structural parts, however. It was a significant discovery that one fairly intuitive way of carving the brain into structural components seemed useful with respect to different structural parts playing a distinctive functional role. But still, it seems that even with respect to neurology the functional parts are what we are more interested in. Consider the language processing areas, for example, which are localised in the left hemisphere for the majority of subjects but in the other hemisphere or in both hemispheres for a smaller minority of subjects. The correct thing to say here seems to be that lan-

guage processing areas are multiply realizable. A language processing area is whatever processes language wherever it is localised but that there was some localisation structurally that underwrote those abilities was kinda cool.

9.2.4 Psychology

Neuroanatomy, cognitive science, folk psychology. What are the psychological kinds? Neuroanatomy often looks for the functional kinds. It is an interesting and significant finding that structural morphology (on one salient way of looking at structural morphology) maps fairly interestingly onto functional similarity. But language processing area and we can see that functional similarity and multiple realizability is allowed.

9.2.5 Medicine

9.2.6 Computer science

There are a variety of computer errors. Fixing the functions seems relatively unproblematic - they are designed by an agent with a certain intention. When the system fails to do what it is designed to do then there is a problem. It is important to note that whether a computer is malfunctioning (in the design sense) isn't as important as the fact that it isn't doing what it is designed to do. If I am getting a blank screen on start up then there is a problem. Even if the problem arises from my failing to plug my desktop in or because there is a power failure because the line is down there is a still a problem with blank screen on start-up. Of course fixing the problem will depend on what caused the problem. If there is a power failure then I need to wait until power is up (or take steps to getting the power up faster. If the contrast setting is too low then I need to adjust the contrast. If the screen is broken I need to take my computer into the store and get someone to fix the chip.

One might think that the appropriate way to classify computer faults would be according to their causes. This doesn't seem adequate for what we want to do with such a classification, however. The person reports the symptoms and the trouble shooting (diagnostics) involve figuring out the relevant causal

processes. The person would need to engage in some trouble shooting (diagnostics) in order to figure out what kind of problem there is if kinds were individuated on the basis of causal processes.

This is especially relevant to medicine / psychiatry with respect to whether a classification should aim to capture superficial / readily observable problems and then trouble shoot causal mechanisms and look at fixing them... Or whether psychiatry should aim to diagnose on the basis of underlying causal mechanisms. What kinds of problems are there really? What are the real kinds? Kinds of superficial problems or kinds of causal mechanisms problems? Depends on what we are interested in... Keep coming back to the behaviour that is problematic.

9.2.7 Cooking and gardening

Folk biological classifications. For the purposes of cooking tomatoes are often regarded as a vegetable. For the purposes of cooking we have italian herbs and so on and so forth. These can come apart from biological classifications. Is it that the folk classification is wrong (so a tomato is really a fruit) or is it that for our purposes of cooking tomato really is a vegetable? We often prioritize the scientific. Water case. Depends on what our interests are, however.

9.2.8 Books

Dennett has an example of classifying books in a bookstore as a nominalist classification system. Fiction and non-fiction and biography etc. There might be borderline cases to be sure. What further fact of the matter will determine the issue? Our interests seem particularly salient in this case.

9.2.9 Psychiatry

In the next chapter I'll talk about the present classification system in psychiatry and some of the assumptions behind the present classification system. I'll also talk about the interests that drive the classification system. The

DSM is explicit in a number of aims and it seems to me that there might not be one system of classification that best answers to all of those deriserata. In particular, I'll show that the classification system that is of most use to researchers could look quite different from the classification system that is of most use to clinicians. That psychiatry could have more than one classification system depending on what features we are interested in capturing / explaining doesn't undermine psychiatry's status as a science anymore than there being different ways of capturing / explaining chemical phenomena undermines chemistry's status as a science, however.

By way of preview issues include: Whether mental disorders are best thought of as morphological (like how biological kinds used to be and how computer errors currently are). Whether they are best thought of as having some intrinsic underlying essence - like chemical kinds. Whether they are best thought of as having some relational essence (cause or perhaps necessary cause like sunburn and biological essences). And... An ongoing theme for me, how much our interests do (and indeed should) drive psychiatric classification such that getting clearer on them is just as important as investigating the world.

One of the thing that I think is happening here is that we are seeing that there is more diversity to scientific projects than we may have supposed. There are problems with the 'levels' view of science especially to the extent that scientists are engaged in different research projects and the field divisions within academia are arbitrary to a large extent. Not unconstrained, but less constrained than the levels approach to science suggests. We get ourselves into a muddle sometime with thinking that there are levels of metaphysical supervenience of the entities and that the fundamental natural kinds are not multiply realized. But there seem to be different ways to carve out essential vs contingent properties. Problems with constitution and causation...

Diagnosis of mental disorder is made on the basis of behaviour - including the verbal behaviour - of subjects. In saying this it is important to note that while there are problems with distinguishing behaviours from related notions such as involuntary movement, growth, development, morphological

change etc we don't have direct access to the mental states of subjects. This is just the behaviourist point that the thought that mental disorder involves a disorder of mental or cognitive processing this is an inference that we make from the behaviours of subjects. While we might not be behaviourists any more (philosophy has moved on from the behaviourist paradigm of thinking about the mind) it does seem to have been taken on board that we don't observe mental states directly.

Of course the main distinction between behaviourism and functionalism (its modern heir) is in whether mental states are to be identified with the behaviours or whether they are the cause of the behaviours (since something can't cause itself identifying mental states with behaviours conflicts with the intuition that mental states cause behaviours). Similarly we can see a current debate between theorists who maintain that one can be an alcoholic (or have a mental disorder) without engaging in drinking behaviours and theorists who maintain that one can't abuse alcohol without drinking overly much. The first notion is that of alcoholism being an inner cause of drinking overly much whereas the second notion is that of alcohol abuse being a description of drinking overly much.

Behaviours are thus very important. Either because (on one view) mental disorders just are to be identified with certain behaviour or because (on another view) mental disorders are the cause of certain behaviours.

9.2.10 A toy model

Now that we have behaviours in place we are now ready to consider a very simple model of mental disorder. On this model we have a neurophysiological state (let us call it A) that is predictive of behaviour A. So individuals who have the behaviour have the neurophysiological state and individuals who don't have the behaviour lack the state. Perfect correlation. Issue: Are mental disorders to be identified with the behaviour or with the neurophysiological cause of the behaviour? It doesn't seem to matter so much so long as there is a perfect correlation (and interfering or preventing one involves

interfering or preventing the other). Insofar as there isn't a perfect correlation (either actually or possibly) then it seems important which we identify the mental disorder with, however. It makes a difference.

The above considered two factors: That of behaviour and that of neurophysiology. There are relationships between other factors that have been considered, however. For instance, the relationship between cognitive states and behaviours or social states and behaviours or environmental states and behaviours or genetic states and behaviour. None of the correlations seem to be perfect but we can talk about the robustness of the relationship with respect to how strong it is.

MODEL A

In this model individuals come to the attention of psychiatric services in virtue of their behaviour (including their verbal reports). In this respect the model is similar to the situation that we find ourselves in. In this case, however, evaluators find it obvious that there are a discrete number of different types of behavioural symptom.

In this model each type of behavioural symptom turns out to be perfectly correlated with a type of neurology. They are also perfectly correlated with a type of aetiology. The course (evolution of symptom over time) is also perfectly correlated. The response to a kind of treatment is, too. WHAT KINDS OF MENTAL DISORDER ARE THERE? This issue is supposed to be obvious.

WHAT IS MENTAL DISORDER?

Is the mental disorder to be identified with the behavioural morphology? With the neurology? With the aetiology? With the course? We might think that the latter two are ruled out because the aetiology is (by definition) the cause of and the course is the effects of. This might just be a way of speaking, however.

Consider two burns that are morphologically identical. One can be a sunburn but another not because only the first is caused by the sun. Here the notion

of sunburn includes that it is caused by the sun. Mental disorders might be similar to this. Similarly the notion of a poison includes that it causes toxic reaction. Mental disorders might be similar to this.

WHAT DIFFERENCE DOES IT MAKE?

Where there is a perfect correlation it doesn't seem to matter. Where there is an imperfect correlation it does matter, though. Different answers will result in different individuals being classified differently.

WHILE SCIENCE CAN DISCOVER CORRELATIONS IT CAN'T DISCOVER IDENTITIES UNLESS IT IS PART OF OUR CONCEPT THAT IT IS TO BE IDENTIFIED WITH THE EMPIRICAL CORRELATES.

WHERE CORRELATIONS ARE IMPERFECT WE CAN: DENY THE CORRELATION (RESIST THE IDENTITY) REVISE OUR CONCEPT SO THE CORRELATION IS PERFECT (ACCEPT THE IDENTITY)

Kraeplin's vision

Kraeplin is often hailed as the father of empirical or scientific psychiatry in virtue of his positing different kinds of mental disorder on the basis of patients file notes. While the types of disorder that he posited haven't survived in his form into the present day his methodology of gathering empirical data from which to posit kinds has resulted in his being hailed as the father of scientific psychiatry.

Mental disorders are diagnosed on the basis of behavioural (including verbal) data. While there are issues around sorting this data into kinds or types of symptoms / signs the behaviour is what is considered problematic and it is this that leads to people coming to the attention of psychiatric services for evaluation. For now I will set aside the issue of how we class behaviours into different symptoms (e.g., delusion, hallucination etc). This problem will be addressed later. If we take as granted for now that symptoms such as delusion and hallucination are readily identifiable types or kinds of symptoms then there is a question we can ask: Are there different symptoms that are clustered together (correlated) enough to constitute different kinds of

disorder. It might be, for instance that it is found that symptoms A, B, C, D, and E tend to co-occur. We might then consider that these symptoms are necessary and sufficient conditions for the disorder. Or there might be a degree of probability involved in the sense that the correlations are higher than would be expected.

Kraeplin thought that behavioural symptoms differed between different kinds of disorders but there may be overlapping. He thought that different kinds of disorders could be differentiated according to their aetiology and course, however. The idea was that schizophrenia and bi-polar were different in the sense that their aetiologies and courses differed. A person presenting with their symptoms might be identifiable as one or the other. Or it might take aetiology to predict the likely course and the course could act as confirmation.

ISSUES ARISING

Let us consider a very simple model of mental disorder - one that does not obtain in the actual world. I will then add various complications to the model so that it comes to approximate the way things seem in the actual world. Along the way he thought that different kinds of disorders could be differentiated on the basis of different aetiologies and that the different aetiologies resulted in different courses or evolutions of the behavioural symptoms over time. So, for instance cluster A and cluster B had different aetiologies and courses.

Ideal mapping.

Consider a very simple model of disorder whereby there are different behavioural symptoms with no overlapping. Each has a distinct aetiology and a distinct course. In this case there seems to be a strong case for there being different kinds of disorders. There is a problem of depth here, with respect to differentiating disorder from aetiology from course - but it does seem that there are different kinds of disorders. It doesn't make a difference so long as individuals who have one aspect (e.g., the aetiology) have the others.

Problems arise with what we are to say when they come apart, however. If

someone has the aetiology but not the rest then do they have the disorder? Here there might be a temptation to say that the aetiology is the cause but not a constituent. Now the issue of differentiating cause from constituent seems to be crucially important for deciding whether the individual has the disorder or not. Causes (not necessary), necessary causes (e.g., the sun for sunburn), constituents (not necessary), necessary constituents. We can make the same manoeuvrer with respect to effects.

The problem also arises if we consider overlapping effects and overlapping causes. The picture becomes much more complicated. One issue of discovering correlations - another issue of teasing out causal relations. Yet another issue of figuring out which of these are necessary for membership. It makes a difference.

The depth and spread problems have been talked about in the philosophy of mind with respect to how concepts get to 'hook onto' stuff in the world. The depth and spread problems are problems for concepts generally and also for the concept of mental disorder in particular. Complications arise. Many factors have been thought to be relevant for disorder. Genes, neurophysiology, cognition / mental, behaviour. Aetiology is often thought to be the cause of mental disorder whereas course is the manifestation of the disorder that may evolve over time. Aetiology occurs prior and course involves being temporally later. Potential problems arise with distinguishing causes from the phenomenon itself.

Consider the phenomenon of sunburn. A sunburn is a burn that is caused by the sun. Sunburn might be a candidate for a concept that has a particular cause or aetiology built into it. Two burns can be morphologically identical but one burn can be a sunburn whereas the other is not because only the former was caused by the sun. Of course an alternative way of speaking is to say that the aetiology of sunburn is not built into the notion of sunburn. The aetiology of sunburn involves the cause of the person coming into contact with the sun such that burn resulted. So, for instance, playing outdoor sports or swimming might be aetiologies of sunburn rather than the contact with the sun being part of the aetiology.

This example shows us that there are issues with spread. We might not think that debate over whether sunburn has aetiology built in or whether there is an independent aetiology of sunburn makes much of a difference, but there are differences in which phenomenon gets to be included or excluded as sunburn. This could well make a difference to things that we find out about it - the generalizations that hold and the ones that don't. Also makes a difference with respect to whether something is a conceptual truth about sunburn (scientists could never discover sunburn was not, in fact caused by the sun) or whether something is an empirical issues.

One might think that natural kinds of mental disorder will be aetiological in the sense that there will be a genetic or a neurological or a cognitive condition that is required for it. Kraeplin had this notion that each mental disorder involved a distinct neurological pathology along with an aetiology and course. So, for instance, mental disorder A will have behavioural profile A, neurological pathology A, and course A. There might be overlap in symptoms, how- ever. Models of Disease

Zachar and Kendler offer six conceptual dimensions which underlie common assumptions about what counts as an adequate category of psychiatric disorder.

i) causalism-descriptivism

Should psychiatric disorders be categorized as a function of their causes (causalism) or their clinical characteristics (descriptivism)?

Discovery of a discrete and unique cause vs accurate description of a conditions signs, symptoms, course, and typical outcome.

At least three different approaches to the role of causalism. Temporizing we will have to settle for descriptive approaches until we understand the real causes. Robust descriptivism the causal structure of psych illness is so complex, resulting from the actions and interactions of many individual causes each typically of small effect, as to be useless to solve nosological questions. The causal model rooted in infectious diseases with one clear

aetiological agent is simply inappropriate for complex conditions like psych disorders. Third, intermediate position argues that despite the complexity of the causes one particular class of causal features (e.g., genes, neurochemistry, structural brain changes) might for practical reasons be given priority when making particular nosologic decisions.

ii) essentialism-nominalism

Are categories of psychiatric disorder defined by their underlying nature (essentialism) or are they practical categories identified by humans for particular uses (nominalism)? Essentialist believes that psych disorders exist independently of our classifications and the job of nosologists is to discover their inherent natures and classify them accurately. E.g., gold, oxygen appear to be entities sharing the same underlying properties. Two approaches to nominalism. Radical nominalist argues that we must pick our categories for their utility with no expectation that they will reflect deeper truths about the world. Moderate nominalist agrees there is some structure of psychiatric illness out there in the world but no one unique categorization that stands above the others on a-priori grounds.

Advocates of moderate nominalism suggest the world is heterogeneous and classification requires highlighting some features and minimizing others. Development of classification involved discovering facts about disorders that allow us to lump, split, weigh, and order them for particular purposes. We discover rather than invent but there are multiple ways to divide up disorders and no way has universal priority for all purposes. Decision as well as discovery.

iii) objectivism-evaluationism

Is deciding whether or not something is a psychiatric disorder a simple factual matter (something is broken and needs to be fixed) objectivism or does it inevitably involve a value judgement (evaluationism)

iv) internalism-externalism

v) entities-agents

vi) categories-continua

Different Kinds of Reference

I wont attempt to define a category at this stage as the notion should get clearer through this section and shall have much more to say about them in later sections.

The first variety of reference that I want to consider is Nominal Reference. When there is nominal reference a concept that is intended to refer to a category turns out not to refer to a category. Griffiths offers the example of Aristotle's notion of a SUPER-LUNARY OBJECT as an example of such a concept. The only property that the instances have in common is the property of falling under the concept and the instances don't share properties in common that are useful for scientific generalisation and prediction. In the face of nominal reference concepts are discarded for scientific purposes. If the concept of mental disorder or a concept of a particular kind of mental disorder turned out to have nominal reference then we should eliminate that concept from the science of psychiatry.

Another way that reference could go would be split reference where the concept refers to more than one category. The most often cited example of split reference is how our concept GREENSTONE turned out to refer to two different categories: jadeite and nephrite. While in this instance we eliminated the concept GREENSTONE from science there are other cases where we retain the concept such as when biologists conclude that there are two species of Tuatara.

Another way that reference could go would be if there turned out to be partial reference. In partial reference our concept is found to refer to refer to more instances than we had taken there to be. When we learned that whales were mammals, for example, then we had to revise our beliefs about mammals. Another way that partial reference could go would be if our concept referred to a category but we also took it to refer to a collection of other instances that turned out not to share generalisable properties with instances of the category. Wakefield criticises the DSM for being too liberal

with the criteria so that many individuals are identified as being mentally disordered when they aren't. He argues this on conceptual grounds because it follows from his harmful dysfunction analysis of the concept of disorder rather than because of the lack of generalisable properties, however. For our concepts to be maximally scientifically fruitful it would be best if we revised our beliefs about them so that we can identify members of categories that share properties in common that allow us to make generalisations and predictions. In the face of partial reference we could eliminate our concept though it would seem more fitting to revise our beliefs about it so we are able to identify members of a category if there is a category in the near vicinity.

We can thus see that if our concept of mental disorder turned out to have nominal, partial, or split reference then one could use this to motivate eliminativism. We also have concepts of particular kinds of mental disorder such as depression, obsessive-compulsive disorder, schizophrenia, autism, and the like. If one or more of these concepts turned out to have nominal, partial, or split reference then one could use this to motivate eliminativism about that particular kind or kinds of disorder. In the case of split reference scientists do sometimes distinguish between higher and lower categories and retain the concepts for the higher category. In the case of partial reference we would also not be forced to eliminate our concept, however, as we could instead revise our beliefs about the concept. Even if there is full reference where we are fairly easily able to identify individuals who are in fact instances of a category there could still be grounds for eliminativism, however. In the rest of the seminar I want to consider the different kinds of categories that could be relevant referents for our concepts of mental disorder and particular kinds of mental disorder and see which of these could lead us to eliminativism about our concepts. What is the purpose of a taxonomy?

Trees and shrubs and grasses in the gardening store. Interested in local conditions only. Would be depressing if conceptual analysis was more like this than science.

Addiction and psychopathy? How much do they share with other instances of mental disorder that are more clearly paradigmatic cases? Might always

remain fuzzy as the example revealed.

The problem here is that whether these conditions are labelled mental illnesses or not has important implications for whether these people are treated or jailed, whether health insurance companies are required to provide treatment or not, whether we are able to discriminate against these people or whether they are covered by mental health laws. It would seem to me that the relationship between mental disorder and right to treatment, moral responsibility, and legal responsibility is a separate issue really It is far from clear that these things are part of the concept or if they are connected so as to feature into the Carnp conditional then this is importantly different (there aren't facts aside from our social practices). What is left to argue about how our social practices should be. For example, it could be possible to proclaim that addiction is a mental disorder and yet addicts should be prosecuted. The interest in these being mental disorders seems to be around social and legal responsibility. We already know these come apart. An anxious person is responsible for murder Dunno The answer to these questions will come from a compex interrelationship of honing our intuitions and empirical investigation. It is nice that people are doing the conceptual analysis thing and it is important to not end up with a brain storm of features where some are redundant or fairly irrelevant but by the same token it is important not to make the issue out to be too black and white and it is also important not to isolate part of the project off from the whole

Implications for sociopathy and addiction. How many features do these conditions share with paradigmatic mental disorders and paradigmatic non-mental disorders? How much do mental disorders really have in common? Problem with the data in that the models seem to assume rather than discover irrationality etc concern about stipulated malfunctions. Natural Kinds - Griffiths maintains that emotions don't form a natural kind. - He maintains that there are three different kinds of emotion. - Primary Emotions (Ekmans Affective Response Programs) That share an underlying essence - Secondary Emotions (Socially Constructed Emotions) - Socially Sustained Pretenses - Prinz maintains that they have something in common in that

they are all brain states whose function is to register body state changes whose function is to represent core relational themes. - Very abstract level of analysis - Basically we can have reference (water, lion) - Split reference (greenstone turned out to be jadeite and nephrite) - Some other kind? Divided? Where we don't have natural kinds. - Some have denied that there are any such things as mental illnesses When they make this claim they are denying that people who we judge to be mentally ill have anything interesting in common People who deny that certain kinds of mental disorders are real seem to be disputing their status as natural kinds (e.g., ADHD) or disputing that there is anything wrong (homosexuality, political dissent etc) Folk Psychology. Behaviours (classifying symptoms, intuitively disordered) Cognitive Psychology (theory of mind deficit, language processing) Neuropsychology (structural abnormality, neurotransmission abnormality) Genetics (markers) Evolutionary Psychology (gives us the function of cognitive and neurological mechanisms. Turns out to be the relevant process for explaining our folk psychological judgements of mental disorder) Developmental Psychology (Development Attachment Neurodevelopmental disorders autism, schizophrenia, personality disorders, social / environmental can come to be represented in the brain. Environment, e.g., virus in the third trimester. Cerebral injury).

Chapter 10

Looping, conclusions, role

In this chapter I want to look at: - Non-natural kinds. Can consider whether anything much follows from this (doesn't seem to matter for the things that concerned us) - decisions we can make with respect to depth, spread, necessary, contingent - What does hang on our decisions? Prevalence, stigma etc.

(Griffiths, 1997; Hacking, 1995) How do aetiology, genetics, neurology, cognitive, psychological, phenomenological, behavioural, sociological, evolution of symptoms over time, and interventions that we have identified relate to disorder? Are they part of the identity conditions or are they merely contingent? Ill point out some problems with making evolution of symptoms essential when we are dealing with human beings.

Dominic Murphy in his book psychiatry in the scientific image has recently stated that in order to progress as a science psychiatry needs to move beyond purely behavioural symptoms and look to the cognitive neurosciences for the causal mechanisms that sustain the behavioural symptoms of psychiatric disorder. I agree with him in this, but I think that not all such causal mechanisms are internal to the agent. While Murphy does consider the rule of social causal mechanisms I think that there is a lot more work to be done on this.

Essential Kinds are thought to be categories that share the same intrinsic, or non-relational essential properties. Paradigmatic examples include water and gold where in order to count as an instance of water the instance must have the property of being H2O and in order to count as an instance of gold the instance must have the property of being atomic number 79. The intrinsic properties are thought to be constitutive of kind membership. Mental disorders could turn out to be essential kinds if it was found that they had a very specific biochemical basis, for example.

Biological Kinds. Are thought to categories that share the same relational, extrinsic essential properties of historical lines of descent. Paradigmatic examples include elms and tigers. There is controversy over whether natural kinds are required to have intrinsic essential properties such that biological kinds don't count as natural kinds; or whether biological kinds are natural kinds and thus natural kinds may have extrinsic, relational essences; or whether membership of a lineage is an internal property to the species as a whole and thus biological kinds are intrinsic essential kinds and thus natural kinds after all. I shan't get caught up in this debate, however. Whether biological kinds are properly thought of as natural kinds or not it seems that they form something of a natural category.

The notion of a natural category is tied up with the notions of generalisability, projectability, and predictive leverage. Natural categories may be thought of as something along the lines of what Boyd calls a homeostatic property cluster. The notion here is that certain properties are found to be clustered together in nature. If we see some properties then we can infer the presence of other properties and thus homeostatic property clusters support scientific generalisations and predictions. We would seem to identify instances of a natural category on the basis of these observable properties. The category of birds, for example, includes such properties as flight and feathers where these properties are superficial properties rather than properties at a lower level of analysis such as genetic. This view seems to be very much in line with the way the DSM provides behavioural symptoms as relatively superficial observable properties that enable clinicians to identify individuals as

having a certain kind of disorder. The majority of diagnoses also do not have essential symptoms and thus members of diagnostic categories exhibit family resemblances of symptoms. A feature of the property cluster view is that different instances have slightly different features and they may be more or less prototypical, for example, not all birds can fly.

While the DSM provides a nosology where clinicians identify mental disorder on the basis of behavioural symptoms it would seem to be a separate issue whether mental disorders are constituted or defined by the behavioural symptoms as Behavioural Kinds, however. If one takes the behavioural symptoms to be definitional or constituitive then there could plausibly be borderline cases where it is indeterminate whether the individual is in fact a member of the kind or not. It would seem, however, that the main reason why it is that certain properties are to be found clustered together in nature is because they share some underlying causal mechanism that are responsible for the properties homeostasis. It is because the causal mechanism is found in the different instances that we are able to make scientific generalisations and predictions. It could also turn out that the same set of behavioural symptoms could be generated in two quite different ways. If we found this to be the case then it would seem better to conclude that there are two distinct kinds of disorders where different interventions are required.

Thus, while we might typically identify or come to believe that instances are members of a certain category on the basis of superficial, observable properties, taxonomy is often revised as we come to define categories on the basis of the underlying causal mechanisms that are necessary for category membership. This is because causal mechanisms seem to be what leads to the properties homeostasis and the more homeostatic a property cluster the more those properties are able to support generalisations and predictions. While Boyd's view focused on internal generative mechanisms it is unclear whether a principled distinction between internal and external generative mechanisms can be sustained. If one views a species as an individual, for example, then lineage would be an internal property to the species. Boyd's homeostatic property cluster view, or something like it, can thus be thought

of as consistent with both the essentialist and relational view of categories.

Wakefield attempts to draw a principled distinction between the right and wrong kind of causes for mental disorder. He maintains that when the harmful behaviours are due to inner malfunction the individual is mentally disordered and when the harmful behaviours are the result of external causal mechanisms the harmful behaviours are not indicative of mental disorder and are instead best thought of as a non-pathological problem in living. It would seem that whether mental disorders are constituted by social causal mechanisms would be an empirical matter rather than one to be settled on intuitive grounds or by stipulation, however. Wakefield is especially focused on the notion of neurological and / or cognitive malfunction which he characterises along the lines of a hardware / software distinction and while he doesn't mention it I don't think he would be opposed to adding genetic malfunction to the mix (supposing that it makes sense to talk of genetic malfunction or kinds of genetic disorder). This way of thinking about inner malfunction seems very much in line with cognitive neuropsychology and it might be the case that the kinds of psychiatric disorder are derived as malfunctions of the causal mechanisms that is identified, at lest in part, by cognitive neuroscientists. Neurological kinds would seem to be fairly straightforwardly thought of as biological kinds. Some theorists have attempted to analyse Psychological kinds as another variety of biological kinds where mental or cognitive states such as belief and desire are the kind of state they are in virtue of what the mechanisms that support the state have evolved to do.

Sometimes theorists (like Wakefield) appeal to current functions instead of evolutionary functions where the effects of a current function are responsible for the mechanism being prevalent in current populations. Treating mental kinds as biological kinds is controversial, however. The natural categories or kinds would seem to be those of normal functions. Psychiatric kinds are breakdowns of normal symptoms and the breakdowns may be unified only by being break- downs of a specific mechanism. There would thus seem to be an open ended class of ways things could go wrong. Attempting to list them all with respect to behavioural symptoms is thus bound to get

unwieldy and more progress might be made by looking at different ways that normally functioning systems can break down. I now want to turn to some of the external mechanisms that might be relevant for mental disorder and Ill consider several different varieties of socially constructed kinds.

Artefacts like pens and chairs are paradigmatic examples of Socially Constructed Kinds. Instances of the category pens count as members of the category in virtue of having the historical relational property of being designed by an agent for a certain function. As such agents designing them for a certain function is necessary and sufficient for or constitutive of category membership. Because they are designed by agents for a certain function pens exhibit a cluster of superficial properties in common. Those properties may enable us to identify instances as instances of the category. If we found something that shared the superficial properties with pens but it grew on a tree or materialised out of a swamp then because it was not designed by an agent with the relevant intention it would not count as a pen, however. While pens are dependent on us for their initial existence once the instances have been brought into being then it is a mind independent fact that the instances are in fact members of the category. Even if we lost our concept of a pen or we no longer used pens to perform their function the instances that still exist would continue to exist as members of the category.

Some other socially constructed kinds aren't dependent on the intentions or mental states of agents so much as their social practices. Something might count as a doorstop, for example, not because it was designed with that intention in mind, but instead because it is currently being used to perform that function. If we accept this reading of what it is to count as a doorstop then it would follow that if we were to stop using the object as a doorstop that it would cease to be a member of that kind. There isn't a science of pens or doorknobs. While we might be able to make generalisations such as that pens usually have ink and that doorstops tend to be sturdy or obstructive it would seem that there are significantly less generalisations and predictions available to us than there is with either chemical or biological kinds.

I now want to turn to another sort of socially constructed kind that is clearly

more relevant to psychiatric disorder. The notion of a Looping Kind was initially introduced by Hacking and it has subsequently been picked up on by other authors such as Griffiths, Mallon, and Murphy. In order to describe the features of looping kinds I need to draw a further distinction between what I shall call explicit looping kinds and implicit looping kinds.

Explicit looping kinds are kinds that are constituted by our social practices. While artefacts like pens are mind independent in the sense that they continue to be pens in the absence of our social practices around them, looping kinds are thought to be causally rather than definitionally or constitutively dependent on our social practices. Our social practices cause them to come into being as instances of the category and if our social practices change then this can cause them to go out of being as instances of the category. It is easiest to see this by way of example. Members of Parliament and Licensed Dog Owners are examples of explicit looping kinds. We have social practices around parliament and the election of members of parliament, for example, and in virtue of those social practices individuals come to be Members of Parliament. Unlike pens explicit looping kinds aren't independent of our social practices because if we alter our social practices so that there isn't a parliament then the individuals would cease to be members of the category Members of Parliament.

Individuals that are Members of Parliament have properties in common such that they may be identified as Members of Parliament. We are able to make generalisations and predictions about Members of Parliament with respect to the properties they exhibit or are likely to exhibit and ways in which they are likely to behave. When the individuals are no longer members of the category Members of Parliament then they lose the properties that they had in virtue of their category membership, however, and we can no longer make such generalisations and predictions about them. These looping kinds are explicit in the sense that we are aware that the categories are dependent on our social practices. We know that there wouldn't be any Members of Parliament if we altered our social practices in certain ways. This doesn't stop us being able to make generalisations and predictions about Members

of Parliament, however. It also doesn't stop the special science of politics from taking them seriously as a category.

Implicit looping kinds are similar to explicit looping kinds except that in this instance we aren't explicitly aware that the instances of the category are instances of the category because of our social practices and instead we regard the category as being a natural (or biological) kind. Hacking maintains that in this case if we were to become aware of their status as a looping kind then it would be inevitable that our social practices would change and this would have the result that the instances would no longer being members of the category. Our awareness and subsequent change in our social practices would also result in an alteration to the properties that the individuals shared as members of the category and thus the generalisations and predictions that were made about individuals in virtue of their category membership would no longer obtain.

Once again, it is probably best to convey this phenomena by way of example. Examples of implicit looping kinds include categories such as demonic possession and being possessed by a wild pig. The notion is that when we believed in these concepts then our belief in them and our social practices around them results in opening up new ways of behaving that are stereotypic of the category. If we take a person to be a member of the category or if they take themselves to be a member of the category then this may cause them to behave in ways that are stereotypic of the category. Members of the category are thus able to be identified as members of the category in virtue of sharing certain stereotypical properties in common. What is supposed to be distinctive about these categories, however, is that they cannot survive our realisation that they refer to looping kinds. The notion is that once we become aware that the properties are due to our social practices then we cease believing in them and we inevitably alter our social practices so that the individuals no longer display those common features. This phenomena is probably best conveyed by way of Ian Hackings characterisation of Multiple Personality Disorder which he takes to be an all too perfect illustration of the feedback effect in implicit looping kinds:

> We tend to behave in ways that are expected of us, especially by authority figures doctors, for example. Some physicians had multiples among their patients in the 1840's, but their picture of the disorder was very different from the one that is common in the 1990's. The doctors vision was different because the patients were different; but the patients were different because the doctors expectations were different. That is an example of a very general phenomenon: the looping effect of human kinds. People classified in a certain way tend to conform to or grow into the ways that they are described; but they also evolve in their own ways, so that the classifications and descriptions have to be constantly revised. (Hacking, 1995, p. 21).

Hacking thus maintains that in the case of implicit looping kinds there is a tension in that possession of the concept and our social practises around this are the mechanism that both stabilises and destabilises the property cluster. With respect to the stabilising function he considers that individuals symptoms are shaped because when the clinician applies the concept to the patient this results in the clinician having either implicit or explicit expectations of the symptoms they expect to find in the patient. This changes the way that the clinician relates to the patient and is thought to lead to the patient exhibiting the symptoms they are expected to exhibit. Another way this can happen is if the clients apply the concept to themselves and thus come to exhibit symptoms that they believe to be stereotypic features of the category. In this way the concept and our social practices stabilise the symptoms that the patient exhibits as they come to behave in ways that are consistent with the stereotype.

Hacking also considers how our social practices can have a destabilising effect, however. He traces how the stereotypical features of Multiple Personality Disorder have evolved through time. Hacking tells a complex story of destabilisation and he draws on a variety of factors including political and theoretical, which lead to our beliefs about the concept evolving and the symptoms evolving in response to this. Some examples he has of this effect

in the case of MPD include how many alters are thought to be typical (one or several or over one hundred); whether there is one or two way amnesia; how long it takes to switch between alters; and reports of abuse. It thus seems that the change seems mostly to be a function of a change in the theoretical views of clinicians. This led to a subsequent change in how they related to their clients and what kinds of symptoms they expected to see. Hacking seems to regard implicit looping kinds as having some homeostasis but the homeostasis is less stable than other kinds of socially constructed and natural kinds in that awareness of their status as looping kinds will result in the dissolution of the category.

Implications of Implicit Looping Kinds for a Scientific Nosology.

In these cases because it is implicit that we are dealing with a looping kind we are unaware of the impact of categorisation, our social practices, our expectations, our ways of interacting with the person, and so forth. If we come to believe that a certain kind of mental disorder is a looping kind then it seems that one of three things could happen: Firstly, it could turn out to be the case as an empirical matter of fact our change in belief does not result in a change in our social practices. While Hacking thinks the relevant social practices are ones that invariably would change if we became aware that the category was a looping kind surely it could be possible that the social practices that are sustaining the phenomena could be resistant to change possibly because they have other beneficial effects. It is unclear whether Hacking would consider this to be an example of an implicit looping kind because it was implicit even though awareness did not result in its dissolution or whether Hacking would consider this to be an example of an explicit looping kind because it does not dissolve in the face of our awareness even though the so called explicit looping kind was implicit for a time.

Secondly, it could turn out to be the case that as an empirical matter of fact that if we came to believe the category was looping and we changed the relevant social practices the stereotypical behavioural features remain. In this case we seem to be left having to conclude that the category wasn't a looping kind after all. While it could still be socially constructed in the sense

that artefacts similarly rely on us for their initial existence the phenomenon wouldn't seem to be dependent on our social practices and thus it would not be an implicit looping kind on Hackings account. The third thing that could happen would be that our awareness of the category as an implicit looping kind could cause the stereotypic features to shift. If we found that a particular kind of mental disorder was an implicit looping kind this isn't to say that all instances of the category are suddenly cured of all symptoms of psychopathology, however. It is just to say that they wont display features of psychopathology that were stereotypic of the looping kind. They may well go on to display stereotypic features of another psychiatric kind, for example. Social constructionists about Multiple Personality Disorder often say that there is no such category as Multiple Personality Disorder there is only Borderline Personality Disorder that has been worked up into Multiple Personality Disorder in response to our social practices around the concept. The notion here seems to be that if we refuse to participate in those social practices the patients will display stereotypic features of Borderline Personality Disorder instead.

What is unclear, however, is whether this would be so because the clinicians expect them to come to display the stereotypical features of Borderline Personality Disorder or whether this is in response to some other mechanism. If clinicians came to believe that there was no such category as Borderline Personality Disorder then would the individuals continue to behave in a way consistent with a diagnosis of Borderline Personality Disorder or would their behavioural symptoms shift so that they met criteria for another diagnostic category? While Multiple Personality Disorder is often one of the favourite categories of those who maintain that we need to look at social causal mechanisms it is unclear whether other, more paradigmatically biological psychiatric kinds could turn out to be looping kinds or to have a looping kind feature to their behavioural symptoms. It could turn out to be the case that mental disorder more generally has a significant looping kind component.

If this was found to be the case then this would seem to have significant implications for both the project of how we identify mental disorders and the

project of how we develop a scientific classification of them. One implication is that focusing solely on behavioural symptoms might be counter-productive. Each subsequent edition of the DSM is praised for making scientific progress with respect to providing categories that better support generalisations and predictions. If the properties relevant for generalisation and prediction are purely behavioural symptoms and if the behavioural symptoms evolve over time in response to the classification system and a new round of expectations by clinicians then it would seem that the DSM approach will be limited insofar as the property cluster is unstable. The DSM may not only describe current symptomatology but it also may have a causal role to play with respect to future symptom development. One consequence of this might be that the DSM and ICD aren't necessarily converging on constructs that are more valid than the old constructs; rather each edition might recover some of the construct validity that the old one had by adequately capturing present symptoms that may, at least partly, have been evoked in response to previous systems of classification. Construct validity on the basis of generalisations and predictions on the basis of behavioural symptoms may be of limited value with respect to a scientific nosology.

If we identify kinds of mental disorders according to causal mechanisms rather than behavioural symptomatology, however, then this enables us to say that the behavioural symptomatology of a particular kind of disorder can evolve over time. This latter approach also allows that there could be considerable cross-cultural variation in the behavioural symptoms of individuals who have the same kind of mental disorder. While the DSM saw purely behavioural symptoms as progress from the causal mechanisms offered by the psychodynamic theorists cognitive neuropsychology would seem to have good prospects for grounding the next stage of scientific development from observational properties towards a scientific nosology of the causal mechanisms that produce psychiatric disorders. It seems plausible to me that more valid constructs may require us to incorporate causes from multiple levels of analysis. While there will be more to social causes than the looping effects that Hacking deals with the looping kind effect is interesting with respect to the

relationship between social cognitive and behavioural facts. If we consider that the cognitive facts are represented within the brains of individuals it seems that whether the cause is inner or outer may be a function of how far back in the causal chain we look.

Implications for Problematic Cases like Addiction and Sociopathy

What is the purpose of a taxonomy?

Trees and shrubs and grasses in the gardening store. Interested in local conditions only. Would be depressing if conceptual analysis was more like this than science.

Addiction and psychopathy? How much do they share with other instances of mental disorder that are more clearly paradigmatic cases? Might always remain fuzzy as the example revealed.

The problem here is that whether these conditions are labelled mental illnesses or not has important implications for whether these people are treated or jailed, whether health insurance companies are required to provide treatment or not, whether we are able to discriminate against these people or whether they are covered by mental health laws. It would seem to me that the relationship between mental disorder and right to treatment, moral responsibility, and legal responsibility is a separate issue really It is far from clear that these things are part of the concept or if they are connected so as to feature into the Carnap conditional then this is importantly different (there aren't facts aside from our social practices). What is left to argue about how our social practices should be. For example, it could be possible to proclaim that addiction is a mental disorder and yet addicts should be prosecuted. The interest in these being mental disorders seems to be around social and legal responsibility. We already know these come apart. An anxious person is responsible for murder. Don't know.

The answer to these questions will come from a complex interrelationship of honing our intuitions and empirical investigation. It is nice that people are doing the conceptual analysis thing and it is important to not end up with

a brain storm of features where some are redundant or fairly irrelevant but by the same token it is important not to make the issue out to be too black and white and it is also important not to isolate part of the project off from the whole.

Implications for sociopathy and addiction.

How many features do these conditions share with paradigmatic mental disorders and paradigmatic non-mental disorders? How much do mental disorders really have in common? Problem with the data in that the models seem to assume rather than discover irrationality etc concern about stipulated malfunctions.

Decisions we made around the criteria have consequences for the under / over inclusiveness of categories. Once we realize that is is problematic whether there is a categorical feature to nature such that we get things right or wrong. Once we appreciate some of the subtlety of the situation then we can be more nuanced. Multiple personality (and the sciences of memory). We can cast the net broadly or narrowly. This has consequences for seriousness. institutionalization. medication. and so on. the answers to these questions is dependent on how to choose to id the individuals to start with. the literature on sociopathy. different ways of defining cast it narrow or broad. cast it broad and study undergraduates. but then problematic relationship to the most serious (which is very rare).

References

References

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: American Psychiatric Association.

Bentall, R. P. (2003). Madness explained: Psychosis and human nature. London: Penguin Books.

Compas, B. E., & Gotlib, I. H. (2002). Introduction to clinical psychology science and practice. McGraw-Hill Higher Education.

Cooper, R. (2005). Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders (Vol. 86). Netherlands: Springer.

Cooper, R. (2007). Psychiatry and the philosophy of science. Acumen.

Davidson, G. C., & Neale, J. M. (2001). Abnormal psychology (8th ed.). John Wiley & Sons.

Davies, P. S. (2000a).Malfunctions. Biology and Philosophy, 15, 19-38.

Davies, P. S. (2000b). The nature of natural norms: Why selected functions are systemic capacity functions. Nous, 34(1), 85-107.

Davies, P. S. (2001). Norms of nature: Naturalism and the nature of functions. MIT Press.

Ellenberger, H. F. (1970). The discovery of the unconscious: The history and evolution of dynamic psychiatry. BasicBooks a Division of Harper-Collins.

Fulford, K. (2000). Teleology without tears: Naturalism, neo-naturalism, and evaluationism in the analysis of function statements in biology (and a bet on the twenty-first century). Philosophy, Psychology, & Psychiatry, 7(1), 77-94.

Griffiths, P. (1997). What emotions really are: The problem of psychological categories. The University of Chicago Press.

Hacking, I. (1995). Rewriting the soul: Multiple personality and the sciences of memory. Princeton University Press.

Murphy, D. (2006). Psychiatry in the scientific image. The MIT Press.

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

Murphy, D., & Woolfolk, R. L. (2000b). The harmful dysfunction analysis of mental disorder. Philosophy, Psychology, & Psychiatry, 7(4), 241-252.

Sadock, B. J., & Sadock, V. A. (2003). Kaplan & Sadock's synopsis of psychiatry (9th ed.). Lippincott Williams and Wilkins.

Shorter, E. (1997). A history of psychiatry: From the era of the asylum to the age of prozac. John Wiley & Sons.

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

Wakefield, J. C. (1992b). Disorder as harmful dysfunction: A conceptual critique of DSM-III-R's definition of mental disorder. Psychological Review, 99(2), 232-247.

Wakefield, J. C. (1993). Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. Journal of Abnormal Psychology, 102(1), 160-172.

Wakefield, J. C. (1999). Evolutionary versus prototype analyses of the concept of disorder. Journal of Abnormal Psychology, 108, 374-399.

Wakefield, J. C. (2000a). Aristotle as sociobiologist: The "function of a human being" argument, black box essentialism, and the concept of mental disorder. Philosophy, Psychology, & Psychiatry, 7(4), 253-269.

Wakefield, J. C. (2000b). Spandrels, vestigal organs, and such: Reply to Murphy and Woolfolk's "the harmful dysfunction analysis of mental

disorder". Philosophy, Psychology, & Psychiatry, 7(4), 253-269.

Wakefield, J. C. (2003). Dysfunction as a factual component of disorder. Behaviour Research and Therapy, 41, 969-990.

Wakefield, J. C. (2004). The myth of open concepts: Meehl's analysis of construct meaning versus black box essentialism. Applied and Preventative Psychology, 11, 77-82.