The Nature of Mental Disorder

There has been a lot of controversy over both the nature of mental disorders in general and also over the particular kinds of mental disorders that there are. The biggest threat to the prospects for a science of psychiatry is eliminativism where eliminativists maintain that we should eliminate our concept of mental disorder, as there is no such thing. One could also be an eliminativist about particular kinds of mental disorder that appear in psychiatric nosology. This variety of eliminativism would be less radical, however, as it is plainly the case that current nosology is a work in progress and it wouldn't seem to undermine the notion that there are categories of mental disorder it is just that we haven't hit upon them at present. There are a variety of ways that one could be led to eliminativism. In this seminar I shall begin by talking about our concept of mental disorder and then turn to different ways our concept of mental disorder and concepts of particular kinds of mental disorder could turn out to map onto the world. There may be grounds for eliminativism here if our concepts don't map onto categories. I shall then turn to investigating different kinds of categories where there are different causal mechanisms that are responsible for generating the properties that make the category useful for scientific generalisation and prediction. It might be the case that social mechanisms are one such mechanism but that by their nature this results in categories that are less stable than the traditional sciences. If this turned out to be the whole story about mental disorder and there weren't any genetic, neurological, or cognitive causal mechanisms that were relevant then one might be led to eliminativism. If this is just part of the story, however, then the science of psychiatry would need to be reformed so as to investigate and better incorporate these social causal mechanisms.

The Concept of Mental Disorder

Defining mental disorder is problematic. The first problem is how mental disorders differ from other varieties of disorders. The second problem is how disorders differ from non-disorders. Another related issue is how we recognise or identify whether an individual is mentally disordered. I shall now address each of these issues in turn.

With respect to the first problem, the issue is sometimes put as the problem of distinguishing mental or psychiatric disorders from non-mental or neurological disorders. 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 psychiatric, however. This move seems to be in line with our common-sense intuitions though it is in tension with our intuition that mental disorders are disorders of cognitive processes as vision would be a paradigmatically cognitive or mental process. Indeed, other visual disturbances such as hysterical blindness and hallucinations are typically regarded as psychiatric rather than neurological. The concept of mental that is employed in both common sense and in current nosology (or taxonomy) thus seems to be under-inclusive. Current nosology might also be thought of as over-inclusive, however. For instance, the essential feature of Tourette's is tics but there wouldn't seem to be anything particularly mental or cognitive about a motor disturbance. Perhaps Tourette's really has an essentially cognitive component that is neglected by current nosology, or perhaps Tourette's is not appropriately classified as a mental disorder and current nosology is over-inclusive with respect to this case.

With respect to the second problem the most influential view of disorder is probably Wakefield's 'Harmful Dysfunction' (HD) analysis of the concept of disorder as it is employed in psychiatry, medicine, and common sense. Wakefield maintains that there are two individually necessary and jointly sufficient conditions for someone having a disease, disorder, or illness. The first condition is that there is an inner malfunction and the second condition is that the effects of the inner malfunction are harmful to the person and / or to society. Wakefield's account is controversial, however. While he takes the notion of malfunction to be determined by facts about biological function as talked about by theorists such as Millikan and Neander other theorists have objected that the notion of malfunction is dependent on our value judgements that the individual's behaviour is in violation of norms of society. Another point of controversy is over whether disorders must be due to inner malfunction. These are issues that I shall return to in a later section.

If we now turn to the issue of how we recognise if an individual is mentally disordered it may well be that our concept of mental disorder is a cluster concept

where there are several features that are relevant but where none of the features are necessary and where there is no clear boundary on how many features are sufficient for our regarding an individual to be mentally disordered. In textbooks on psychopathology the following features are commonly listed: 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 of worries constantly about his financial situation'. The Clinician's handbook *The Diagnostic and Statistical Manual of Mental Disorders* similarly attempts to define mental disorder by incorporating all of the above features and then conceding that no definition seems adequate to capture the phenomena. While Wakefield attempts to defend his 'Harmful Dysfunction' analysis arguing that it **is** an adequate analysis of our concept of disorder and that the DSM should revise its definition so it is in keeping with his account the DSM might be less interested in defining mental disorder and more interested in training clinician's to agree in how to classify individuals.

The DSM lists behavioural symptoms for each category and when a person meets a specified number of them they meet the diagnostic criteria for that disorder. Ian Hacking maintains that even more important than the DSM symptom lists is the accompanying case book that provides case studies of people who are prototypical instances of someone meeting a certain diagnostic category. Clinical judgement is thought to consist of experience with a variety of more or less prototypical cases so that a clinician's judgement falls in line with the judgement of other health professionals.

We also have our common sense conception of mental disorder. Our common sense conception seems to be similarly formed around exemplars of people who are considered to be mentally disordered where these exemplars form something of a prototype or stereotype. Prototypes seem to play a significant role in our intuitive judgements as to who is and who is not mentally disordered and also in what kind of mental disorder they have. Our common sense intuitions seem to have evolved as our conception becomes more informed by the categories of mental disorders offered by the DSM.

My main reason for dwelling on our conception of mental disorder is so we are in a better position to assess when we should be eliminativists about our concept of disorder when we see how the world turns out. I now want to say a few words about how our concepts can map onto categories and then I'll turn to the main issue of this paper: the issue of the different kinds of categories that mental disorders could turn out to be. Along the way I'll consider varieties of reference and categories that could lead one to be an eliminativist about our concept of mental disorder in general or about our concept of a particular kind of mental disorder.

Different Kinds of Reference

I won't 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 science.

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: jadite 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 concept of mammals. This is why it is important not to get too caught up in conceptual analysis when one is interested in the nature of the world. Another way that partial reference could go would be if our concept referred to a category but also 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 diagnosed 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 they were revised so as to allow us to 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 concept so it falls in line with a category if there is one 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 conception so that it did refer to a category.

Even if there is ***full reference*** where the concept fairly straightforwardly refers to 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.

Kinds of Categories: Part One

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. While essential kinds aren't particularly relevant for psychopathology it is easier to understand other categories by way of contrast.

A kind of category that would seem to have better prospects for psychopathology would be the category of ***Biological Kinds.*** Paradigmatic examples of biological kinds include elms and tigers. I shall consider two different accounts that have been offered of the nature of biological kinds before considering how biological kinds could be relevant to psychopathology. The first account is probably the most widely accepted and on this account the essential properties for biological kinds are thought to be relational, extrinsic properties of historical lines of descent

Mallon considers a second account of biological kinds. He maintains that some theorists have regarded biological kinds as homeostatic property clusters. 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. Some theorists maintain that the reason why certain properties are found to be clustered together in nature is because they share some underlying causal mechanism that is responsible for the properties homeostasis. It is because the causal mechanism is found in the different instances that results in our being able to generalise and make accurate predictions and if there were no common causal mechanism then we wouldn't have the generalisation and prediction power that we do. This move would seem to lapse back to either essentialism or the relational historical view, however, depending on whether one attempted to cash out the relevant mechanism as intrinsic or extrinsic. It might be the case that intrinsic mechanism result in the tightest property cluster as in chemical kinds, for example, while extrinsic or relational causal mechanisms result in weaker property cluster as in biological kinds for example,

though these property clusters are still useful for science. Woolfolk considers that some theorists maintain that biological kinds are homeostatic property clusters even though they do not share intrinsic or relational essential properties. These theorists aren't eliminativists about biological kinds, however, as they maintain that the biological property clusters are still useful for science with respect to generalisation and prediction.

There are a couple of interesting features of this latter view. The first is that it seems to be very much in line with the way the DSM carves up different kinds of mental disorder. I have already said how the DSM provides a number of behavioural symptoms and in order to meet criteria for a given disorder the person must meet some specified number of those symptoms. The majority of diagnoses do not have essential symptoms and thus members of diagnostic categories exhibit family resemblances of symptoms. This is similar to the property cluster view in that there can be some variation or family resemblance in the properties exhibited by individual members of the category. A feature of the property cluster view is that different instances have slightly different features and they may be more or less prototypical. Not all birds can fly, for example. There can also be borderline cases where it is unclear whether the instance is in fact an instance of the category or not.

Another feature of interest is that the properties that are of interest in the DSM are behavioural symptoms. One might thus consider the DSM to be treating mental disorders as ***Behavioural Kinds*** where each kind of disorder is a property cluster of behaviours. Theorists who adopt the property cluster view as I have outlined it about biological kinds often take similarly superficial, observable properties to be the relevant properties while remaining agnostic as to the underlying causal mechanism. 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.

A problem that one might have with the homeostatic property cluster view as I have outlined it is that causal mechanisms seem to matter. We would like to know *why* it is that these properties are found clustered together and *what* caused the features or symptoms to maintain homeostasis. While an important part of science involves

observation and description, as science progresses it starts to develop theories of the causal mechanisms responsible for the phenomena. Often our taxonomy is revised as we delineate kinds on the basis of causal mechanisms rather than superficial similarities. This is because if we classify on the basis of causal mechanisms the properties are likely to be better suited to generalisations and predictions.

Wakefield, in particular, pushes the intuition that causation matters and that even if there is a cluster of behavioural features if the 'right kind' of causes are absent then we do not regard the person as mentally disordered. One example he offers is a case of a person meeting DSM criteria for reading disorder. He maintains that if we find the person can't read because nobody ever instructed him how to read that we wouldn't regard the person as mentally disordered. If we found that the person had received adequate instruction and yet could not read because of some kind of inner malfunction, however, then we would conclude that the person was mentally disordered.

Wakefield thus characterises the 'right kind' of causes to be ones that are internal to the person. He 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). ***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. In many respects they have more to do with some of the other varieties of kinds I want to consider: Socially constructed kinds.

Kinds of Categories: Part Two

Artefacts like pens and chairs are paradigmatic examples of ***Socially Constructed Kinds.*** Instances of the category pens could be thought of as instances of the category in virtue of having the relational property of being designed by an agent for a certain function. As such agents designing them for a certain function is a necessary and sufficient cause for category membership. Alternatively, one could characterise token pens as being instances of the category in virtue of having a cluster of properties in common. We certainly identify pens on the basis of these properties. The cluster of properties that the instances have in common in virtue of their being designed by agents allow for generalisations and predictions to be made about pens, though it might be that there are less of these available to us than there are about chemical kinds or biological kinds. Socially constructed categories are distinctive from those other categories in the sense that people's intentional states are a necessary cause of the instances. Once the instances have been brought into being, however, then it is a mind independent fact that the instances are in fact members of the category. Even if our social practices changed so that we no longer used pens to write with or even if we lost our concept of a pen so that we couldn't identify instances as pens the instances of the category would continue to be instances of the category in virtue of their being designed by an agent with a certain intention. As such the intentions of agents play a necessary causal rather than a constituent role with respect to category membership.

While pens and chairs are socially constructed categories it is hard to see why someone would want to be an eliminativist about our concept of a pen. While someone might want to eliminate pens in the sense of eliminating instances of the category and making people write with pencils or crayola crayons this is not a case of eliminativism about our concept PEN. If we find that a category is socially constructed we are not thereby required to be eliminativists. It is useful to have this notion of a socially constructed category as a backdrop for understanding some of the other socially constructed categories that might seem more relevant for mental disorder.

The notion of a ***Looping Kind*** was introduced by Hacking and this notion has subsequently been picked up on by other authors such as Griffiths and Mallon. 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 dependent on our social practices in the sense that the instances wouldn't have existed as instances of the category if our social practices had been different in certain respects from what they were. They are thus constituted by our social practices and they are different from artefacts in the sense that if we altered our social practices in certain ways then the instances would no longer share the properties that are characteristic of their category membership. It is easiest to see this by way of example. Members of Parliament and Licensed Dog Owners are examples of explicit looping kinds. The category Members of Parliament relies on our social practices not only with respect to the instances sharing properties in common but also with respect to the instances continuing to share properties in common. The category is constituted by our social practices. If we altered our social practices so that we no longer had parliament, for example, then while the instances of the category would continue to exist they would lose the properties that are relevant for membership in that category.

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. While one might well want to eliminate the instances as instances of the category Member of Parliament by blowing up either parliament or politicians, for example, this is not eliminativism about our concept of Member of Parliament. People could similarly want to eliminate mental illness by curing it or by eugenic policies but this is also not the relevant notion of eliminativism. The relevant notion of eliminativism would be to advocate that we eliminate our concept because Members of Parliament are dependent on the continuation of our social practices and that causal mechanism is the wrong kind of causal mechanism for categories. We accept this same causal mechanism in the case of pens, however, and Members of Parliament share properties in common that are useful for the special science of politics and thus we are not required to be

eliminativists about explicit looping kinds. Indeed, eliminativism about members of parliament would seem to undermine politics as a special science.

Implicit looping kinds are similar to explicit looping kinds except in this instance we are unaware of their status as looping kinds and if we were to became aware of this then Hacking maintains our social practices would change and as a result the instances would no longer share the properties that support generalisation and prediction. Examples of implicit looping kinds include categories such as demonic possession and being possessed by a wild pig. The notion is that when our social practices legitimated these categories people came to behave in such ways and thus we could have pointed to the properties that instances of the category shared. Members of the category were identified as instances of the category because they shared certain properties in common. What is supposed to be distinctive about these categories, however, is that they cannot survive our realisation of their status as looping kinds. The notion is that once we realise that these individuals display their common features in virtue of our social practices then 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 Hacking's 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 maintains that in the case of implicit looping kinds there is a tension in that our social practises 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 are stereotypic features of the concept. 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.

This story seems to be causal, but it is also thought to be constitutive in the sense that Hacking maintains that if we become aware that a category is a looping category then this will lead to our changing our expectations and social practices and thus the properties will no longer occur. While our intentions were thought to be necessary causes of artefacts our intentions could alter and artefacts would continue exhibit the properties in virtue of which they are members of the category. Thus, while our intentions are necessary causes of the properties that artefacts have they are not sustaining causes of those properties. Our intentions can alter but pens continue to display the properties in virtue of which we can make generalisations and predictions about them. With looping kinds our concept seems to play a causal role once more but in this case it may be thought to be constitutive in the sense that our intentions and social practices are sustaining causes of the properties the instances have in common. If our intentions and social practices alter then the instances no longer exhibit the properties in virtue of which the instances were members of the category.

Hacking thus also considers how our social practices can have a destabilising effect. 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 the concept evolving and the symptoms evolving in response to the evolution of the concept. 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 clinician's. This led to a subsequent change in how they related to their clients and what kinds of symptoms they were interested in seeing. 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 perhaps because our concepts evolve much faster.

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, or 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 it 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 properties 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 maintained by our social practices.

The third thing that could happen would be that the defining properties of the category could shift so that there wouldn't be any properties that the instances shared that were of any use for scientific generalisation or prediction. In this latter case we would be left with a ***Nominal Category***. Nominal categories aren't really categories at all from

a scientific point of view and thus we would eliminate the concept from science. It thus seems that we will end up being eliminativists about implicit looping kinds if implicit looping kinds are kinds such that being aware of their status is enough to effect social change which is enough to destabilise the property cluster so that it is no longer scientifically fruitful. This is the way Hacking characterises them though it seems that those dimensions might be teased apart. 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 won't display features of psychopathology that were stereotypic of the looping kind. Eliminativists about Multiple Personality Disorder often say that there is no such thing 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 our concept of Multiple Personality Disorder. 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. What is unclear, however, is whether this would be so because the clinician's expect them to come to display the stereotypical features of Borderline Personality Disorder or whether this is in response to some other mechanism.

While Multiple Personality Disorder is one of the favourite categories of those who maintain we should be eliminativists 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 the behavioural symptoms. I want to end with a question: If looping kinds are the kinds that mental disorders often are, then what are the consequences for the science of psychiatry?