CHAPTER 2

ARE MENTAL DISORDERS NATURAL KINDS?

Whether a condition is a disorder is partly a value judgement, but the distinctions between types of disorder might still depend solely on psychological and biological facts. If this were the case then the domain of mental disorders would be analogous to the domain of weeds. Weeds are unwanted plants, thus whether a daisy is a weed is at least in part a value judgement. Still, the distinctions between kinds of plants generally considered weeds are fixed by the nature of the world. Botanical facts make it the case that daisies and thistles are genuinely distinct types of plant.

A fundamental assumption of the D.S.M. project is that empirical research can tell us how mental disorders ought to be classified. When the A.P.A. committees developed the D.S.M.-IV they reviewed thousands of empirical studies.[1] These studies examined matters such as the biochemical correlates of disorders, how people with different disorders respond to particular treatments, and whether a particular disorder disproportionately affects people of a certain age or sex. The assumption is that by examining all this data it will be possible to construct a classification system that at least approximately reflects the true natural similarities and differences between cases of mental illness. Whether the correct classification of mental disorders should be dimensional or categorical, whether there are three or five different types of schizophrenia, and whether there is such a thing as caffeine withdrawal syndrome are all problems that it is thought could potentially be solved by empirical research.

The similarities and differences between types of mental disease are assumed to be not only objective but also of great significance to psychiatric theory. This is why psychiatric research generally examines groups of patients with the same diagnosis; these patients are assumed to be similar in some fundamental way. It is supposed that fundamentally different pathological processes underlie different disorders, and that different disorders can best be treated in different ways.

Thus the A.P.A. can be seen as aiming to produce a classification system very much like those found in biology or chemistry. Like the differences between the chemical elements and biological species, the differences between types of mental disorder are thought to be objective and theoretically important. In short, mental disorders are assumed to be "natural kinds". "Natural kind" is a technical term used by philosophers to refer to the kinds of thing or stuff studied by the natural sciences. Sodium, fleas, dandelions, and electrons are all examples of natural kinds. Members of a natural kind are naturally similar to each other, and there is some explanation for this. Fleas, for example, are all similar in that they jump, drink blood, and are

[1] These studies are summarised in the D.S.M.-IV *Sourcebook*. Widiger et al. 1994, 1996, 1997.

poisoned by flea-spray, and fleas are alike in these respects because they are similar in some more fundamental way, plausibly because they are all genetically similar.

In this chapter I ask whether the D.S.M. committees are right to assume that types of mental disorder are natural kinds. This is an important question for several reasons. First, and as already mentioned, if mental disorders are natural kinds then this implies that the domain of mental disorders has a natural structure that it should be possible to discover via empirical research. In addition, whether mental disorders are natural kinds is important because natural kinds and natural laws are linked.[2] The behaviour of members of a natural kind is governed by natural laws. For example, it is a law that copper melts at 1083°C, and a law that Syrian hamsters have a sixteen day gestation period. As a consequence of being governed by laws, natural kinds can function in explanations (for example, "Miffy is afraid of dogs because she is a rabbit"), and support inductive inferences (for example, we can conclude that cow_n will eat grass like all other cows). As such, whether mental disorders are natural kinds matters. If mental disorders are natural kinds then there will be laws, explanations, and sound inductive inferences in psychiatry – in short psychiatry will be a genuine science. If on the other hand mental disorders are not natural kinds, whether psychiatry is a science must be questioned.

Throughout this discussion it should be borne in mind that the question of whether types of mental disease are natural kinds is completely distinct from the question of whether the super-category of mental disease forms a natural kind. The category "weed" is not a natural kind, but types of plant that are commonly considered weeds are natural kinds. I will argue that the situation with diseases is similar. "Disease" is not a natural kind – because whether a condition is a disease depends on whether it is bad thing. Nevertheless I will argue that many of the conditions that are generally considered diseases – tuberculosis, Huntington's chorea, and so on - are natural kinds.

The remainder of this chapter falls into three sections. The first assesses accounts of natural kinds. Various different accounts of natural kinds have been proposed, and whether mental disorders are natural kinds will depend on the account of natural kinds adopted. As such, before I can argue that some types of mental disorder are natural kinds, it is necessary to establish which account of natural kinds is correct. Once a satisfactory account of natural kinds has been outlined, in the second section I go on to refute arguments that purport to prove that types of mental disorder cannot be natural kinds. The arguments that I attack are philosophically more technical than the other material dealt with in this book. To readers who find this section difficult, I can only say that all three of the arguments I refute need refuting. All have been influential and have convinced some people that mental disorders cannot be natural kinds. As such, when asking whether mental disorders are natural kinds, there is nothing for it but to tackle these arguments head on. Finally, with an account of

[2] Classic formulations of the links between kinds, laws, explanations, and inductive inferences can be found in Quine 1969 and Nagel 1979, pp.30-1, footnote 2. For more recent discussion of these links see Bird 1998. It should be noted, however, that there is also a distinct Aristotelian tradition that thinks of talk about natural kinds as being important chiefly for debates about identity and change. See, for example, Lowe 2002.

natural kinds in place, and arguments that mental disorders are not natural kinds refuted, in the last section of the chapter I argue that some types of mental disorder actually are natural kinds.

1. ACCOUNTS OF NATURAL KINDS

Whether mental diseases can be considered natural kinds may well depend on the account of natural kinds adopted. As such, before going further it is necessary to ask which account of natural kinds is most plausible. Traditional accounts of natural kinds centre around ideas of "essences" or "essential properties".[3] Popular candidates for such essential properties are the atomic numbers of chemical elements and some kind of genetic property in the case of biological species. The essentialist claims that all members of a natural kind possess the same essential property. Thus, all samples of gold have an atomic number of 79, and all water is H_2O. The essential property fulfils two roles. First, possession of the essential property determines membership of the kind - to belong to the natural kind "gold" it is necessary and sufficient to have an atomic number of 79. Second, the essential property largely determines the other properties possessed by members of the kind - it is a lawful consequence of having an atomic number of 79 that a piece of material will be metal, will conduct electricity, will be solid at 20°C, will be largely inert, and so on.

Importantly for the essentialist, natural kinds provide a basis for our inductive inferences. So long as the background conditions are kept constant, this sample of gold will melt at the same temperature as other pieces of gold. This is because all members of a kind possess the same essential property, and the essential property lawfully determines the behaviour of the entity. It is because all members of a natural kind behave similarly that natural kinds are of interest to science. Measuring the melting point of *this* sample of gold is worthwhile because it provides one with information about all pieces of gold. Similarly, when a biologist dissects an organism they learn about the physiology of all organisms of that kind, not just about the individual.

In recent years traditional essentialist accounts of natural kinds have come in for fierce criticism. A major difficulty is that for biological species, which are traditionally considered amongst the best examples of natural kinds, no plausible candidates for the essences can be found. Several different criteria may be employed by biologists seeking to delineate species: morphological features, evolutionary lineage, the criteria of reproductive isolation, or genetic features. On examination none of these appear suitable candidates for being the essential properties of biological species.

In practice, most organisms are sorted into kinds on the basis of their morphological characteristics. Species X is known to have such and such wing markings, species Y has tail feathers shaped just so, and so the species to which individual organisms belong can be identified. Could morphological features, such

[3] For a recent defence of traditional essentialism see Wilkerson 1995. Other prominent essentialists include Kripke 1980 and Putnam 1970.

as possessing particular tail markings, serve as the essential properties of biological species? Saul Kripke presents the classic reasons for thinking not in *Naming and Necessity*.[4] For the sake of argument, suppose that tigers essentially are four-legged, large animals with yellow and black stripes. Then, argues Kripke, tigers would necessarily possess these features. However, in actual fact there are three-legged tigers, and albino tigers, and so we can see that our initial supposition was wrong and that tigers do not necessarily have to look like tigers. Nor is it sufficient for something to be a tiger that it looks like a tiger. Suppose, argues Kripke, that we discovered that some of the creatures we'd considered to be tigers were actually reptiles. Admittedly, this would be a highly surprising discovery, but then people were surprised to find out that whales are not fish. In this situation, Kripke says, we'd be forced to conclude that we'd made a mistake and that these reptilian "tigers" weren't tigers at all, rather they belonged to some other species, Fools' Tiger say. As looking like a tiger is neither necessary nor sufficient for being a tiger, morphological features cannot be the essential properties of biological species.

Nor can the essential property of a species be its evolutionary lineage. John Dupré shows this in his paper "Natural Kinds and Biological Taxa". Relationships of ancestry cannot be the essential properties of species because "Any sorting procedure that is based on ancestry presupposes that at some time in the past the ancestral organisms could have been subjected to some kind of sorting".[5] The point is that in order to make sense of claims such as "Cats are the offspring of cats, while dogs are the offspring of dogs" one must have some way of distinguishing the ancestor cats from the ancestor dogs. Relations of ancestry are only of any use once the parent organisms have been sorted into kinds. As such, sorting on the basis of ancestry must always be a secondary, parasitic method of sorting. When we are seeking the essential properties of species it is thus more appropriate to look to the basis of the primary method of sorting, whatever it might be, rather than to relations of ancestry.

For well-known reasons the criterion that members of a species be able to successfully interbreed will not do either. Some members of any species will be infertile and so unable to successfully breed with any other organisms. As such the criterion that members of a species be able to interbreed is too strong. It is possible to weaken the criterion so that infertile organisms can be accommodated. A revised criterion might require only that reproductive links exist between all members of the species. Infertile organisms satisfy this condition, as some member of the species must have given birth to them. Now, however, the revised criterion is too weak. There may well exist hybrid organisms that have been produced by matings between members of different species, and these individuals will now also count as members of the species. No criterion concerning interbreeding can be formulated that will both count infertile organisms as members of the species and also exclude hybrid organisms. Moreover, criteria concerning patterns of breeding are of no use when considering species that reproduce asexually.

[4] Kripke 1980 pp.119-121
[5] Dupré 1981 p.88

Given the current state of biology, genetic properties are left as the best candidates for being the essential properties of biological species. These properties seem appropriately theoretically important. However, as Dupré points out,[6] there are reasons for thinking that often there will be no one genetic property or set of properties shared by all members of a species. Most importantly, evolutionary theory suggests that it will be beneficial for there to be variation in the genes possessed by members of a species as this will facilitate quick adaptation if the environment changes. The existence of genetic diseases gives another reason for thinking that the genetic properties of members of a species will vary.

As no plausible candidates for the essential properties of biological species can be found, the claim that biological species have essential properties is thrown into doubt. Dupré concludes that as biological species are paradigmatic examples of natural kinds, and yet it is plausible that the members of a species need share no essential property, essentialist accounts of natural kinds must be rejected. [7] Instead he advances a view that he terms "promiscuous realism". Dupré has outlined promiscuous realism in a number of different works. The versions differ slightly, and it seems that over the years Dupré has become less promiscuous. Here I shall first outline Dupré's original account, as presented in his 1981 paper and 1993 book. Then I shall show why this original account is unacceptable, and go on to consider Dupré's more recent, revised account.

In his original account, Dupré asks us to imagine the individual entities of some domain (he considers biological organisms but his ideas can be generalised) mapped out on a multidimensional quality space. He claims that in such a map we would find numerous clusters corresponding to groups of similar entities. In many cases the clusters will not be discrete, but will be messy and hard to make out. Some clusters will correspond to traditional natural kinds – plausibly dogs will all cluster together, for example. At different levels of resolution different clusters will be discerned – as well as a cluster that corresponds to dogs, there will be finer clusters corresponding to dog-breeds, and, at a finer level still, to particular strains of pedigree dogs. Different clusters can also be generated by restricting our attention to particular dimensions of the map. If we restrict our attention to the dimensions that code for nutritional value, for example, we will find a cluster of things that are poisonous to humans. The task of the taxonomist is to pick out areas of relatively high density in the quality space. Dupré's account is realist because the clusters in the quality space reflect the real structure of nature. His account is promiscuous because many different classification systems can be extracted from the pattern of clusters in the space and none is privileged over the others.

Promiscuous realism salvages the idea that the divisions between classes of entities exist in an external world and can be discovered. Crucially, however, the distinction between natural kinds and groups of accidentally similar entities is lost. On Dupré's account any class of similar entities counts as a natural kind. But this is problematic. Consider the tins of tomato soup in Mr Smith's shop. These tins are all about three months out of date, all slightly dented, and all priced at 59p. The tins of

[6] Dupré 1981 pp.84-85
[7] Dupré 1981, 1993

tomato soup are similar to each other and so will form a cluster in the quality space. However, this cluster arises accidentally. There are no laws linking the fact that Mr Smith forgot to check the date labels, that the cleaner is heavy handed and knocked the tins over, and that the Saturday boy priced them all at 59p. In contrast the properties possessed by members of a natural kind, rabbits for example, tend to be lawfully linked. The properties of having long ears, being born blind, and being susceptible to myxomatosis are all found together for a reason.

The distinction between natural kinds and classes of entities that just so happen to be similar to each other must be maintained. It is only because the properties of members of a natural kind are lawfully linked that we can make inductive inferences about the members of natural kinds. It is because the properties of rabbits are lawfully connected that we can infer that anything that looks like a rabbit will also possess other rabbity features. Similarly, only natural kinds can support counterfactuals and function in explanations. As Dupré's clusters of similar entities merely exist as brute facts and need not be supported by natural laws, his account leaves our inductive and explanatory practices in limbo. For this reason, I suggest that Dupré's original account is an unsatisfactory alternative to traditional natural kind accounts and should be rejected.

Dupré thinks my tins of tomato soup example is unfair to him. He says that it is also part of his account that natural kinds should serve some kind of investigative or explanatory goal.[8] I don't think this comes over in the early versions of his account. In his 1981 paper Dupré says, "there are many sameness relations that serve to distinguish classes of organisms in ways that are relevant to various concerns… promiscuity derives from the fact that none of these relations is privileged". As an illustration he cites the texture of frogs' legs as being a quality that might interest gourmets.[9] In his 1993 book Dupré still holds that kinds might be distinguished on the basis of properties including those that are "economically useful or strikingly noticeable… [or]… of interest for further theoretical reasons",[10] which again seems to me too liberal. Kinds distinguished on the basis of properties that are merely economically useful or strikingly noticeable are little better than accidental kinds and as such will also often be incapable of doing the work required of natural kinds. Contrast a kind picked out on the basis of properties that are of economic significance, "clothes once worn by Princess Diana", with a kind that I would consider natural, "made of 100% cotton". Knowing that a piece of clothing was once worn by Diana tells one little about it – apart from giving an indication that it's likely to be worth money. In contrast, knowing that something is made of 100% cotton provides a host of information – it tells one what kinds of chemical and physical properties the cloth will have.

This being said, in some of his later work, Dupré is less promiscuous and suggests that kinds must provide a basis for scientific theorising.[11] So, we should turn to consider the question of what would happen if, to the account described here,

[8] Dupré personal correspondence 1999
[9] Dupré 1981 pp82-83
[10] Dupré 1993 p.113
[11] Dupré 2002

Dupré adds a condition stating that genuine natural kinds can be distinguished from accidental kinds on the basis that they serve an investigative or explanatory function. Then, I suggest, his account would become something like the account that I think is the correct account of natural kinds, to which we can now turn. The account of natural kinds that I shall outline and endorse isn't particularly novel – it really is just Dupré's account tweaked. My aim here isn't to provide an account that is excitingly new, but merely to provide an account that is right.

1.1 The Right Account

I suggest that the right account of natural kinds claims that members of a natural kind possess similar important properties. These important properties are important because they determine many of the other properties possessed by members of the kind. For this reason I will call them "determining properties". The determining properties of members of a natural kind must be similar, but not necessarily identical; thus this is not an essentialist account. The determining properties lawfully determine many other properties of the members of the natural kind. Of course, many of these laws will be ceteris paribus laws, that is they apply all other things being equal, and so background conditions will also be important. Still, as the determining properties of members of a natural kind are similar, so long as environmental factors are kept constant, members of a natural kind end up being similar in many respects.

To present the account somewhat differently, we can imagine all the entities in some domain plotted on a Dupré-style multidimensional map in which Dupré's quality dimensions are replaced by determining-property dimensions. For different domains different numbers of determining-property dimensions will be required. For chemical isotopes, for example, the necessary dimensions would plausibly be atomic number and neutron number. For biological species it is plausible that dimensions corresponding to various genetic properties would be required (or even more probably dimensions corresponding to particular genetic properties plus whatever environmental factors are important in determining how they are expressed). Members of a natural kind will cluster together in such a space.

I begun by thinking that the determining properties would always be microstructural, "underlying" properties, such as having an atomic number of 79. I am grateful to Dupré for pointing out to me that this might be needlessly restrictive. It is possible that in some cases determining properties might shape members of a kind "from above" rather than "from below". For example, prey animals might all be significantly similar because they have all evolved under similar pressures. In response to being hunted, they may all have evolved to become timid, have large litters, and blend in to their environment. In such a case they would have been shaped by a determining property that acts "from above". In his 1995 paper, "A different kind of natural kind", Crawford Elder develops in much greater detail the idea that the "members of some natural kinds may reliably present the same distinctive packet of observable properties, not because of anything that is

distinctive about their insides, but because of the moulding of a common environment".

I claim that members of a natural kind all possess similar determining properties. Obviously similarity is a matter of degree. In saying that members of a natural kind all possess similar determining properties I accept that this implies that there will be borderline natural kinds whose members possess determining properties that are quite, but not very, similar. Plausible examples are higher level biological kinds, such as "plant" or "vertebrate". Whether or not a property is important enough to count as a determining property may also be indeterminate, and so there will also be borderline natural kinds whose members are similar with regard to some moderately important property. The kind "red things" might be an example. Red things have something in common, namely being red, and there are some laws that concern red things, for example "In standard conditions red things appear red to normal observers", and maybe "Red berries are normally poisonous". However, being red isn't a property that is lawfully linked to much else. It doesn't seem important enough to count "red thing" as a proper natural kind. Members of borderline natural kinds will be similar to each other in fewer respects than members of more typical natural kinds.

At this point my account may be clarified by briefly discussing those classes of entities that will not qualify as natural kinds. Classes of entities completely fail to be natural kinds when they are not actually similar to each other in any way. Thus the class of things on my desk almost certainly fails to be a natural kind because "being on my desk" is highly unlikely to be a genuine property. Genuine properties, such as possessing negative charge, endow entities with particular causal powers, and ground objective similarities. The pseudo-property of "being on my desk" does neither of these. It is worth noting that when entities are artificially produced this does not necessarily preclude them forming a natural kind. The "natural" in "natural kind" should be read as in "natural law" rather than as in "present in the Garden of Eden". Plausible candidates for artificially produced natural kinds include plutonium and nylon.

So far I have only considered cases in which the determining properties of some class of entities cluster in all dimensions of the determining-property space. However, there may also be more complicated shapes produced. Some classes of entities might possess sets of determining properties only a sub-set of which cluster. Consider, for example, a set of entities that are similar with respect to all but one determining property. In such a case a fuzzy line lying in the direction of the dimension representing the varying property would be found. If a set of entities are similar with respect to all but two determining properties then a fuzzy plane would be produced. Such sets of entities are like natural kinds up to a point; the entities will all be similar with regard to whatever properties are determined by their similar determining properties, and the kinds will support inductive inferences concerning these properties. As these kinds can do part of the work of natural kinds I propose to call them "partial kinds". I suggest that partial kinds are very common. Those chemical elements that form isotopes will all be partial kinds, as samples of the elements possess the same atomic number but different neutron numbers.

Sometimes it is assumed that natural kinds must be discrete, that is that any two kinds have natural boundaries between them and that intermediate forms do not occur.[12] I do not make this assumption. On my account, the determining properties do all the work when it comes to making inductive inferences and grounding explanations. It is because members of a natural kind all have similar determining properties, and the determining properties determine the other properties of the entities, that we can predict that all members of a natural kind will behave similarly. The "gaps" between natural kinds, where there are any, do no work. Thus there seems to be no reason to claim that natural kinds must be discrete, and abandoning this claim has the advantage that kinds that are not discrete, such as alloys, can be accommodated within a natural kind account. Alloys can plausibly be considered natural kinds, I suggest, because knowing that that a sample is a particular alloy is as useful, and useful in the same kinds of ways, as knowing that it is a 100% pure metal. Once one knows that a sample is a 40% zinc and 60% copper alloy one can predict how the sample will behave just as well as if one knew it to be pure copper.

I claim that natural kinds do not need to be discrete. In order to count as co-members of a natural kind, entities just need similar determining properties, and whether or not kinds are discrete makes no difference to their ability to fulfil this criterion. As a consequence, I think that debates as to whether mental disorders are separated by "zones of rarity" (that is, whether there are gaps between them in quality space) are not relevant to the question of whether they are natural kinds.[13] It might well be the case that types of depression and of anxiety disorder merge into each other, for example. Conceivably this might occur because the genetic bases of both depression and anxiety disorders are similar, if not identical, and similar environmental stressors are risk factors in both cases.[14] In such a situation, when plotted in a multi-dimensional determining-property space, cases of the disorders would not form distinct clusters. Still, cases of depression could form a natural kind in my sense – all cases might be fundamentally similar, and this might also be the case for anxiety disorders.

Robert Kendell and Assen Jablensky (2003) have recently claimed that psychiatric diagnoses can only be considered valid if a zone of rarity separates each disorder from others. They take "validity" to mean "well-founded…sound…against which no objection can fairly be brought",[15] which clearly implies that it would be a bad thing if psychiatric diagnoses fail to be valid. In contrast, I hold that the absence of zones of rarity would not be that important. Indeed, Kendell and Jablensky seem to admit as much when they say that even if diagnoses fail to be valid they may yet have "utility" and, furthermore,

> …provide invaluable information about the likelihood of future recovery, relapse, deterioration, and social handicap … guide decisions about treatment; and …provide a wealth of information about similar patients encountered in clinical populations or community surveys throughout the world – their frequency and demographic

[12] For example, Mill 1973 p.123, De Sousa 1984 p.565, Haslam 2002

[13] Haslam 2002 provides an example of someone who runs these two questions together.

[14] This example is taken from Kendell and Jablensky, 2003, pp.9-10.

[15] Kendell and Jablensky 2003 p.8

> characteristics, their family backgrounds and premorbid personalities, their symptom profiles and their evolution over time; the results of clinical trials of several alternative therapies; and research on the etiology of the symptom.[16]

Faced with this list, I suggest that an "invalid" diagnosis that has "utility" can give us all we might want. Disorders that are "invalid" in Kendell and Jablensky's sense, can still count as natural kinds on my account.

Where there are gaps between natural kinds, I claim that these gaps are of no metaphysical significance. However, I accept that when a domain is "gappy" this can be epistemically advantageous. In order to predict how an entity will behave we need to know at least roughly what sort of an entity it is. In gappy domains, the problem of identification becomes a multiple-choice problem (is this an X, or a Y, or a Z) rather than one with a potentially limitless number of possible answers. This can make correct identification easier.

Whether a domain is continuous may also make a difference to whether a dimensional or categorical classification system should be preferred. *If* maximal information retrieval is the sole aim of a classification system, then a dimensional classification system is best for a continuous domain, and a categorical classification system is best for a discontinuous domain. Following this reasoning, Kendell and Jablensky assume that if there fail to be zones of rarity between disorders then psychiatric diagnosis should employ a dimensional system.[17] This conclusion need not follow, however. In practice, classification systems do not only need to supply information about the classified entities, they must also possess other virtues, such as being easy to use. For this reason a categorical classification system may be used when classifying a continuous domain – as is the case with British degrees that are classified as being third, second, or first-class. Thus, even if there are no zones of rarity between mental disorders, a categorical classification system, such as the current D.S.M., may still be the best option. Whether a classification should be categorical or dimensional, and whether a domain is discrete or continuous, are distinct questions, and I claim that both issues are separate from the question of whether a domain consists of natural kinds.

The outlines of my account of natural kinds are now in place. The central claim of my account is that members of a natural kind all possess similar determining properties, where the determining properties of an entity are those properties that determine its other properties. On this account the links between natural kinds and natural laws can be easily understood. A determining property is a property that determines many other properties. As such, determining properties appear in many natural laws and so are the kinds of properties in which scientists are likely to be interested.

Although determining properties are like essential properties in some respects they are also importantly different. Essential properties and determining properties are similar in that both are said to determine many of the other properties possessed by an entity. Essential properties and determining properties differ in that the essentialist claims that all members of a natural kind must share some identical

[16] Kendell and Jablensky 2003, pp.9-10.
[17] Kendell and Jablensky 2003 p.8.

essential property, while I only claim that members of natural kind must possess *similar* determining properties.

It is worth noting that the account proposed here has some similarities to that proposed by Richard Boyd. Boyd argues that members of a kind possess a cluster of regularly co-occurring properties and that this is for a reason, there is some "homeostatic" mechanism that makes it the case that these properties re-occur. [18] Boyd's account works well for biological species. Members of a species possess clusters of co-occurring properties and this is as a result of homeostatic mechanisms, such as gene flow between the organisms and pressures that arise from the fact that all members of the species must survive in similar environments.

I think that Boyd's account is adequate for some natural kinds, and indeed his account can be accommodated by my own. I can say that it is because homeostatic processes act on them that members of a species possess similar determining properties. Still, I do not regard Boyd's account as suitable as a complete account of natural kinds. His account may work well enough for biological species, but it does not lend itself to dealing with other types of natural kind such as types of fundamental particle. There are no homeostatic mechanisms that make it the case that the properties of fundamental particles hang together. A particle with the mass of an electron might be negatively charged (and so be an electron), or it might be positively charged (and so be a positron). Fundamental particles are paradigmatic examples of natural kinds and any satisfactory account of natural kinds should accommodate them. For this reason I reject Boyd's suggestion that there must be a homeostatic mechanism that makes it the case that members of a kind possess a particular cluster of properties. On my account all that is needed is that members of a kind share a cluster of determining properties – whether there is a homeostatic mechanism behind the co-occurrence of the properties is unimportant.

To summarise this section: I have proposed an account of natural kinds according to which members of a natural kind possess similar determining properties. The determining properties of a member of a kind determine many of its other properties. As a result, members of a kind can be expected to behave similarly in similar circumstances.

2.ARGUMENTS AGAINST MENTAL DISORDERS BEING NATURAL KINDS

Various arguments have been put forward that purport to show that mental disorders cannot be natural kinds. These arguments all aim at showing that mental diseases cannot be natural kinds on the traditional, essentialist understanding of natural kind. Still the arguments, if sound, would show that mental diseases could not be natural kinds on my account either. When describing the arguments I have altered the terminology to make it consistent with that used in formulating my account (i.e. changed "essential property" to "determining property" where arguments against the existence of one will also be arguments against the existence of the other). This

[18] Boyd 1988, 1991

section is devoted to refuting these arguments that mental diseases cannot be natural kinds.

2.1 The Historical Argument

Some recent work in the history of medicine has aimed at showing how disease entities have been constructed via the interaction of various technologies, institutions, and social interests. To take an example, in *The Harmony of Illusions* Allan Young claims to show how Post-Traumatic Stress Disorder arose out of the interaction of lobbying by Vietnam veterans and the various tests and treatment programmes which arose for diagnosing and treating the disorder. Young and the authors of other such case studies argue that as a disease entity has been artificially manufactured it cannot be a natural kind.[19]

There are two possible ways of replying to such arguments and the appropriate response varies from case to case. In some cases, one can agree that the disease has been created but argue that natural kinds can be artificial in this sense. The key to seeing that natural kinds can be artificially created is to remember that the "natural" in "natural kind" should be read as in "natural law" rather than as in "present in the Garden of Eden". Plutonium is an example of a manufactured natural kind, and doubtless some highly "social" story could be told concerning its creation. Diseases that are artificially produced – in the sense of being produced by modern ways of living – such as types of drug addiction and, arguably, Post-Traumatic Stress Disorder could similarly be both artificially manufactured and natural kinds.

In other cases, one can argue that the disorder itself has not been artificially produced. Rather, the disease has always existed and it is only the means of *recognising* it that is recent. This is a plausible response when retrospectively the disease can be seen to have afflicted people throughout history. Some authors consider this to be the case with Post Traumatic Stress Disorder. For example, R. Daly has claimed that the mental symptoms recorded in Samuel Pepys' diary indicate that he suffered from Post Traumatic Stress Disorder after having witnessed the Great Fire of London.[20] Turning to more recent times, a 1956 textbook describes a condition highly reminiscent of Post Traumatic Stress Disorder called "traumatic psychoneurosis".[21] Given such evidence it can plausibly be claimed that people have

[19] Young 1995 p.5 "The disorder is not timeless, nor does it possess an intrinsic unity. Rather, it is glued together by the practices, technologies, and narratives with which it is diagnosed, studied, treated, and represented and by the various interests, institutions, and moral arguments that mobilised these effects and resources." As another example, Aronowitz 1998 ch.3 argues that Lyme disease has been socially constructed and is thus not a natural kind.

[20] Daly 1983

[21] Henderson and Gillespie 1956 p.207 The disorder is described as one in which "The [triggering] experience is nearly [always], if not always overwhelming... The symptoms that follow the fright are usually insomnia with terrifying dreams in which the patient wakes again and again; these dreams representing the accident in more or less distorted form. Anxiety symptoms occur during the day, especially lack of concentration or uneasiness of mind, and the bodily discomforts associated with anxiety, such as tremor or palpitation. Such symptoms may appear even in the most stable individual if the experience is severe enough..."

suffered from Post Traumatic Stress Disorder throughout history. On such a view, rather than Young having documented the social factors that led to the construction of Post Traumatic Stress Disorder he has just documented the social factors that led to the condition being diagnosed.[22]

Amongst some historians of medicine, however, it has become fashionable to argue that studies that seek to show that people in times past suffered from the same disorders that afflict people today are anachronistic. Both historians of medicine and transcultural psychiatrists often claim that people from other cultures or times cannot be said to suffer from the same diseases as twentieth century Westerners.[23] Andrew Cunningham presents the clearest argument for this position in his paper "Transforming plague: the laboratory and the identity of infectious diseases". Cunningham claims that the meaning of disease terms, such as "plague", is partly fixed by the ways in which the disease is identified. On such an account, as we identify plague with laboratory tests while people in earlier times identified plague by its symptoms, "plague" as it was used by historical figures and "plague" as used by us have quite different meanings. To avoid confusion we could write "plague_m" or "plague_h" to indicate whether the modern or historical meaning is being signified. If we accept Cunningham's claims, then as "plague_m" and "plague_h" have quite different meanings we should not talk of historical figures as having suffered from the same disease that afflicts people today; they suffered from plague_h while we can only suffer from plague_m. If Cunningham is right, people from other cultures and times cannot have suffered from the same diseases that afflict people currently.

I am happy to accept Cunningham's claim that we should be sensitive to the meanings of disease terms as they were used by historical figures, but I will argue that he goes too far in claiming that we should not talk of historical figures as having suffered from contemporary diseases. My argument depends heavily on the distinction between terms appearing in "referentially opaque" and in "referentially transparent" contexts. In statements such as "Mary believes that x", "Mary hopes that x", and "Mary is afraid that x", the phrase x appears in a referentially opaque context. Characteristically, in such cases the truth value of a statement may alter when a term is swapped for another that refers to the same entity. For example, suppose Fred Bloggs is the masked man, but Mary does not know this. In such a case, "Mary is afraid of the masked man" may be true, while "Mary is afraid of Fred Bloggs" is false. In contrast, when phrases appear in referentially transparent contexts, they may be swapped for other terms with the same reference, and the truth value of the statement is always preserved. For example, if it is true that Bloggs has a broken arm and acne it will also be true that the masked man has a broken arm and acne.

[22] Young is, of course, aware of cases such as that described in the 1956 textbook, but he claims that they are not cases of Post Traumatic Stress Disorder. On this, however, reasonable people can disagree.

[23] An example of a transcultural psychiatrist holding these views can be found in Fernando 1991 pp. 131-132. Fernando complains that "The basic assumption underlying the IPSS [International Pilot Study of Schizophrenia] is concerned with the meaning of "schizophrenia": it is assumed to be an objective entity...and, moreover, an entity that is "present" in objective form all over the world with a universally similar, if not identical, meaning irrespective of culture."

Returning to Cunningham's argument, I agree with Cunningham that a consideration of actors' beliefs is important when dealing with terms as they occur in opaque contexts. Thus if we are trying to decide whether the statement "I believe I've got the Black Death" can be adequately translated by the statement "I believe I've got plague" the beliefs of the actor whose mental states we are trying to describe must be considered. Given Cunningham's evidence that "Black Death" was a disease identified by its symptoms while contemporary "plague" is a disease identified by its cause, I accept that the suggested description of the patient's belief should be rejected. Even if the historical patient uses the word "plague" rather than "Black Death", it would be misleading to describe his belief as being that he had plague. "Plague" as used by the patient is "plague_h" which, Cunningham has shown, has radically different implications from "plague" as we use the word.

I break company with Cunningham when it comes to talking of plague in contexts that are referentially transparent. I claim that statements such as "There was a lot of plague around in the 18th century", "Maybe the plague bacillus was more virulent in the past", and "Mr Smith died of plague in 1756" would, if true, be perfectly respectable statements. This is because the "plague" in such statements has our modern meaning, and as such refers to all cases of disease that are caused by the plague bacillus. The beliefs of the various sufferers are neither here nor there; all that matters is whether or not the bacillus was present in their bodies. Similarly, while a statement like "Priestley believed he breathed oxygen" is anachronistic, a statement like "Priestley breathed oxygen" is admissible. The difference is that in the first case we are trying to describe Priestley's mental states and so must use concepts with which he would have been familiar; whereas in the second case we are just making statements about the actual gas that he breathed. Everyone breathes or breathed oxygen regardless of whether they have or had any beliefs about it.

As such, Cunningham's claim that historical figures cannot be said to suffer from contemporary diseases should be rejected. Thus there is no reason why historical and transcultural studies that make this assumption should be considered illegitimate. In some cases such studies can give us reason to believe that a particular disease has occurred throughout history and that only the means for recognising it are of recent origin. For this reason, and because natural kinds can in any case be artificially created, I conclude that historical studies, such as Young's, do not show that a disease is not a natural kind.

2.2 Hacking's Argument

In a series of papers written between 1986 and 1995 Ian Hacking developed an argument that purports to show that types of mental disorder cannot be natural kinds.[24] Since then, Hacking seems to have changed tack and, although he does not give reasons for rejecting his earlier work, Chapter Four of his *The Social Construction of What?* (1999) raises the possibility that at least some types of mental disorder are natural kinds. Hacking discusses autism as a possible example. Here I

[24] Hacking 1986, 1988, 1992, 1995a., 1995b.

will be concerned mainly with refuting Hacking's earlier argument that mental disorders cannot be natural kinds. Although Hacking may have moved on, his argument has been influential and so is still worth considering.

In his earlier work Hacking introduces the term "human kind" to refer to the kinds of people - child abusers, pregnant teenagers, the unemployed - dealt with by the human sciences. "Human kind" is a term chosen by Hacking to contrast with "natural kind". Hacking argues that classifying and describing human kinds results in feedback that alters the very kinds under study. The resulting feedback means that human kinds have histories totally unlike the histories of natural kinds, leading Hacking to conclude that human kinds are not natural kinds. Hacking's case studies include Multiple Personality Disorder and autism, thus it is clear that he considers kinds of psychiatric patient to be human kinds.

Here I argue that Hacking's argument is flawed, and that he has failed to show that types of mental disorder cannot be natural kinds. The feedback that Hacking claims makes human kinds so very different from natural kinds is supposed to operate at two levels, a cultural level and a conceptual level. I will examine each type of feedback in turn, and show that in so far as feedback occurs it is compatible with human kinds being natural kinds.

2.2.1 Cultural Feedback

Feedback at the cultural level is dependent on the description of a kind of person entering popular culture. Often human kind terms carry heavy moral overtones; consider for example, "sexual pervert", "alcoholic", and "normal". Being classified in a certain way may also carry institutionalised benefits or costs. For example, students classified as being dyslexic may receive extra time in exams, and one may have to be certified "psychologically fit" before being employed in certain roles. As a consequence, people are motivated to attempt to alter the ways in which they are classified and, as their behaviour changes, so do the kinds under study. Consider, for example, the kind "obese person": The characteristics of both obese and non-obese people are affected by attitudes towards obesity. When obesity becomes stigmatised obese people will tend to become socially isolated and unhappy, and go on diets, while non-obese people will start making jokes about obesity and worry about becoming obese themselves. Attitudes towards obesity also result in new human kinds, such as "people with stapled stomachs", coming into existence. Hacking claims that the existence of *such* feedback shows that human kinds cannot be natural kinds.

J. Bogen has interpreted Hacking as claiming that human kinds are not natural kinds because the classification of human kinds results in feedback.[25] Hacking rejects such an interpretation. In any case, as Bogen points out, such an argument would fail because our classificatory practices also result in feedback that alters some natural kinds. For example, because marijuana is classified as illegal the plants are grown in attics and wardrobes altering their physical appearance. As another example, the characteristics of domestic livestock change over time because

[25] Bogen 1988

particular animals are classified as being the "Best in Show" and are used in selective breeding – sheep and pigs would now look very different if it weren't for our classificatory practices.

Hacking's argument that human kinds are not natural kinds must rest not merely on the fact that feedback occurs, but rather on the fact that it occurs in a particular way. The difference, Hacking claims, is that feedback in human kinds occurs because subjects become aware of the ways in which they are being described and judged.[26]

This idea needs working on before it can become an argument that human kinds cannot be natural kinds. As it stands Hacking has merely claimed that human kinds can be affected by a mechanism to which other kinds of entity are immune. Although this shows that there is some difference between human kinds and other kinds, it is not sufficient to show that this difference is of any fundamental significance. After all many other types of entity can be affected by mechanisms to which only entities of that type are vulnerable. While it is true that only human kinds are affected by the subjects' ideas, it is also true that only bacteria are affected by antibiotics, and that only domestic animals can be selectively bred. But no one would cite this as evidence that "bacterial kinds" or "domestic animal kinds" are not natural kinds.

The fact that only human kinds are affected by the subjects' ideas will only be a reason for thinking that human kinds are distinct from natural kinds if an extra premise is added to the effect that being affected by ideas is of greater metaphysical significance than being affected by, say, antibiotics. In places Hacking suggests that feedback caused by the subject's awareness of being classified is important because it results in feedback occurring at a faster rate than that which affects natural kinds.[27] The thought seems to be that the speed with which change occurs confounds our attempts to use human kinds in inductive inferences. Such a claim is questionable. Do human kinds really change more quickly than bacteria and viruses mutate? In any case, a difference in the rate of feedback is inadequate to mark a fundamental metaphysical distinction between human kinds and natural kinds. If it were true that the characteristics of human kinds shifted more rapidly this would imply that human kinds are not particularly useful natural kinds, not that human kinds cannot be natural kinds at all.

Alternatively, idea-dependence might be thought to matter because it betrays the subjective nature of a kind. The argument then would be that while natural kinds are objective, human kinds are affected by ideas and so subjective, and that thus human kinds cannot be natural kinds. Hacking gives no indication that this is a route he would wish to go down; however it is the most obvious option for someone who wishes to claim that idea-dependence is metaphysically significant and so worth pursuing here.

However, entities can be idea-dependent in two fundamentally different senses. And, as I will argue, idea-dependence in only one of these senses is indicative of

[26] Hacking 1997 p.15

[27] Hacking 1992 p.190 suggests feedback in natural kinds is different because "it works not at the level of individuals but through a great many generations, be it for microbes or mammals."

subjectivity. Compare two senses in which ideas of female beauty "affect" entities: In one case a woman, influenced by images of the "ideal female form", decides she is too fat and so slims. Her altered shape is idea-dependent in the sense that her ideas concerning her weight caused her to slim. The development of Concorde was dependent on the ideas of its developers in much the same kind of way; the developers had ideas about aeroplane designs, and these ideas feature in the causal history that culminated in the building of Concorde. Nevertheless, despite being in a sense "idea-dependent" the reduction in the woman's weight, and the building of Concorde, are both perfectly objective. Idea dependence of this type results in objective changes in entities and is compatible with a kind being objective.

On the other hand consider the case where we look at old photos of the first Miss World. Miss World looks rather plump and short by today's standards, nevertheless presumably at the time she looked fine. Miss World's looks are also idea-dependent, but this time nothing about the photo of Miss World has actually changed. Rather the ideas prevalent in popular culture have made the properties of the photo appear different solely by acting on the viewers. The change is a relational change only. Such relational changes indicate that a kind, such as "attractive women", is merely a subjective kind and so not a natural kind.

Hacking has shown that human kinds are idea-dependent. In order to show that this means that human kinds are subjective and so cannot be natural kinds, it would need to be shown that human kinds are idea-dependent in the way that produces relational as opposed to genuine changes. All Hacking's examples, however, seem to be of cases where ideas produce genuine changes in people's behaviour. Take, for example, the case of Multiple Personality Disorder.[28] When patients with personalities of the opposite sex and animal personalities started to appear on American chat shows, more and more patients started presenting with similar symptoms. The ideas prevalent in popular culture affected the symptoms typical of Multiple Personality Disorder. Still, here it seems that ideas about Multiple Personality Disorder caused a genuine change in patients' symptoms. Patients really did start barking. Such a claim need no more incriminate the kind "Multiple Personality Disorder" than the claim that changing views on animal welfare have resulted in fewer dogs having their tails docked incriminates the kind "dog". In order to show that the changes in the symptoms of Multiple Personality Disorder indicate that it is not a natural kind Hacking would need to show that barking, like beauty, is in the eye of the beholder, and he makes no suggestion that this is the case. Hacking's examples of cases in which human kinds are affected by ideas are all cases in which the ideas cause genuine changes. Such feedback is compatible with the kinds being natural kinds.

2.2.2 Conceptual Feedback

Hacking's argument for feedback at a conceptual level is dependent on Elizabeth Anscombe's account of intentional action. In her 1957 monograph, *Intention,* Anscombe considers the circumstances under which an action can be said to be

[28] Hacking 1995a.

intentional. Her solution is that an action X can be said to be intentional when the actor could respond to the question "Why are you doing X?" by giving a reason for acting. Suppose I lock my office as I leave, for example. If you ask me why I do this, I will tell you that I believe there are thieves about and I don't want my things stolen. I give a reason for my action, which thus passes The Why Test and counts as an intentional action. In other cases I will not be able to answer the "Why?" question. Maybe I am not aware that I am doing X, for example I have accidentally stood on someone's toe. Maybe I know I am X-ing but only because I have observed it, for example I am blushing. Maybe I know I am X-ing but the cause is presumed to be purely non-mental, for example I suffer from a muscle spasm. In these cases the behaviour does not pass The Why Test and is not an intentional action.

On Anscombe's account an action is only intentional under a description because occasionally when we ask an agent "Why are you X-ing?" he may fail to recognise his action under certain descriptions. For example, I am in the kitchen X-ing, where X may be either "cooking" or "getting in the way of my flat-mates". I recognise my action only under the description of "cooking", as I have not noticed that I am getting in the way. The action passes The Why Test, and thus is an intentional action, only under the description of "cooking".

Following Anscombe, Hacking uses the slogan "Intentional actions are actions under a description" in his argument that feedback occurs in human kinds:

1. Intentional actions are actions under a description.
2. Intentional actions make us the kind of person we are.

New descriptions allow new intentional actions which allow new kinds of person.

If Hacking is correct then the creation of new descriptions makes logically possible the creation of new kinds of person.[29] In creating new terminology the human sciences would make it possible for people to act in new ways. Here, however, I shall argue that Hacking's argument fails because he has misinterpreted Anscombe's phrase "under a description".

The phrase "under a description" occurs throughout Anscombe's monograph. However its use is idiosyncratic and in her 1971 paper called "Under a Description" Anscombe explains how she intended it to be understood. She writes, "under a description is 'qua'...in modern dress". Anscombe gives an example indicating that she uses "qua" in the usual manner, she writes "A may, qua B, receive such-and-such a salary and, qua C, such-and-such a salary."[30]

If Anscombe in fact meant "Intentional action is only intentional qua some aspect" why did she use the misleading phrase "under a description"? Anscombe worked in the ordinary language tradition. She sought to gain philosophical insight from considering the ways in which we use everyday language. As such, her monograph aims to give an account of what we say about commonplace actions.

[29] That Hacking's claim concerns logical possibility comes out most clearly in Hacking 1986.
[30] Anscombe 1971 p.208 in 1981 reprint

Within Anscombe's domain of the everyday the possibility of something being intended qua X where there is no description that refers to X does not have to be considered, as it is fair to assume that all commonplace intentions have already been described. Thus Anscombe can treat "under a description" and "qua" as equivalent.

In contrast, Hacking is interested precisely in the situations that Anscombe can ignore. Hacking wants to consider the new possibilities for action created by a new description; he needs to compare what was possible before the description was invented with what is possible after. In such cases the interpretation of "under a description" becomes key. Consider Ug the caveman, sitting in his cave at the dawn of time before language developed. According to Hacking, Ug cannot intentionally light a fire, go outside, or hum himself a tune - as there are no descriptions, Ug must wait for them to develop before he can intentionally do anything.

If, on the other hand, we take "under a description" to merely mean "qua", Ug is free to intentionally act in many ways. Ug can intend his banging flints qua a way to make a fire, rather than qua a way to make a noise. Although we cannot use The Why Test to find out what Ug intends to do, there are other ways in which we can decide what it is that he intends. We can consider Ug's probable motives: if it's cold he'd have reason to make a fire, if other people are banging drums he probably wants to make a noise. We can watch Ug's response when we intervene in his action - if he intends to light a fire bringing in wood will tend to make him smile, if he's starting a music session singing would probably be more welcome.

Such an approach fits in well with Anscombe's discussion of the intentional actions of non-verbal agents. In "Under a Description" she discusses a bird who lands on a twig that happens to be both covered in bird lime and near some seeds.[31] The bird, she says, lands on the twig with the intention of reaching a seed but not with the intention of landing in the bird lime. We infer the bird's intention by attributing intentions that are appropriate for the bird given its perceptual apparatus, its intelligence, and typical bird behaviour. We think that birds can identify seeds, that they get hungry, and that typically birds try to get seeds, and so we attribute the intention of getting the seed to the bird.

The problem of deciding how an action is intended arises because one bodily behaviour can help fulfil several different possible goals. Thus we cannot decide what someone intends merely by looking at their movements. Hacking presumes that the conditions under which an intentional action can be performed are identical to the conditions under which an observer can infer the actor's intentions. He sets about asking when intentional actions are possible via asking how an observer can determine what it is that an agent intends, and assumes that if one cannot tell what an agent intends then no intentional action is possible. This is only permissible if some verificationist principle is adopted. Even if such a principle is considered acceptable, however, if "under a description" is interpreted as "qua" there is no reason to think that intentional actions are logically dependent on the existence of descriptions. Asking an actor to explain his actions is one way, but not the only way, to discover what an agent intended. Using the method of asking the actor requires

[31] Bird lime was a sticky substance put on twigs. Birds that landed on it became stuck, and so could be caught.

descriptions, but as there are other means of inferring an actor's intentions which do not depend on descriptions, it cannot be concluded that descriptions are essential for intentional actions. Ug can intend to make a fire, and the bird can intend to land on the twig, without any descriptions being required. In such cases Hacking is simply wrong to claim that descriptions are required for intentional action.

Of course not all actions are so contingently linked to language. Consider the act of marrying someone, or the act of promising.[32] In order to marry a man one actually has to say, "I hereby take this man to be my lawfully wedded husband". Similarly, someone can only promise to do something if they say, "I promise to do X". Without the descriptions relating to marriage and promising, there can be no such actions. I suggest, however, that such actions form an unusual class. Such actions are peculiar in that they are defined in pseudo-legal ways, and the law, of course, unlike everyday thinking, dislikes ambiguity. It is extremely important to people that they have a way of being sure whether or not they are married, and of being sure when they have been promised something. That these actions are defined as being tied to the utterance of descriptions acts to reduce possible sources of doubt as to whether an intentional action has occurred or not. That one actually has to say particular sentences in order to get married makes it extremely unlikely that one could find oneself considered married by accident. In contrast, in everyday life we are able, and forced, to tolerate uncertainty, and accept conventions whereby we can say that this or that person intended to do X or Y even though there is a chance that we are wrong. In short while there is a class of pseudo-legal actions that are logically tied to their descriptions, such actions are only a sub-set of all actions. I can accept that the logical link between such actions and descriptions means that kinds such as "promisee" and "husband" will not form natural kinds, but still argue that no such logical link between actions and descriptions affects kinds such as "autistic person", "obese person" and "homosexual".

I accept, in addition, that there might be contingent links between descriptions and the ability to perform certain types of intentional actions. Some actions might be too complicated to perform without the aid of a description, for example, cooking certain complicated dishes might require a recipe describing what is to be done at each stage. It might also be true that actors only act in certain ways because certain descriptions exist in a culture, for example, it might be true that the existence of a tradition of limerick writing in a sense makes it possible for us to intend to write a limerick, as without the tradition no individual would ever think of doing such a strange thing. In such cases, however, our ability to perform certain intentional actions is only contingently dependent on the existence of certain descriptions. The descriptions in the culture may feature in the causal histories that culminate in our acting in certain ways, but they are not needed for it to be logically possible for us to act in these ways. Here Hacking's conceptual feedback collapses back into his cultural feedback and, as I have already argued, the existence of such feedback does not show that human kinds are not natural kinds. I conclude that Hacking's argument fails and he has not shown that human kinds are not natural kinds. Thus, despite Hacking's argument, types of mental disorder might be natural kinds.

[32] I am grateful to Martin Kusch for these examples.

As mentioned previously, Hacking himself has moved on since his work arguing that human kinds cannot be natural kinds. His earlier work still needed to be considered here, however, because once proposed arguments can take on a life of their own. While Hacking now accepts that types of mental disorder can be natural kinds, other people (convinced by Hacking's earlier argument) think they cannot. Still, now his earlier argument that human kinds cannot be natural kinds has been refuted, for completeness, I will briefly discuss Hacking's current views.

In *The Social Construction of What?* (1999) Hacking replaces talk of "human" and "natural" kinds with talk of "interactive" and "indifferent" kinds. Like the old human kinds, interactive kinds are affected by feedback that stems from the fact that the subjects classified are aware of how they are classified. In contrast, "indifferent kinds" are those kinds that are not aware of how they are classified. Hacking gives "quark" as an example, "the classification 'quark' is indifferent in the sense that calling a quark a quark makes no difference to the quark."[33]

The key difference between the claims made in Hacking's earlier work and in *The Social Construction of What?* is that Hacking now thinks that a kind can be both interactive and indifferent, or, using the old terminology, both human and natural. In Chapter Four, "Madness: Biological or constructed?", Hacking discusses childhood autism as a kind that is plausibly both interactive and indifferent. In an earlier essay, Hacking had shown that the symptoms typical of autism have plausibly shifted over time as a consequence of autistic people responding to the ways in which they are classified – autism can thus be considered an interactive kind.[34] At the same time, childhood autism is a disorder that will plausibly turn out to have some underlying biological cause. We can imagine that one day scientists will announce that they have discovered the abnormality that causes autism (whether it be genetic, neurological, or whatever), let us call it P. In such a scenario, Hacking suggests, the newspapers can fairly report, "Autism is P". P will be an indifferent kind, "the neuro-geno-biochemical state P is not aware of what we find out".[35] Thus autism might well turn out to be both an interactive kind and an indifferent kind. The challenge that Hacking sets himself is to show how this might be possible; how might a kind be both interactive and indifferent?

Hacking proposes a "semantic resolution" to his problem. This resolution makes use of the theory of meaning developed by Hilary Putnam.[36] Putnam suggests that we think of the meaning of a term as being a vector made up of syntactic markers, semantic markers, a stereotype, and the extension. The syntactic and semantic markers say what kind of a word the word is. For example, "water" is a mass noun and natural kind term. The stereotype is that which any competent speaker must know if they are to be said to understand the term. In the case of "water" one must know that it is thirst-quenching, colourless, in rivers, and so on. The extension is the class of things to which the term applies. In the case of "water" the extension is all samples of H_20.

[33] Hacking 1999 p.105
[34] Hacking 1995b.
[35] Hacking 1999 p.117
[36] Putnam 1975

Hacking suggests that we think of the meaning of "autism" as being a Putnam-style vector, but that in to this vector we put an enriched stereotype. The stereotype should include "the current idea of autism – prototypes, theories, hypotheses, therapies, attitudes, the lot".[37] The extension of "autism" will be the class of people with P. With this semantic apparatus in place, Hacking claims that we can understand how someone might write a paper titled "The Social Construction of Childhood Autism":

> The author could perfectly well maintain (a) there is probably a definite unknown neuropathology P that is the cause of prototypical and most other examples of what we now call childhood autism; (b) the idea of childhood autism is a social construct that interacts not only with therapists and psychiatrists in their treatments, but also interacts with autistic children themselves, who find the current mode of being autistic a way for themselves to be.[38]

According to Hacking, "autism" refers to P, which is an indifferent kind. At the same time the autistic children form an interactive kind – their behaviour is affected by the stereotype of autism, and, given time, the stereotype will have to be revised if it is to continue to describe them.

I agree with Hacking that a kind can be both interactive and natural. It is worth pointing out, however, that his account of how this can be the case differs from my own. Hacking thinks that the underlying pathology, P, is a natural kind that is unaffected by feedback. Whatever feedback occurs is thus limited. For Hacking, feedback can affect ideas about the kind. It can also affect the behaviour manifested by members of the kind (as the manifestation of the underlying disorder will be shaped by the social environment). But, feedback cannot affect the underlying pathology itself. Hacking thus makes room for interactive kinds to also be natural kinds via limiting the power of feedback. It is because Hacking thinks that feedback doesn't go "all the way down" that he can claim that there is an underlying, unchanging, biological, natural kind beneath the surface complexity.

I am happy to grant that some natural kinds of mental disorder may be as Hacking describes. To take an example, the content of the delusional beliefs of psychotic people is clearly socially influenced. Once people thought themselves possessed by demons, now they think they are controlled by C.I.A. agents. Still, plausibly, the same basic pathology underlies the delusions in both cases, and I am happy to say, along with Hacking, that this may form a natural kind.

Still, the difference between Hacking and myself is that I want to go further, and argue that even if the underlying pathological cause of a disorder is affected by feedback that disorder may still be a natural kind. I have claimed that natural kinds can be affected by our classificatory practices (for example, domestic animals are shaped by selective breeding). I have further argued that there is no reason to think that feedback that stems from the members of a kind being aware of how they are classified is of any greater metaphysical significance than any other kind of feedback. Thus, I think that even if people's ideas about how they are classified affect the basic pathology that underlies disorders (for example, because people start

[37] Hacking 1999 p.121
[38] Hacking 1999, p.121

using genetic engineering to alter the genetic bases of conditions) that these disorders can still be natural kinds. While I am pleased that Hacking has changed his mind and now thinks that interactive kinds can also be natural kinds in those cases where the effects of feedback are limited, I would urge him to go further and to concede that feedback that goes "all the way down" can also be compatible with a condition being a natural kind.

2.3 McGinn's Arguments

In *The Problem of Consciousness* (1991) Colin McGinn presents two arguments against mental or psychological kinds being natural kinds. McGinn is mainly interested in mental kinds of the type appealed to by folk-psychology - "beliefs that it is raining", "tingling sensations", "pains", and such like. Still, despite taking his examples from folk-psychology, McGinn takes his conclusions to rule out the possibility of there being any psychological kinds what-so-ever. As such, McGinn's arguments are relevant to the issue of whether types of mental diseases might be natural kinds. Here I shall refute each of McGinn's arguments in turn.

2.3.1 McGinn's Argument From Multiple Realisation

Many philosophers are attracted towards a functionalist account of mind.[39] Functionalists claim that mental states should be characterised in functional terms, that is solely in terms of sensory inputs, behavioural outputs, and relations with other mental states. On such a view, mental states can be multiply realised - that is the same mental state can be realised by different physical, or indeed possibly non-physical, systems. A human can have the belief that chocolate is the best flavour ice-cream, for example, and this functional state be realised by brain activity. A Martian could have the very same belief and this be instantiated by some configuration of the green gunge that fills Martian heads. In order to count as beliefs that chocolate is the best flavour ice-cream, all the states have to do is behave like beliefs that chocolate is the best flavour ice-cream (the state must prompt the believer to choose chocolate when given a choice of ice-cream flavours, and so on). Multiple realisation means that examples of the same mental state don't need to have similar physical properties – at the physical level, the Martian's and the human's beliefs have nothing in common.

This leads McGinn to doubt that psychological kinds can be natural kinds. Members of prototypical natural kinds have similar determining physical properties. All samples of gold are similar in having an atomic number of 79, all cats are genetically similar. In contrast, at the physical level, members of a psychological kind can be completely different. This means that if members of a psychological kind possess similar determining properties, these properties cannot be physical properties.

[39] McGinn himself is not a functionalist, but he presents this argument from the functionalist's point of view.

What about the possibility that members of a psychological kind instead possess determining properties that are functionally defined – that the determining property of a belief that chocolate is the best flavour ice-cream just is having a particular causal role? McGinn rules out this possibility on three grounds.[40] First, we must remember that the causal dispositions of mental states operate holistically. In other words, what a particular mental state does depends on the other mental states possessed by an agent. If you know that I am afraid of bulls, for example, you cannot simply conclude that when faced with a bull I will run. Maybe I have read that running from bulls antagonises them and that it's best to stick one's ground and look them in the eye. Because what a mental state does depends on the other mental states of the agent, McGinn suggests that mental states of a kind may fail to have a similar causal role.

Second, McGinn points out that determining properties are paradigmatically properties linked to the internal constitution of entities (for example, water is characterised as being H_2O, cats all have similar genetic properties). Functionally defined properties, on the other hand, are specifically not linked to the internal constitution of entities.

Third, McGinn claims that any specification of a mental state's causal role would be definitional and a priori. It is not an empirical discovery that those who believe chocolate is the best flavour ice-cream pick that flavour when given the option (all other things being equal), rather that they do this is true by definition. In contrast the properties that characterise natural kinds can only be specified a posteriori. Empirical work was required before we knew that all samples of gold have an atomic number of 79, or that water is H_2O, or that all cats are genetically similar.

If functionalism is correct, and many philosophers find it an attractive view, then psychological kinds cannot be characterised in terms of the physical properties they possess. However, claims McGinn, kinds characterised by functionally defined properties would differ so much from prototypical natural kinds that they would not be natural kinds at all. It follows, he thinks, that psychological kinds cannot be natural kinds.

Here I will not argue with McGinn's claim that natural kinds cannot be characterised in terms of functionally defined properties. I am not myself committed to this position, but am not sure how to argue against it. Instead, I will employ a different tactic. I will accept, for the sake of argument, that natural kinds cannot be characterised in terms of functionally defined properties, but will give two arguments that show that even if this is the case types of mental disorder could still be natural kinds.

First I will argue that McGinn fails to show that types of mental disorder cannot be natural kinds because it is consistent to be a functionalist about normal mental states while holding that types of mental disorder are not functionally defined. This is because systems that are functionally equivalent when they are working properly generally display different patterns of breakdown. Consider electrical resistors: These can be made out of many different materials, for now let us restrict our attention to carbon resistors and wire-wound resistors. When these resistors work

[40] McGinn 1991 p.134

properly they are functionally equivalent and either kind of resistor can be used in an electrical circuit. However, the different types of resistor can be broken in different ways. Carbon resistors are brittle and can be smashed. Wire-wound resistors are tougher but will eventually break at very high temperatures. Although "resistor" is functionally defined, different kinds of resistor breakdown in different ways, and so types of resistor-breakdown cannot be functionally defined.

On analogy with the resistor case, it should be clear that even if normally functioning mental states can be functionally defined, this does not imply that the same can be said for types of mental disorder. We can think of humans as suffering from particular types of mental disorder because of design weaknesses in the way that humans are made. So, to offer a couple of only moderately controversial examples, in humans a serotonin deficiency tends to produce feelings of misery, anhedonia and sleep problems, while a traumatic experience can produce flashbacks and nightmares. Now, as Martians don't have serotonin they can't suffer from a disorder that is caused by a serotonin deficiency, and as they are differently wired there is no reason to expect a traumatised Martian to suffer from nightmares. Disorders that arise from weaknesses inherent in the design of human brains will be specifically human disorders.

This being said, it is possible that in some cases the minds of Martians and humans might breakdown in ways that are functionally equivalent. To use the computer analogy popular with functionalists, mental disorders might be caused not only by problems with the mind's hardware but also by problems with the mind's software. Many theorists think that autism occurs when something goes wrong with a human's "Theory of Mind Module", for example. Now, while it seems to me unlikely, it is possible that for some reason all minds tend to be functionally equivalent at all levels of design. That is, maybe the best design for any mind is one that incorporates systems that are functionally equivalent to a human theory of mind module and to whatever other modules and sub-systems human minds possess. If this were the case then Martians would also have theory of mind modules and, when such modules breakdown, they would suffer from some of the symptoms typical of autism. Would such a disorder in a Martian be a case of autism? Here my second argument against McGinn comes into play. I will argue that it is consistent to take a functionalist approach to some mental state talk, but to reject functionalism as an acceptable account of mental state talk in technical contexts.

To expand, I suggest that the functionalist is plausibly right in saying that there is a sense in which aliens, or computers, could have mental states. A Martian might have a inner state that fills the same causal role that folk psychology attributes to beliefs about chocolate ice-cream, say, and in such a case it seems right to say that in a sense the Martian has a belief about ice-cream. However, in technical contexts, claims are made about mental states that go beyond anything known to folk psychology. A psychologist may claim that desires can be modified by Pavlovian conditioning, or that true beliefs can usually be extracted under the influence of amytal, for example. While the Martian's beliefs and desires may fulfil the causal role allocated to such states by folk psychology, they are unlikely to behave entirely as the academic psychologist expects. We might thus be left wanting to claim that (a) it is true that beliefs can be extracted under amytal, and (b) there's no reason to

think that Martian beliefs can be extracted under amytal. I suggest that this apparent contradiction can be resolved if we recognise two senses of "belief". As "belief" occurs in commonsense contexts it is functionally defined. In such contexts a belief can be said to be any state that acts as folk psychology says beliefs should, and robots and aliens can have beliefs. However, when "belief" occurs in the mental sciences it should be taken as shorthand for "normal human belief", and such states are not functionally defined. The same goes for all mental states.

To return to the autism case, a Martian might suffer from a breakdown in their theory of mind module (supposing they had one) but I suggest they would not be said to suffer from autism. "Autism" is a term that is not part of folk psychology, but instead belongs to the mental sciences. As "autism" occurs in such sciences it is tacitly assumed that autism is a disorder of humans, not of Martians. As a consequence it makes sense for psychologists to search for the genes that cause autism, or to look for drug treatments that alleviate the symptoms of the disorder. I conclude that even if we are functionalists about mental state talk as it occurs in non-technical contexts, we should not be functionalists about mental state talk as it occurs in the mental sciences. In such technical contexts mental states are assumed to belong to humans. This implies that mental disorders are not functionally defined, and that McGinn has given us no reason to think that they cannot be natural kinds.

2.3.2 McGinn's Argument From Language

It has come to be generally accepted that our use of natural kind terms differs from our use of other terms in various ways. McGinn examines our use of mental terms, finds that this differs from our use of natural kind terms, and takes this to be a reason for doubting that mental kinds are natural kinds. After presenting McGinn's argument in greater detail I shall argue that he has not in fact shown that mental diseases are not natural kinds. This sub-section presupposes some familiarity with debates in the philosophy of language concerning the semantics of natural kind terms. Unfortunately, these debates are too complex for me to be able to outline them quickly here. Readers who have not heard of "twin-earth" are advised to accept my word that McGinn fails to show that types of mental disorder cannot be natural kinds, and to skip this sub-section.

From the literature on natural kind terms, McGinn draws up a list of nine characteristics of their use:

> (i) our initial criteria of recognition for membership in the kind are epistemically contingent; (ii) our original naive classifications of objects into natural kinds are susceptible of revision in response to scientific investigation of the kinds; (iii) there is the prospect of eliminating (ordinary language) natural kind terms in favour of nomenclature drawn from a scientific theory of the kinds; (iv) the equivalence relation that collects objects into a given natural kind is a theoretical relation; (v) we can construct plausible "Twin earth" cases for natural kind terms; (vi) the extension of a natural kind term is not fixed by the concepts speakers associate with the term ("meanings are not in the head"); (vii) natural kind terms exhibit a high degree of division of linguistic labour; (viii) a causal-historical theory of reference seems

> applicable to natural kind terms; (ix) the extension of a natural kind term is typically fixed by ostension (natural kind terms are indexical in some way).[41]

Our talk of mental states, claims McGinn, fails to display many of these features: Twin earth scenarios cannot be constructed for folk-psychological kinds; if, in another possible world, there are beings who display all the signs of believing that chocolate is the best ice-cream flavour, then even if they don't possess neurons, they really do have this belief. No one could convince us that we use folk-psychological concepts wrongly, as we are all experts in this area. As such, there is no division of linguistic labour, and folk-psychological terms could not be eliminated in favour of more "scientific" terms.

I do not wish to take issue with McGinn's claims regarding folk-psychological terms. Instead my criticism of his argument will focus on the way in which McGinn considers examples drawn from everyday mental talk and then takes his conclusion to be applicable to all mental talk what-so-ever. In taking his conclusions to have such broad scope McGinn has made a mistake. While I accept that everyday mental talk may be as McGinn claims, in line with my previous argument against McGinn, I will claim that our use of psychiatric terms, and indeed our use of terms from scientific psychology, is quite different

In contrast with our use of folk-psychological terms, our use of psychiatric terms displays precisely the features that McGinn considers to be characteristic of natural kind terms. It is common for lay-people to allow specialists to correct their use of psychiatric terms. Indeed popular magazine articles about mental illness regularly begin by chastising readers for using "depressed" as a synonym for "miserable", or for taking "schizophrenia" to refer to a condition in which someone has more than one personality. Psychiatrists are considered to be experts concerning the use of psychiatric terms. Moreover, as psychiatric knowledge increases classifications of mental diseases become vulnerable to revision. There is the prospect that ordinary language terms may be eliminated in favour of terms drawn from a scientific theory, and mental diseases can be classified on the grounds of theoretical relations.

Constructing a "Twin-earth" thought experiment for a mental disorder is rather difficult because while everyone agrees that water is H_2O the fundamental natures of most mental disorders are unknown. Still, it seems possible to construct twin-earth scenarios for those few diseases that are well understood. The twin earth thought experiment for water depends on the intuition that unless a liquid is H_2O then, even if it has all the superficial features of water, it isn't water. Similarly it is plausible that someone with the characteristics of Down Syndrome who had no chromosome abnormality wouldn't be said to suffer from Down Syndrome, and that someone can't be an alcoholic unless their condition is caused by alcohol. That a twin earth scenario can be constructed implies that a causal-historical theory of reference seems applicable to mental disease terms. This in turn implies that the extension of the kinds is not fixed by the concepts speakers associate with the term, and that the extension of the term is fixed by ostension.

[41] McGinn 1991 p.156

Thus I conclude that McGinn has been wrong in supposing that our use of all mental terms is like our use of folk-psychological terms. McGinn may be correct in claiming that our use of folk-psychological terms is unlike our use of natural kind terms, but our use of psychiatric terms is consistent with them being natural kind terms.

I have finally refuted all the arguments that purport to show that mental diseases *cannot* be natural kinds. It is now time to move on and consider whether it is likely that at least some mental diseases *actually are* natural kinds.

3. ARE TYPES OF MENTAL DISORDER NATURAL KINDS?

Before considering whether it is plausible that some types of mental disorder are natural kinds it will be useful to remind ourselves of the account of natural kinds developed earlier in the chapter. I argued that the best account of natural kinds is one according to which members of a natural kind possess similar, although not necessarily identical, important properties. These important properties determine many of the other properties possessed by the member of the kind. As such, I called them "determining properties". As members of a natural kind all possess similar determining properties they will have many other properties in common.

So far the examples of natural kinds considered - biological species, chemical elements, types of fundamental particle - have all been types of thing or stuff. Diseases should not be thought of in this way. True, some diseases are caused by entities (bacteria, viruses and so on) invading the body, but the disease itself should not be identified with these entities; if one has a colony of bacteria growing in a petri dish one doesn't have a colony of disease, but only a colony of disease causing entities. Rather diseases should be thought of as types of process. A cold, for example, is a process that occurs when the immune system fails to fight off cold viruses and a particular sequence of typical symptoms results.

The account of natural kinds I have proposed can be readily adapted to deal with natural kinds of process. To cope with natural kinds of process the dimensions of the multidimensional determining-property space must be taken to represent properties-at-a-stage. Instances of a natural kind of process will then be close together in the space. In addition to types of disease, natural kinds of process might include particular chemical reactions, for example rusting, and biological processes, for example the metamorphosis of some particular species of caterpillar into a butterfly.

For types of mental disease to be natural kinds the determining properties of instances of the disease must all be similar. Unfortunately many mental disorders are insufficiently well-understood for it to be possible to know whether or not this criterion is met. Plausibly, however, there are at least some mental disorders that meet this condition. Take Huntington's Chorea, for example. Huntington's Chorea is caused by a single dominant gene on chromosome four. Symptoms generally appear in middle-age and include jerky involuntary movements, behavioural changes, and progressive dementia. Plausibly Huntington's Chorea is a natural kind of mental disorder; in all cases an identical determining property, the defective gene, produces characteristic symptoms.

In addition to some mental diseases being natural kinds, it is plausible that some will turn out to be partial kinds (in the sense introduced earlier in the chapter). Diseases will be partial kinds where the determining properties of instances of a disease are similar in many, but not all, stages of the disease. To take a fairly well-understood physical disease as an example, a case of meningitis caused by alcohol and a case of meningitis caused by a viral infection have different causes although the remainder of the disease process is very similar. As another example, the set of determining properties of a particular case of A.I.D.S. might consist of infection with H.I.V., producing a reduction in the efficiency of the immune system, leading to infection with tuberculosis, leading to death. Again the determining properties of cases of A.I.D.S. will not be similar at all stages; sufferers will succumb to different infections. Where cases of a disease share some, but not all, determining properties the disease will not be a natural kind, but only a partial kind. Inductive inferences based on those determining properties that are similar in all cases will be sound, those that are based on determining properties that differ will not. Thus it is safe to infer that all A.I.D.S. sufferers will have low white blood cell counts, but not that all A.I.D.S. sufferers will develop red rashes.

It is plausible that some types of mental disorder are natural kinds, and that others are partial kinds. In addition there will almost certainly turn out to be some categories of mental disorder that are neither, that is mental disorders where cases do not possess any similar determining properties at all. Most obviously "rag-bag" diagnoses included in the D.S.M, such as "Sexual Disorder Not Otherwise Specified", will fall into this category, as there is no reason to think that cases that receive such a diagnosis will be similar to each other in any interesting way. In addition, future research may well find that cases receiving other more "respectable" diagnoses actually have nothing interesting in common. Many researchers hold that this is likely to be the case with schizophrenia, for example.[42]

To conclude this chapter, I have produced an account of natural kinds on which it is possible for types of diseases to be natural kinds. I have refuted arguments that purport to show that mental disorders cannot be natural kinds, and I have suggested that at least some types of mental disorder actually are natural kinds.

Many authors writing about kinds have worried about whether our classifications cut nature "at the joints".[43] On my account of natural kinds seeing the problem primarily in these terms is something of a red herring. The key worry is whether our classifications group together entities that are genuinely similar to each other. To stick with the meat metaphor, these days we know there are far worse things a butcher can do than cutting the joints poorly. If the meat we're given was originally some continuous piece of animal, cut at the joints or not, then we're doing quite well. The more worrying possibility is that our "joint" is made from reformed off-cuts, originating from numerous different beasts, and with water, rusk and additives thrown in. Similarly, we should primarily worry whether categories, such as

[42] *Schizophrenia Research* 1995 **17** pp.133-175 is devoted to the question of whether schizophrenia is a heterogeneous disorder.

[43] Haslam 2002 and Zachar 2000, for example, see the question of whether mental disorders are natural kinds in these terms.

"schizophrenia", group together cases that are actually similar to each other at a fundamental level, or whether we are lumping together cases that are fundamentally radically different. If schizophrenia fades into some other category, such as schizotypal personality disorder, then it will of course be as well to know this, but so long as cases of schizophrenia are fundamentally similar, schizophrenia is still a natural kind on my account. Even if schizophrenia fails to be discrete, knowing that someone suffers from schizophrenia could still be used as the basis of inductive inferences and function as an explanation.

This point can be made clearer by considering an analogy. British degrees are classified as third, second, or first class. There are two possible worries about this practice. The first, and radical worry, is that degree classifications might fail to group like with like. Marking might be fundamentally subjective. One marker might award a first, where another would award a third. If this were the case, degree classifications would be unsound. The second potential worry is that distinctions between degree classifications are artificial in the sense that they split a continuous range of student marks into arbitrary categories. Degree classifications fail to "cut nature at the joints". I suggest that this second worry is far less serious than the first. Admittedly dividing degrees into categories carries some risks – there is a danger that people come to believe that a second class and first class degree are radically different, when at least at the boundary they are not. Still, so long as marking is sound, all those who obtain a first are alike.

Whether and when it is a best to have a classification system that reflects the natural structure of a domain (that reflects the true joints, where there are any) will typically depend on a range of factors. It should be remembered that classification systems should not only provide information about the entities they categorise, but also need virtues that will enable them to be used in practice. In some cases it may be best to reflect the natural structure of a domain, in other cases it will be better to employ categories that make sharp divisions where naturally there are none. Debates over whether degrees should be classified, or raw marks recorded, illustrate that many factors may be involved in such decisions, that the factors to be considered will be particular to each case, and also, I suggest, that the issues will be largely empirical and pragmatic rather than philosophical. Here I have been concerned to show that in at least some instances it is plausible that all cases of a mental disorder are fundamentally similar. This implies that types of mental disorder can support inductive inferences, and function in explanations. They can be natural kinds on my account. Whether the best classification of mental disorders should be categorical or dimensional is a separate question from this, and not, I think, one that a philosopher can contribute much to answering.

Plausibly, I suggest, there are natural kinds of mental disorder. On my account of natural kinds, for a mental disorder to be a natural kind entails only that cases of that disorder are fundamentally similar. I take this to be a plausible, and fairly weak, claim. Despite this many people find it controversial. On several occasions when I have presented papers at conferences, people have come up to me afterwards and suggested that it is dangerous to hold that some mental disorders are natural kinds. Largely this seems to be because they think that claiming that mental disorders are natural kinds has unacceptable political consequences.

I accept that if types of mental disorder are natural kinds this may have political and social implications. Whenever it is discovered that some class of people behave in a characteristic way our reaction depends heavily on whether we think that these people form a natural kind. Consider the public reaction to statistics showing that children in some particular area do poorly at school and get into trouble with the police. We presume these children are fundamentally much the same as other children. We do not think that children in local authority X form a natural kind. As such, if children in local authority X are found to differ from other children in certain respects the cause of this difference is thought to lie in their circumstances rather than in them. Thus, statistics that show that such children do poorly are taken to show that there is something wrong with the local education system rather than that there is something wrong with the children.

In contrast, when it is reported that, say, seven-year old girls get better exam-results than seven-year old boys we think that there is at least a possibility that the cause of this difference lies in the girls themselves. Seven-year old girls and seven-year old boys are different kinds of person and so there is the possibility that their different results are the result of some intrinsic difference between them - maybe, for example, girls' brains mature faster than boys'. Often, of course, we won't actually be able to tell whether the difference is due to differences between the kinds of people or to a difference in circumstances. Thus, the difference between girls and boys may be due to girls' brains maturing faster, but then again it may be due to parents having greater expectations for girls, or to teachers spending more time with girls, or whatever. Still, that boys and girls are members of different natural kinds makes it possible that the inequalities between them are the result of their intrinsic differences. To put it crudely, when it is discovered that the members of a particular natural kind have different life-experiences from other people there is at least the possibility that this is no one else's fault. In contrast, if some disadvantaged group of people do not form a natural kind then it is likely that the cause of their differences lies not in them, but in their circumstances.

If types of people form natural kinds this has political implications as differences between members of distinct natural kinds may be due to intrinsic differences between them. These political implications are not as great nor as sinister as some have suggested, however. On occasion one is given the impression that those who believe that there are natural kinds of human have begun to travel down a short and slippery slope leading straight to fascism. In *The Disorder of Things*, for example, Dupré claims that when types of people are considered to form distinct natural kinds "it is inevitable that any systematic differences that are found will be taken to be explained, or explicable, in terms of the intrinsic differences between members of the two kinds."[44] This leads to the "legitimation of conservative politics and to the discouragement of proposals for significant social change".[45] Here Dupré is overstating the link. It is perfectly consistent to hold that men and women, say, form distinct natural kinds and to think that some of the differences between them are produced by sexist social structures. Thus someone who thinks that men and women

[44] Dupré 1993 p.253
[45] Dupré 1993 p.256

form distinct natural kinds might claim, quite sensibly, that men suffer from a far higher rate of testicular cancer because they have testicles, but also hold that sexism results in women being paid less, and sexually harassed more, than men. If types of people form natural kinds this opens up the possibility that differences may be due to their intrinsic natures, but it is by no means that case that all differences must be so explained. Similarly, it is possible to hold both that types of mental disorder may be natural kinds, and also to think that many of the problems faced by psychiatric patients are caused by social prejudice.

To sum up, there are plausibly natural kinds of mental disorder. As such, it is reasonable for psychiatrists to use empirical data in an attempt to find categories that group together cases that are fundamentally similar. The D.S.M. project of using empirical research to guide classification makes sense.

The upshot of the argument of the book so far is that we should think of mental disorders in a way analogous to the way we think about weeds. Weeds are unwanted plants, thus whether a daisy is a weed is at least in part a value-judgement. Still, types of plant that are generally considered to be weeds – daisies, buttercups, stinging nettles – are natural kinds. Similarly, I argue that the claim that schizophrenia is a disorder is in part a value-judgement, but that it may well be the case that schizophrenia and depression are natural kinds.

Still, there may be difficulties constructing a classification that reflects the natural similarities between types of mental disorders. In the next two chapters two potential sources of difficulty will be considered. These arise from the possibility that observation in psychiatry is theory-laden, and from the fact that the D.S.M. is shaped by pressures emerging from the various ways in which it is used in practice.