

This is an author's version of a chapter published in *Oxford Handbook of Philosophy and Psychiatry*. Fulford, K., Davies, M., Graham, G., Sadler, J, Stanghellini, G. & Thornton, T. (eds.) O.U.P. (2013). pp.950-964.

<https://global.oup.com/academic/product/the-oxford-handbook-of-philosophy-and-psychiatry-9780199579563?cc=gb&lang=en&>

Natural Kinds

Rachel Cooper

What are natural kinds? Are mental disorders natural kinds? Why does it matter? Let's start with rough and ready answers, and then assess complications later.

Paradigmatically, natural kinds are the kinds of thing or stuff that are classified by the natural sciences. The Periodic Table provides perhaps the best example of the potential importance of natural kinds for science. The Periodic Table provides a classificatory basis for chemistry that enables different types of stuff to be classified, and via this classification, for them to be understood and controlled. Thus, once I have determined that a particular chemical sample is lead, say, I know how it will behave and how to treat it if I wish to use it in various ways. Classification grounds explanations and predictions, and enables us to control a domain. If mental disorders are natural kinds, perhaps we can hope that one day psychiatric classification will ground psychiatric theory and practice in a way that approaches the successes of the Periodic Table in grounding chemistry.

In the philosophy of psychiatry, debates over whether mental disorders can be natural kinds emerge because kinds of mental disorder are manifestly different from chemical kinds in various ways. While chemical kinds are precise, psychiatric kinds are fuzzy. While chemical kinds are objective, the identification of psychiatric kinds is value-laden. Psychiatric classification involves classifying people, and unlike chemical elements, those people can respond to being classified in various ways. Later in this chapter I will go through these differences, one-by-one, and argue that despite them, mental disorders may be natural kinds.

Thus we have our rough and ready answers: natural kinds are kinds picked out by the sciences. Identifying natural kinds is worthwhile because such kinds can ground explanations and predictions and enable us to gain control over a domain. Although kinds of mental disorder differ from the kinds recognised by sciences such as chemistry in various ways, they may yet be natural kinds (though this will not be shown until later in this chapter). Now for the complications:

1. What are natural kinds? Three traditions distinguished.

In this chapter we will focus on natural kinds understood as the kinds that are picked out by scientific classifications. However, the literature on natural kinds can be hard to navigate as different authors mean different things when they talk of natural kinds and are interested in different sorts of problem. Other authors have also noted the heterogeneity of natural kind concepts and suggested various classifications (Haslam 2002a; Murphy 2006; Zachar forthcoming). I suggest that we can usefully divide the literature into three traditions:

First, and I think most importantly for the philosophy of psychiatry, there is the tradition on which we will focus, call it the kinds-in-science tradition (e.g. Dupré 1993, 2001, 2006). This tradition is impressed by the power of classification in science and is interested in those kinds that facilitate such successful classifications. Paradigmatic examples of natural kinds are taken to be chemical kinds and biological species. When writers in this tradition seek to understand natural kinds they seek to understand kinds like these, and how they can be employed in scientific practice.

Second, there is an Aristotelian tradition. In the Aristotelian tradition, talk of natural kinds is taken to be of importance not only for explaining the behaviour of members of a kind (as in the kinds-in-science tradition) but also for making sense of problems concerned with identity, development, and change (Brody 1973; Ayers 1981; Megone 1998). For Aristotelians, the character of an individual depends on what kind of thing it is, and the ways in which individuals can change while yet retaining their identity thus depends on the natural kind to which they belong. Thus a caterpillar changing into a butterfly continues to be the same individual, because such changes are part of the natural development of individuals of that type, while a caterpillar that is eaten by a bird ceases to be. Aristotelians take biological kinds to be key examples of natural kinds. Within the philosophy of psychiatry, Chris Megone employs Aristotelian traditions of natural kinds in making sense of mental disorder (1998, 2000). Megone argues that humans are essentially rational animals and that mental disorders can be understood as states that inhibit human flourishing. Aristotelian approaches

might also be used to make sense of some of the problems that mental disorders can raise for questions relating to the identity of persons. In some dissociative conditions, for example, we may wonder whether identity is destroyed or fragments. In so far as the Aristotelian tradition makes use of natural kind talk in understanding the development and destruction of individuals it might prove useful for exploring such issues. To date, and as far as I am aware, however, such work has yet to be undertaken.

Third amongst our traditions of natural kinds, there are New Essentialists (e.g. Ellis 2001, 2002). New Essentialists are principally interested in essences. An “essence” or “essential property” is a property that all members of a kind share that determines their nature. In the case of chemical elements, for example, the essence would plausibly be the atomic number. While Aristotelians also talk of essences they can be distinguished from New Essentialists as their very different metaphysical stance leads them to nominate very different candidates for “essences”. Aristotelians will suggest that “being a rational animal” might be the essential property of humans. New Essentialists think of essences as being the properties fundamental physics and chemistry find explanatory. New Essentialists have principally been interested in the metaphysical implications of a kind having such an essential property; for example, some have argued that natural laws are necessary. They have restricted their interest to those kinds, such as fundamental particles and chemical elements, that plausibly do have essences in their sense. In so far as other kinds, such as biological species, fail to have such essences, thinkers working in this tradition simply lose interest in them. This tradition is of the least interest for the philosophy of psychiatry as it is highly unlikely that kinds of mental disorder will have essential properties in the same sort of way as chemical elements.

These three traditions use the term “natural kind” slightly differently and are concerned with slightly different issues. Within the philosophy of psychiatry, confusions between them have resulted in much misunderstanding. Misunderstanding between those adopting a kinds-in-science approach (according to which natural kinds may or may not have essences) and essentialist approaches (on which natural kinds must have essences by definition) has resulted in much discussion failing to get beyond the stage where one author takes the plausible absence of essences to show that mental disorders cannot be natural kinds (e.g. Zachar 2000; Haslam 2002b), while another argues that mental disorders can be considered natural kinds on some non-essentialist account of natural kinds (e.g. Cooper 2005). In order to avoid such misunderstandings, when talking about kinds it is best to be explicit about what one has in mind, and also to bear in mind that there are various different usages in circulation.

Returning to the kinds-in-science tradition - How does identifying natural kinds play a role in science?

How does identifying natural kinds play a role in science? If we limit ourselves to thinking about kinds such as the chemical elements, the answer to this question may at first seem clear. Why is it we can expect all samples of some element to behave similarly? Because, all samples of an element share an “essential property”; they all have the same atomic number, and this ensures that they will have the same chemical properties. In theoretically important respects, all samples of a particular element are interchangeable.

Turn to biological species, however, and we will soon see that thinking in terms of essential properties will not quite do. Classifying biological individuals into different species has proved highly successful as a classificatory strategy; members of a species can be expected to behave in similar ways. However, plausibly it is not the case that all members of a species share some essential property. Within a species, diversity is the rule at both the genetic and phenotypic level. As John Dupré (1981, 1993) has powerfully argued there are simply no essential properties to be found.¹

Within the kinds-in-science tradition on which we are focussing, several accounts of kinds have been developed with the aim of explaining how it is that kinds like biological species can successfully ground explanations and inductive inferences even though members of the species do not share some essential property. In so far as any kinds of mental disorder might be expected to be rather like other biological kinds these accounts are of particular interest for the philosophy of psychiatry.

John Dupré has offered an account that he calls promiscuous realism (1981, 1993). He asks us to consider the entities of some domain mapped into a multidimensional space where the

¹ Dupré argues that there are no necessary and sufficient criteria for species membership. It will not do to say that members of a species can interbreed. Not only is such a criterion inapplicable to asexual species, but it also runs into problems dealing with sterile organisms, hybrid organisms etc. It will not do to rely on criteria of ancestry. While it is true that rabbits have rabbit ancestors and hares have hare ancestors, this is not enough to distinguish rabbits from hares, as some other criterion will be required to distinguish the ancestor rabbits from the ancestor hares. Nor can measures of genetic or phenotypic similarity be used to pick out co-members of a species. Some species are more heterogeneous than others, so there is no level of difference that is necessary and sufficient to mark species boundaries.

different dimensions map onto different properties (as in cluster analysis). Entities that are similar to each other will form clusters in such a space. Dupré suggests that kinds such as biological species can be identified with some such clusters. Of course, in the multidimensional space, not only biological species, but also multitudes of other clusters may be identified – some will correspond to classifications at levels higher or lower than species, for example, families, and varieties, will also be identifiable. The key claim for Dupré is that the world is such that some individuals are objectively similar to each other.² They share similar properties and will thus behave alike. In my 2005 *Classifying Madness* I argue that a Dupré-style account can fruitfully be applied to kinds of mental disorder.

In another cluster-type account, Richard Boyd has argued that we might usefully think of biological species as being “homeostatic property clusters” (1988, 1991). Like Dupré, Boyd argues that members of a species share a cluster of properties, but in addition Boyd emphasises that this is for a reason. Homeostatic mechanisms ensure that members of the kind will continue to be alike – in the case of biological species these mechanisms include gene-flow between members of the species, and environmental pressures that mean that those organisms which survive must all be capable of surviving in the same environmental niche. The difference between Dupré's account and Boyd's is that Boyd requires homeostatic mechanisms to “glue together” a property cluster, whereas Dupré requires no glue. In the philosophy of psychiatry, Dominic Murphy suggests that Boyd's account of natural kinds might accommodate certain mental disorders (Murphy, 2006, pp.338-341).

A further account of biological kinds has been produced by Ruth Millikan (1999). She emphasises the role of copying mechanisms that make it the case that biological kinds are fundamentally historical kinds, with the similarity of organisms of a species ensured by the fact that copying mechanisms make offspring like their parents. Turning to mental disorders,

² The claim that some pairs of entities are objectively more similar to each other than other pairs is common to all accounts of natural kinds. To illustrate, two twin tigers would be said to be more similar to each other than some other pairs of entities, for example a tiger and a balloon. Such similarities are seen as objective features of the world. The tiger twins share more properties than do the tiger and the balloon. Such claims are compatible with many metaphysical accounts of properties, but not with all of them. In particular, there are certain nominalist positions on which the idea that some pairs of entities are more similar than others makes no sense (for example, Goodman 1972). Discussing the details of the various accounts of properties is far beyond the scope of this chapter. Interested readers might consult Mellor and Oliver 1997, or Armstrong 1989.

copying mechanisms may also play a role in explaining why cases of a kind are alike. Ian Hacking has developed a number of case studies of epidemic mental disorders where unconscious copying mechanisms result in similar cases occurring (Hacking 1995a, 2010). Marion Godman (2011) is currently developing the idea that certain kinds of mental disorder can best be understood as historical kinds.

Plausibly different accounts might work best for different mental disorders. Millikan-style copying, for example, will clearly have a greater role to play in some mental disorders than others. There is still work to be done figuring out exactly which account of kinds will work best for which kinds of mental disorder.

2. Reasons why natural kinds of mental disorder might seem problematic

A number of writers have suggested that kinds of mental disorder cannot be natural kinds. In this section I examine their arguments.

On Gaps

It is frequently assumed that natural kinds should be discrete – that is, when the members of any two natural kinds are plotted in a multidimensional space, there should be a gap between them (Mill, 1973, p.123; DeSousa, 1984, p.565; Haslam 2002b; Reznik, 1987, p.42; Samuels 2009). I suggest, however, that gaps, where they occur, are not important. The important thing about natural kinds is that members of a natural kind are all objectively similar to each other. The basic idea is that the causal structure of the world is such that certain entities are to a large extent interchangeable, in the sense that their similar properties mean that they can be expected to behave in much the same fashion. Thus, once I have learnt how to grow one radish seed, I will be able to grow any radish seed, because they really are all much the same – the similar causal natures of the seeds mean that they will need the same sorts of environmental conditions to flourish. When it comes to grounding predictions and explanations and enabling us to control the world, it's the similarities between members of a kind that do all the work. Some kinds are gappy (e.g. chemical elements, as atomic numbers only come in integer numbers) and some kinds vary along dimensions (e.g. alloys), but this difference doesn't much matter. Alloys provide nice examples of continuously varying kinds that can yet ground explanations and predictions. If I know the make-up of a sample of alloy I can predict its properties just as accurately as if I

know the identity of a sample of pure metal. For this reason I suggest that we should consider kinds that vary along dimensions to also be natural kinds. Given that such dimensional kinds can do all the important work of traditional discrete natural kinds there is no benefit in restricting the term “natural kind” to discrete kinds.

Turning to consider mental disorders, discussion by those who argue that mental disorders might in some cases be better represented by a dimensional, as opposed to a categorical, classification system has revolved around two sorts of case. First, there are cases where one type of disorder seems to merge into another – thus for example, depressive disorders might run into anxiety disorders. Second, there are cases where a disorder fades into the normal range. Once again depression provides an example, as there seems to be no natural dividing line between normal unhappiness and mild depression. In both cases, I suggest, that the really important question is whether cases that are classified together genuinely share properties. Whether there are any sharp boundaries that can be drawn between the kind “depression” and other kinds is then a distinct, and less important, question.

On Values

On many accounts, a condition is only a disorder if it is a bad thing (Flew 1973; Fulford 1989; Reznek 1987; Wakefield 1992; Cooper 2002). Given that disorders are defined partly in value terms, but that natural kinds need to be defined with regard to natural properties, it may thus look like types of disorder cannot be natural kinds (as an example of someone who takes this line of argument see Peter Zachar (2000b.)).

We can respond to this worry by thinking through an analogy. Weeds are unwanted plants, and so whether a particular plant is considered a weed or a flower can vary with the tastes of the gardener. The umbrella category “weed” is defined in terms of values and is not a natural kind. However the different species of weed, such as dandelion and dock, are still natural kinds. Although whether a particular plant counts as a weed depends on values, the fact that it is a dandelion, or a dock, depends solely on its natural properties. Similarly, while the category “mental disorder” is value-laden and does not form a natural kind, conditions that are commonly disorders – schizophrenia, depression, and so on – may still be natural kinds. To complete the analogy, let's imagine that some particular process underpins cases of schizophrenia. Let's suppose that such a process occurs within some individual, but in that person the process does no harm – they hear voices but are not harmed by their condition. In such a case, I suggest we could say that the individual has schizophrenia, but not a disorder. While schizophrenia is frequently a kind of disorder, in cases where it does no

harm, it might simply be considered a kind of difference. In the same way, while dandelions are generally weeds, the dandelions in my wild flower garden are not weeds, though they are still dandelions. I conclude that types of condition that are usually mental disorders might be natural kinds even though the umbrella category “disorder” is not a natural kind.

My reasoning here would imply that someone could have schizophrenia and yet not be mentally ill. Some would take terms such as “schizophrenia” to themselves be value-laden and would say that someone biologically and psychologically of the “schizophrenic-type”, but who is not harmed by their condition does not have schizophrenia. I suspect that current concepts of “schizophrenia” are insufficiently defined for it to be clear whether the term is itself value-laden, or whether it is a purely descriptive term that falls under a value-laden umbrella category (as the “weed” analogy would suggest). Building on work by Joseph Laporte (2004), I think it likely that the extension of such terms will become more precise in the future as the relevant linguistic communities reach a consensus on how such terms should be used.

In his book *Natural Kinds and Conceptual Change* (2004), Laporte uses case studies to examine controversies that have emerged in the history of science because the extension of terms is sometimes not as precise as emerging circumstances require. For example, he considers how the scientific community reacted to the discovery that samples of jade fall into two chemically distinct kinds. Laporte argues that prior to the discovery it was indeterminate whether “jade” referred to all samples of a particular chemical structure or to all samples with particular superficial characteristics. Following the discovery that samples of “jade” fall into two chemical varieties it was necessary for the fuzziness of the extension to be clarified and it was eventually decided that “jade” would apply to both varieties.

I suggest that the discovery that some voice-hearers are not harmed by their condition brings out indeterminacies in the extension of “schizophrenia” in a way analogous to that in which the chemical discoveries brought out indeterminacies in the extension of “jade”. Whether one should think of terms like “schizophrenia” as value-laden or, as the weed analogy suggests, as a purely descriptive term that falls under a value-laden umbrella term will ultimately be a matter for decision by the relevant linguistic communities (primarily mental health professionals, researchers, and service users). The factors to be weighed in making such a decision will be complex. Still, the weed analogy shows that it would be possible to precisify terms like “schizophrenia” and “depression” in such a way that they became confirmed as natural kind terms.

On Cultural Shaping

The “natural” in “natural kind” should be read as in “natural law” as opposed to “present in the garden of Eden”. Some natural kinds are manmade; plutonium is an example. Still, there might be thought to be something problematic about the extent to which kinds of mental disorder are shaped by culture. Plausibly, mental disorders have varied greatly across cultures and history. This may lead one to doubt that natural kinds of disorder can be picked out. Maybe the disorders that are found in one context are simply different to those that are found in another? Depending on the sorts of cultural shaping that occur, different responses to this worry are appropriate.

Superficial variation

As an example of superficial variation consider how the content of delusions varies with time and place. In Europe, in the early modern period, there were people who believed themselves to be made of glass or earthenware (Speak 1990). Nowadays deluded people have different fears. Such variation is easy to understand. It's commonly the case that the superficial properties of members of natural kinds vary with environmental conditions. For example, apple trees can be grown tall or flat against walls depending on how they are pruned. Variation at a superficial level is fully compatible with types of mental disorder being natural kinds.

Deeper cultural moulding

More profound types of cultural moulding may also occur. In *Creating Mental Illness* (2002) Allan Horwitz makes a convincing case that “most nonpsychotic symptoms stem from general underlying vulnerabilities that may assume many different overt forms, depending on the cultural context in which they arise...Cultural processes, not the unfolding of natural disease, structure the overt manifestation of symptoms into recognizable entities” (p.108). Horwitz argues that whether a vulnerable and distressed person manifests a disorder characterised by depression, or anxiety, or somatization, or some other symptom, depends on their cultural context. If Horwitz is right, then not only “superficial” properties shift with cultural setting.

In thinking through such cases of “deep moulding”, considering some of the kinds that occur in other natural historical sciences can be illuminating. Specifically, let us consider the different sorts of igneous rocks that are recognised by geologists. These rocks are all formed

from magma. All the different igneous rocks are made from the same basic stuff, but their characteristics vary depending on the conditions under which they were formed. The size of the crystals in the rock depends on the rate of cooling for example. Igneous rocks are classified according to their chemical composition and their history (both of which can vary continuously). Classifications of rock are complex. Still, the different kinds of rock can be considered natural kinds. Samples of a kind of rock are objectively similar to each other and distinguishing rock kinds is useful for grounding explanations and inductive inferences.

If Horwitz is right and different anxiety-depression type disorders are formed into distinct entities by cultural context, then we can think of such disorders as being kinds analogous to the different kinds of rock distinguished by geologists. Admittedly such historical natural kinds may only occur under certain conditions (in *Mad Travelers* (1998) Ian Hacking shows this is the case for fugue, for example). Still, though such disorders may occur for a limited time or in limited places, within those constraints the kinds operate like normal natural kinds. Historical natural kinds – such as kinds of rock, and culturally formed type of mental disorder, can usefully be considered natural kinds, I suggest, because the kinds can support explanations and inductive inferences and feature in law-like generalisations. The individuals that fall into such kinds are “repeatables” in the sense that any two specimens of basalt, or any two cases of fugue, can be expected to have much in common.

At this point some may worry that in suggesting that even some culturally formed mental disorders can be considered natural kinds, I have come a very long way from what many have meant when they talk of natural kinds. My kinds need not have essential properties, can vary along continua, and can be historically contingent, in that they may only arise under certain historical conditions. The reason I think it's reasonable to call such kinds natural kinds, is that they are up to the job of grounding explanations and predictions. To take an example, anorexia may plausibly be a culturally formed mental disorder, and yet is the sort of kind that can help ground psychiatric science. We can know that anorexia is hard to treat, anorexia is very dangerous, many women with anorexia will cease menstruation, and so on.

Interaction

Over the last few decades, Ian Hacking's work has stressed the importance of the fact that humans respond to being classified in ways that other classified entities do not (1986, 1988, 1992, 1995a, 1995b). A child who is told they are stupid may stop trying at school and fall behind yet further; a diagnosis of “problem drinking” may come to motivate abstinence; a whole class of people may respond to a classification with new forms of resistance, as in “fat

pride". Such interactions between classifications and behaviour mean that "human kinds" - the kinds classified by the human sciences - become moving targets. No sooner has a kind been picked out than behaviours shift and classifications have to be revised.

One of Hacking's best developed examples of such "looping effects" concerns Multiple Personality Disorder (1995a). When cases of Multiple Personality Disorder were first reported, a multiple would typically possess just two or three clearly distinct personalities. Over time, however, the symptoms of multiples shifted. Hacking makes a convincing case that the shift in symptoms was in part caused by changing prototypes of the disorder being made available in the media. The media tended to report more florid cases, and over time multiples started to present with more and more personalities, and as their numbers increased, these personalities became more diverse and also more fragmentary. Note that Hacking's claim is not that patients intentionally copy the symptoms of publicised cases. Rather the mechanism is more subtle and subconscious, but still the consequence is that a distressed individual will most likely manifest distress in ways that are culturally recognised.

At certain points in his work, Hacking has claimed that interaction between kinds and their classification, as seen in the case of Multiple Personality Disorder, marks an important distinction between natural kinds and human kinds, such that human kinds cannot be natural kinds. Previously, I have argued that Hacking is wrong on this point (Cooper 2004). The gist of my argument is this: It is true, as Hacking, has claimed, that human kinds shift in response to classificatory practices, and this requires classifications to be updated. However, this is not sufficient to show that human kinds cannot be natural kinds. Other types of natural kind also shift in response to pressures that only affect kinds of their particular type. For example, types of domestic animal and plant shift as a result of selective breeding and only types of domestic animal and plant can be selectively bred (Boyd 1991). It is of course important to note that particular types of kind are vulnerable to shifting under different types of pressures, but there is no reason to think that these differences mark any fundamental metaphysical distinctions.

Hacking also proposed a supplementary argument, which used Elizabeth Anscombe's claim that intentional actions are intentional under a description to argue that the new descriptions formulated by the human sciences made new types of action logically possible (Hacking 1986, 1995a, Anscombe 1957). I have argued that this argument is based on a misinterpretation of Anscombe's work (Cooper 2004). Her claim that intentional actions are only intentional under-a-description should be interpreted as being equivalent to the claim that an intentional actions is only intentional qua some aspect (an example she gives is one

where a bird intends to land on the twig qua a way to get the seed, but not qua a way to land in the bird trap) (Anscombe 1971). This translation of Anscombe's claim makes it clear that formulating new descriptions does not make new actions logically possible.

In his most recent work, Hacking has himself shifted away from talking of natural and human kinds on the basis that talk of "natural kinds" has become so laden with metaphysical baggage that the term is now best avoided (2007). This is a claim with which I am sympathetic, although the approach suggested in this chapter is that rather than jettisoning the terminology one should be explicit about what one has in mind when talking about natural kinds.

Functionally Defined Kinds

A number of writers have argued that psychological kinds cannot be natural kinds because they are functionally defined (McGinn 1991). Functionalists about the mind claim that mental states are characterised by their causal role. That is, the nature of a mental state is fixed by the types of stimuli that typically produce it, its causal relations with other mental states, and the types of behaviour that it typically produces. Thus, for example, fear is a state that is characteristically produced by stimuli like charging bulls, snarling dogs and aggressive gun men, interacts with other mental states, such as the belief that help can be summoned, and leads to behaviour like screaming for help and running away. Functionalism implies that mental states can be multiply realised. Any state that fits the right causal role counts as a mental state, no matter what its physical realisation. Thus, while my fear of dogs is realised by some neural state, your fear might be realised by some quite different brain state, and a robot's fear would be realised by electronics. Given that cases of the same psychological kind (e.g. fears) can be physically unlike, these kinds look very unlike prototypical natural kinds, where the similar behaviour of members of the kind occurs because the members are physically similar.

This problem can be dealt with in at least two ways. First, and most simply, we can note that the claim that mental states are theoretically multiply realisable is compatible with all human mental states being realised in much the same way (Kim 1993, pp.305-335). In robots and martians fears may be realised by all sorts of different systems, but in humans all fears may be linked to some particular anatomy. This means that human psychological states of a kind may all be physically alike.

Second, we can note that even when the kinds of some domain are functionally defined when they are working properly, the kinds of breakdown that occur need not be functionally defined. Consider electronic components for example. These are functionally defined – anything that behaves like a capacitor is a capacitor, and capacitors can be made of different materials. Still, the ways in which capacitors can break down depend on the physical stuff that different types of capacitor are made of; for instance some are brittle while others are not. In so far as we might think of mental disorders as arising when normal mental functioning breaks down it is consistent to think that, even if normal mental states are functionally defined, abnormal ones might not be. For example, I might be a functionalist about normal beliefs and desires, and yet also think that human mental states are vulnerable to certain types of disruption that are characteristically caused by drinking too much alcohol. In the same sort of way that only brittle capacitors are vulnerable to breaking by smashing, only thinkers with a certain biology will be vulnerable to certain sorts of mental disruption. Being a functionalist about normal mental states is thus compatible with thinking that kinds of mental disorder may be natural kinds.

Admittedly, being a functionalist about normal mental states is also compatible with thinking that mental disorders are functionally-defined. In order to motivate this position, though, some further reason for thinking that mental disorders are functionally-defined would need to be provided. David Papineau (1994) presents an argument for thinking that any disorders that can be characterised as stemming from dysfunctional patterns of learnt behaviour and thinking will be functionally defined. However, the scope of Papineau's argument is limited, as many mental disorders cannot plausibly be seen to have their origins in faulty learning.

3. On finding natural kinds of mental disorder

When seeking natural kinds the aim is to find categories that map the causal structure of the domain being classified. How might natural kinds of disorder best be identified? The distinction between the context of discovery and the context of justification has fallen from favour in recent philosophy of science, but appealing to something like it is here useful. The basic thought is as follows: The key task is to group cases together in such a way that co-members of a category really are importantly similar to each other. Co-members of a category should share properties that mean they can be expected to behave in similar ways. Depending on one's account of kinds, one might also require that these similarities can be explained by the existence of homeostatic mechanisms or via copying. How such a classification is achieved doesn't matter.

Within current psychiatry, research traditions seek to construct classifications in various ways. Most dominant is the approach associated with the DSM, but there are also competing traditions that propose that classifications might be developed using the methods of numerical taxonomy, or propose radical overhauls to classification on the basis of some theory or other.

Though they may not describe their aim in these terms, I take it that research programmes such as that associated with the DSM aim at discovering natural kinds of mental disorder. A basic assumption of the DSM project is that discovering kinds of mental disorder will be important for grounding psychiatric theory. Furthermore, the procedures for revising the DSM assume that fixing the boundaries between kinds will be informed by empirical evidence. In seeking natural kinds of disorder, the DSM has tended to rely on tradition which is then revised as more and more empirical data is found. The sorts of evidence appealed to when the DSM is revised (rates of co-morbidity, family studies, drug response, differences in age of onset etc) can reasonably be hoped to enable us to map the causal structure of the domain of mental disorders. One might have concerns about the ways in which non-scientific factors might affect the process of DSM revision. Plausibly the classification has been affected by lobbying that is politically or financially motivated (Cooper 2005, Kutchins and Kirk 1997). Since the days of the DSM-III, however, the processes for revising the DSM have become less open to distortion. For example, committees are now expected to publish details of the literature on which they have based their decisions (in the DSM-IV *Sourcebooks* (Widiger, Frances and Pincus 1994,1996,1997), and online for the DSM-V) and those serving on the committees responsible for revisions are expected to limit their financial links with the pharmaceutical industry, (see e.g. guidelines of committee membership APA 2010). One might still worry that, in so far as the default position is that disorders remain between successive editions of the DSM, problematic categories inherited from DSM-III will remain. Still, the basic approach of the DSM-system to seeking natural kinds – start with a classification system and revise it as new evidence suggests – is a reasonable way to seek to achieve a classification of natural kinds, at least so long as one assumes that the traditional classification system from which the DSM has tried to progress via incremental stages is on roughly the right tracks. One worry is that if the initial classification was thoroughly misguided then the DSM process of revision, which allows revisions only when an advance over the existing classification can be proven, may not allow the classification system to ever reach an optimal state. Rather the classification could get stuck at a suboptimal point, in the same way in which evolving organisms can get stuck at local maxima in fitness space.

Worries such as this lead those who do think that the starting point for the DSM is likely unsatisfactory to suggest full scale overhaul. Some theorists have suggested classifications based on some overarching theory— maybe, evolutionary theory (Murphy 2006) or developmental approaches. In so far as such classifications depend on the theoretical approach used to develop them they can only be expected to be as good as the theory behind them. Alternative approaches to classification involve the use of statistical methods to find kinds of disorder from raw data. On occasion, the proponents of “numerical taxonomy” have claimed that their approach is purely empirical and generates theory-free classification systems (e.g. Sokal and Sneath 1963). The claim that the techniques of numerical taxonomy are theory-free is misguided. Before the techniques of numerical taxonomy can be applied one must decide which properties will be entered into the analysis, and decide which of the various statistical techniques to apply. One’s theories will shape decisions at both these levels (Cooper, 2005, ch. 3). Still, though they are not theory-free, the techniques of numerical taxonomy offer one approach to seeking natural kinds of disorder.

What if the categories developed by different classificatory approaches fail to correspond to each other? For example, what if a classification that is developed on the basis of treatment response fails to correspond to that developed by geneticists, which in turn fails to correspond to that used by those taking a developmental perspective? We can note that such a situation also occurs in other sciences. Within biology, for example, Dupré has convincingly argued that the species concepts that are required in different areas of biological research fail to correspond to each other (Dupré 2001). While ecologists find it most useful to classify species on the basis of current characteristics, evolutionary theorists find it better to classify on the basis of ancestry. Dupré suggests that in such a situation different scientific sub-disciplines should be free to classify as they find most useful. On Dupré’s metaphysical picture, the world is a complex place. Many categories can usefully be picked out for different scientific purposes, and so there are multiple sets of natural kinds that different subdisciplines might find it useful to classify.

4. Which disorders aren’t natural kinds

In this chapter we have come a long way from the traditional idea that natural kinds will be eternal, discrete, and possess essential properties. I have argued that a looser notion of natural kinds is sufficient to give us kinds that can do the important work of grounding inductions, explanations and predictions. We can say that natural kinds are groups of entities that are genuinely importantly similar to each other (and where, depending on one’s’

account, these similarities might be explained by the existence of homeostatic mechanisms or via copying). If we take this approach, are there any kinds of mental disorder that will fail to be natural kinds?

Finding natural kinds of mental disorder can still be expected to be difficult, and some current categories of mental disorder will fail to be natural kinds because they fail to group together cases that are similar to each other in any causally important respect. Most obviously, ragbag categories included in the DSM for completeness, such as Psychosis NOS, will fail to be natural kinds for this sort of reason. It may also turn out that some prima facie more respectable diagnoses fail to pick out natural kinds of disorder because they lump together heterogeneous cases. For example, if schizophrenia turns out to be an umbrella term for a number of conditions with differing underlying causal structures then schizophrenia would fail to be a natural kind.

5. Implications of mental disorders being natural kinds

If types of mental disorder are natural kinds, what are the implications? Occasionally, it is claimed that if types of people fall into kinds then there are ethical or political implications. In *The Disorder of Things*, 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” (1993, p.253) This leads “to the legitimization of conservative politics and to the discouragement of proposals for significant social change” (1993, p256). Here, I think Dupré is simply mistaken. Take an example of human natural kinds – men and women - and consider some of the systematic differences between them. On average, women give birth to more children and are paid less than men. Here we have no problems recognising that some but not all of the differences are due to intrinsic differences, and that some but not all of the differences might be ameliorated by progressive social policies. Believing in human natural kinds is compatible with holding any range of political views.

One implication that I think is important is that if types of mental disorder are natural kinds then this means that there may be grounds for optimism that one day successful therapies will be developed that will enable the mass treatment of disorders. If mental disorders are natural kinds, then this means that one case of a kind can be expected to behave like other cases of that kind. All cases of a kind will be alike in important respects. This means that a treatment that works for one of the kind can be expected to work for all. As a consequence, if mental disorders are natural kinds, then we can hope that “black-box” therapies may one

day be available. A black-box technology is one that a consumer can simply buy off-the-shelf (so named because they are typically sold in a black box) (Mackenzie 1993). Black-box technologies may originally have been hard to develop, but have now been perfected so that they can be produced on an industrial scale, and delivered in a form that can be used reliably by people who don't understand how they work. Lasers offer an example. Originally getting lasers to work was very difficult, but now they can be bought off-the-shelf. Successful drug therapies would provide the most straightforward example of black-box therapies. Developing drugs is of course difficult. However once the right chemical has been found, drug treatments can ideally be refined to the stage where they can be produced on an industrial scale and taken with reliable effect by people with little understanding. Think of paracetamol, or the contraceptive pill.

It will only be possible to develop black-box therapies for mental disorders if the disorders are natural kinds. The therapies can only be developed to work reliably in so far as the problems of those in the treatment group are all fundamentally similar. Note that although drug therapies offer the clearest promise of black-boxability, other forms of therapy might also be black-boxable. Suppose it turns out that depression can reliably be cured if a person plays football for half an hour a day, and spends an hour talking to others. Such a therapy would be black-boxable in my sense, as it is the sort of therapy that can be packaged such that it can be reliably reproduced by unskilled therapists (or reliably used for self-treatment). Ultimately the reason why it matters whether mental disorders are natural kinds is that if it is possible to classify mental disorders in such a way that cases that are importantly theoretically similar are classified together, then it may be possible to develop treatments that can successfully treat all cases of a kind.

6. Conclusions

In this chapter we have examined accounts of natural kinds, and asked whether types of mental disorder might be natural kinds and why it matters. We started by noting that different theorists define "natural kind" in different ways and are interested in different problems. One of the key claims of this chapter is that the variety of uses of the term "natural kind" means that it is important to be explicit exactly what one means when talking about natural kinds and to be clear what points are at issue. I have suggested that if we are interested in natural kinds in so far as they support explanations, inductions and predictions then the key question for the philosophy of psychiatry is whether it will be possible to classify mental disorders in a way that maps the causal structure of the domain of mental disorders. The aim is to classify cases together when they are similar to each other in causally important ways

and to classify them apart when there are no such similarities. Depending on the models of mental disorder that turn out to be correct, the similarities between cases of mental disorder that are important might be similarities with regard to neurotransmitter levels, or genetic abnormalities, or developmental history, or patterns of learnt responses, or whatever, or some combination of such similarities.

We examined various objections to mental disorders being natural kinds (in this weak sense) and showed how they could be overcome. We argued that classifications that distinguish natural kinds of mental disorder might be created in a variety of ways. We saw that the implications of there being kinds of mental disorder are important and yet ethically and politically limited. If there are natural kinds of mental disorder then we can hope that a treatment that successfully treats one-of-a-kind might also treat others-of-that kind. Ultimately this is why searching for natural kinds of mental disorder is worthwhile.

References

American Psychiatric Association (2010) Board of Trustee Principles available at <http://www.dsm5.org/about/Pages/BoardofTrusteePrinciples.aspx>

Anscombe, E. (1957) *Intention*. Oxford: Basil Blackwell.

Anscombe, E. (1971) Under a description. Reprinted in E. Anscombe (1981) *Metaphysics and the Philosophy of Mind*. Oxford: Basil Blackwell.

Armstrong, D. (1989). *Universals: An Opinionated Introduction*. Boulder: Westview Press.

Ayers, M. (1981) Locke versus Aristotle on natural kinds. *Journal of Philosophy*, 78, 247-272.

Brody, B. (1973) Why settle for anything less than good old-fashioned Aristotelian essentialism. *Nous*, 7, 351-365.

Boyd, R. (1988) How to be a moral realist. In G. Sayre-McCord (ed.) *Essays on Moral Realism*. Ithaca: Cornell University Press. Pp181-228

Boyd, R. (1991) Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127-148.

Cooper, R. (2002) Disease. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 33, 263-282.

Cooper, R. (2004) Why Hacking is wrong about human kinds. *British Journal for the Philosophy of Science*, 55, 73-85.

Cooper, R. (2005) *Classifying Madness*. Dordrecht: Springer.

De Sousa, R. (1984) The natural shiftiness of natural kinds. *Canadian Journal of Philosophy*, 14, 561-580.

Dupré, J. (1981) Natural Kinds and Biological Taxa. *The Philosophical Review*, XC, 66-90.

- Dupré, J. (1993) *The Disorder of Things*. Cambridge, Massachusetts: Harvard University Press.
- Dupré, J. (2001) In defence of classification. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 32, 203-219.
- Dupré, J. (2006) Scientific Classification. *Theory Culture Society*. 23, 30-32.
- Ellis, B. (2001) *Scientific Essentialism*. Cambridge: Cambridge University Press.
- Ellis, B. (2002) *The Philosophy of Nature*. Chesham: Acumen.
- Flew, A. (1973) *Crime or disease?* London: The Macmillan Press.
- Fulford, K.W.M. (1989) *Moral theory and medical practice*. Cambridge: Cambridge University Press.
- Godman, M. (2011) Grounding natural kinds in the “special sciences”. PhD thesis submitted to King's College London.
- Goodman, N. (1972) Seven strictures on similarity. In *Problems and Projects*. Indianapolis: Bobbs-Merrill. Pp.437-447.
- Hacking, I. (1986) Making up people. In T. Heller, M. Sosna and D. Wellberry (eds.) *Reconstructing Individualism*. Stanford, California: Stanford University Press. pp.222-236.
- Hacking, I. (1988) The sociology of knowledge about child abuse. *Nous*, 22, 53-63.
- Hacking, I. (1992) World-making by kind-making: Child abuse for example. In M.Douglas and D.Hull (eds.) *How Classification Works*. Edinburgh: Edinburgh University Press. pp.180-238.
- Hacking, I. (1995a.) *Rewriting the Soul*. Princeton: Princeton University Press.
- Hacking, I. (1995b.) The looping effects of human kinds. In D. Sperber, A. Premack (eds.) *Causal Cognition*. Oxford: Clarendon Press. pp.351-394.
- Hacking, I (1998) *Mad Travelers*. Charlottesville: University of Virginia Press

Hacking, I. (2007) Natural kinds: Rosy dawn scholastic twilight. *Royal Institute of Philosophy Supplement*, 61, 203-239.

Hacking, I. (2010) Pathological withdrawal of refugee children seeking asylum in Sweden. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 41, 309-317.

Haslam, N. (2002a) Kinds of kinds: A conceptual taxonomy of psychiatric categories. *Philosophy, Psychiatry, Psychology*, 9, 203-217.

Haslam, N. (2002b) Natural kinds, practical kinds, and psychiatric categories. *Psychology*, 13. Available at <http://psychprints.ecs.stonon.ac.uk/>

Horwitz, A. (2002) *Creating Mental Illness*. Chicago: University of Chicago Press.

Kim, J. (1993) *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kutchins, H. and S. Kirk (1997) *Making Us Crazy*. New York: The Free Press.

Laporte, J. (2004) *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

MacKenzie, D. (1993) *Inventing accuracy: a historical sociology of nuclear missile guidance* MIT Press.

McGinn, C. (1991) Mental states, natural kinds and psychophysical laws. In his *The Problem of Consciousness*. Oxford: Basil Blackwell. Pp126-152

Megone, C. (1998) Aristotle's function argument and the concept of mental illness. *Philosophy, Psychiatry, Psychology*, 5, 187-201.

Megone, C. (2000) Mental illness, human function and values. *Philosophy, Psychiatry and Psychology*, 7, 45-65.

Mellor, D.H. and A. Oliver (1997) *Properties*. Oxford: Oxford University Press.

- Mill, J.S. (1973) *Collected Works of John Stuart Mill*. Edited by J. Robson. London: Routledge.
- Millikan, R.G. (1999) Historical kinds and the “special science”. *Philosophical Studies*, 95,45-65
- Murphy, D. (2006) *Psychiatry in the Scientific Image*. Cambridge, Massachusetts: The MIT Press.
- Papineau, D. (1994) Mental disorder, illness and biological dysfunction. In A. Griffiths (ed.) *Philosophy, Psychology and Psychiatry. Royal Institute of Philosophy Supplement 37*. Cambridge: Cambridge University Press. pp. 73-82
- Reznek, L. (1987) *The Nature of Disease*. London: Routledge and Kegan Paul.
- Samuels, R. (2009) Delusion as a natural kind. In M. Broome and L. Bortolotti (eds) *Psychiatry as Cognitive Neuroscience*. Oxford: Oxford University Press. pp49-79
- Sokal, R. and Sneath, P. (1963) *Principles of Numerical Taxonomy*. San Francisco: W.H. Freeman and Company.
- Speak, G. (1990) An odd kind of melancholy: reflections in the glass delusion in Europe (1440-1680). *History of Psychiatry*, 1, 191-206.
- Wakefield, J. (1992) The Concept of Mental Disorder - On the Boundary Between Biological Facts and Social Value. *American Psychologist*, 47, 373-388.
- Widiger, T., A. Frances, H. Pincus et al. (1994) *DSM-IV Sourcebook Vol. 1*. Washington: American Psychiatric Association.
- Widiger, T., A. Frances, H. Pincus et al. (1996) *DSM-IV Sourcebook Vol. 2*. Washington: American Psychiatric Association.
- Widiger, T., A. Frances, H. Pincus et al. (1997) *DSM-IV Sourcebook Vol. 3*. Washington: American Psychiatric Association.
- Zachar, P. (2000a.) Psychiatric disorders are not natural kinds. *Philosophy, Psychiatry, Psychology*. 7,167-182.
- Zachar, P. (2000b.) *Psychological Concepts and Biological Psychiatry: A Philosophical Analysis*. Amsterdam: John Benjamins.

Zachar, P. (forthcoming) Beyond natural kinds: Toward a “relevant” scientific taxonomy in psychiatry. In. H. Kincaid and J. Sullivan (Eds.), *Mental Illness and Natural kinds*. Cambridge, MA: The MIT Press.