

# Psychiatric Diagnosis and Hermeneutical Injustice: The Impact of Biomedical Diagnoses on Personal Narratives.

## **Richard Hassall**

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy (PhD).

University of Sheffield

2022

# Psychiatric Diagnosis and Hermeneutical Injustice: The Impact of Biomedical Diagnoses on Personal Narratives.

# **Richard Hassall**

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy.

University of Sheffield Faculty of Arts and Humanities Department of Philosophy

September 2022

# **Declaration**

I, the author, declare that this thesis is my own work. I am aware of the university's guidance on the use of unfair means. This work has not been presented for any other award at this university or at any other institution.

Richard Hassall

Word count: 73,411 words, including abstract and footnotes, but excluding references.

## **Publication from this thesis**

A short summary of the claims presented in Chapters 5 and 6 has been published as: Hassall, R. (2020). The impact of a psychiatric diagnosis on the self-narrative of the recipient. History & Philosophy of Psychology, 21, 4-10.

# **Contents**

Declaration Declaration	3
Publication from this thesis	4
Acknowledgements	7
Abstract	8
Introduction	9
Chapter 1: The natural kind status of diseases	
1.1 Introduction	12
1.2 Conceptions of natural kinds	12
1.2.1 Essentialist accounts	14
1.2.2 Property cluster accounts incorporating causal processes	17
1.2.3 Non-causal property cluster accounts	25
1.2.4 Objections to non-causal accounts of cluster kinds	28
1.3 Diseases as natural kinds 1.4 Conclusion	32
1.4 Conclusion	40
Chapter 2: The natural kind status of psychiatric diagnostic categories	
2.1 Introduction	41
2.2 The natural kind status of mental disorders	43
2.3 Traditional approach to the validation of DSM categories	44
2.4 Critique of the traditional approach to validation	50
2.5 Understanding of DSM categories in clinical practice	54
2.6 Explanatory deficiencies of DSM categories	57
2.6.1 Maung's arguments	58
2.6.2 Murphy's "exemplar" argument	60 63
<ul><li>2.7 The natural kind status of schizophrenia</li><li>2.8 Conclusions</li></ul>	68
2.6 Conclusions	00
Chapter 3: The biomedical model in psychiatry	60
3.1 Introduction	69
3.2 Models of mental disorder	70
3.2.1 The biomedical/disease model in medicine	72
3.2.2 The biomedical/disease model in psychiatry	74
3.3 What sorts of things are diseases?	80
3.3.1 Conceptions of disease 3.3.2 The causal basis of diseases	80 86
3.4 Conclusion	89
J.4 Colletusion	0)
Chapter 4: Diagnosis in medicine and psychiatry	0.1
4.1 Introduction	91
4.2 The explanatory function of diagnosis in somatic medicine	92
4.2.1 Different meanings of 'diagnosis	92
4.2.2 Disease identification and explanation	93
4.3 Objections to the view of diagnosis as explanation	97
4.3.1 Non-explanatory diagnoses	97

4.3.2 Diagnosis as one element in clinical reasoning	100
4.4 The social consequences of diagnosis	102
4.5 Diagnosis as a speech act	105
4.6 Diagnosis as an institutional fact	108
4.7 Psychiatric diagnoses	112
4.8 Conclusion	115
Chapter 5: The effect of a psychiatric diagnosis on the recipient's self-narrative	
5.1 Introduction	117
5.2 Accounts of narrativity	118
5.2.1 Intelligibility and meaning in narratives	119
5.2.2 Narratives constituting the self	122
5.2.3 Summary of the argument for narrativity	125
5.3 Strawson's objections to narrativity theory	128
5.3.1 The psychological narrativity thesis is not true of everybody	129
5.3.2 If the psychological narrativity thesis is true at all, it is a trivial thesis	132
5.4 The effects of a psychiatric diagnosis	133
5.4.1 What kinds of effects may be produced in the recipient?	133
5.4.2 How can psychiatric diagnoses generate such effects?	138
5.5 Conclusion	141
Chapter 6: Hermeneutical injustice as a consequence of a psychiatric diagnosis	
6.1 Introduction	143
6.2 Epistemic injustice	145
6.2.1 Hermeneutical injustice	146
6.3 Epistemic injustice in medicine and psychiatry	150
6.3.1 Epistemic injustice in healthcare contexts	150
6.3.2 Epistemic injustice in psychiatry	153
6.4 Effects of a diagnosis on intellectual self-trust	156
6.5 Evidence of hermeneutical in justice in testimonies of service-users and	
former service-users	159
6.6 Possible objection: can a diagnosis enhance the patient's hermeneutical resources?	166
6.7 Conclusion	169
Chapter 7: Conclusion	171
References	174

### Acknowledgements

I would firstly like to thank my supervisors, Luca Barlassina, Megan Blomfield, and Steve Laurence for their invaluable help, critiques and guidance in writing this thesis.

I would additionally like to thank Steve Laurence for encouraging me to apply to do a PhD at Sheffield following the completion of my MA.

Many other people, both within the University of Sheffield Department of Philosophy and elsewhere, have helped me in various ways during the years I have been writing this. The Department has been a very stimulating intellectual environment, in which I am enormously grateful to have had the opportunity to study. The postgraduate seminars that I have attended have helped me think about how to present my own ideas in a cogent manner. I am also grateful for the opportunity my time in the Philosophy Department has given me to educate myself a little more about other areas of philosophy unconnected to the subject of my thesis.

I would also like to acknowledge the stimulation and helpful feedback I have gained from attending and presenting some of my work at several of the philosophy of psychiatry work-in-progress meetings organised annually by Rachel Cooper and her colleagues at Lancaster University.

I am indebted to conversations I have had with Philip Rau who drew my attention to the work of Marya Schechtman and Şerife Tekin. I am also grateful to Philip for warning me about making the unwarranted assumption that narratives somehow constitute the "self".

I am grateful to Lucy Johnstone for her comments on an earlier draft of Chapter 2. I also wish to acknowledge a comment she made about evidence that some former service-users trace the start of their recovery from the moment they rejected their diagnosis, in her presentation at the David Smail Memorial Conference in 2015, which started me thinking about the issue of hermeneutical injustice arising following a diagnosis.

I am also grateful to many other people, too numerous to list, with whom I have had conversations over many years, including during my career as a clinical psychologist, that have contributed to my thinking on the subject of this thesis.

Finally, I am deeply grateful and indebted to my wife, Margy, for her continuing support and patience, when I have been less available for domestic and family matters than I should have been during the years it has taken me to prepare this thesis.

# <u>Psychiatric Diagnosis and Hermeneutical Injustice: The Impact of</u> <u>Biomedical Diagnoses on Personal Narratives</u>

#### **Abstract**

Users of mental health services are at risk of becoming victims of epistemic injustice, as described by Miranda Fricker. In this thesis, I claim that they can be victims specifically of hermeneutical injustice, which can occur due to the diagnosis they receive. A psychiatric diagnosis is often taken to represent some kind of discrete disease, as if it connotes a natural kind of disease entity. I argue that many physical diseases can be understood as natural kinds in medical science. However, most psychiatric diagnostic categories cannot be so understood. They do not offer any explanation for the patient's condition. Much of modern psychiatry is based on the biomedical model of diseases. This model seeks the causes of patients' illnesses in biological abnormalities in the body. Accordingly, the receipt of a psychiatric diagnosis can convey a biomedical narrative about the nature of the patient's condition, one that tends to locate the cause of the condition in abnormal brain processes of some kind. The effect of this can be understood in terms of the self-narratives that individuals construct for themselves. I discuss several accounts of narrativity which explain how individuals gain meaning in their lives through their self-narratives. These narratives can be changed by the person's social circumstances and by extraneous events. Receiving a psychiatric diagnosis is one such event in some people's lives which can change the recipient's self-narrative about their life and their difficulties. The medicalization implicit in psychiatric diagnoses conveys a biomedical narrative which may conflict with or diminish the recipient's previous selfnarratives at a time when they will be experiencing significant emotional distress and disturbance. As such, the recipient's own hermeneutical resources for making sense of their experiences can become marginalised. This can result in the recipients of psychiatric diagnoses becoming victims of hermeneutical injustice.

#### Introduction

I claim in this thesis that a psychiatric diagnosis, based on the DSM diagnostic system (American Psychiatric Association, 2013), can lead to the recipient of the diagnosis becoming a victim of epistemic injustice, and specifically of hermeneutical injustice. The claim being made is not that this is an inevitable result of the diagnosis, but rather that the recipient is vulnerable to such a consequence occurring. This can come about because the biomedical narrative associated with a psychiatric diagnosis can cause the recipient to doubt his own prior self-narrative and believe that his condition is in some way caused by an abnormality in his brain. A psychiatric diagnosis is a very powerful statement which influences how recipients see themselves. Much of its power derives from the epistemic authority of the medical profession and the status it has in the institutions of medicine and healthcare. Moreover, the diagnostic process itself, which is aimed at deciding upon a diagnostic category for the patient can lead to the latter being hermeneutically marginalised.<sup>1</sup>

Unlike diagnoses of physical illnesses which typically have some explanatory value, psychiatric diagnoses do not generally provide any explanation for why the patient's mental condition is what it is. The DSM-based diagnostic categories (with a few exceptions) are not individuated by means of their aetiologies, which in most cases are highly heterogeneous. Nevertheless, people suffering from mental disorders and under the care of mental health services are frequently told that they have "an illness like any other" – in other words, that their condition is analogous to a physical disease. This raises the question of whether the conditions indicated by the psychiatric diagnostic categories really are analogous to physical diseases. One way of posing this question is to ask whether these categories represent natural kinds in medical or psychiatric science, and whether they differ from diagnoses of physical diseases in that respect. I therefore address this question in Chapters 1 and 2.

In Chapter 1, I start by discussing different conceptions of natural kinds that have been proposed, focussing in particular on essentialist and property cluster kinds. I argue that property cluster kinds, in which the properties that constitute a natural kind are causally associated with each other, as in the accounts described by Richard Boyd and Muhammed Ali Khalidi, can have broad applicability across the sciences. According to these accounts, natural kinds reflect the causal structure of the world and, as such, they can ground

-

<sup>&</sup>lt;sup>1</sup> In this thesis, I also use the term 'service-user' on occasions. I use the terms 'patient' and 'service-user' interchangeably.

explanations and inductions in science. On the basis of this conceptualisation, I argue that many diseases in somatic medicine can be seen as natural kinds due to the explanatory function they have in medical science.

I discuss psychiatric diagnostic categories more directly in Chapter 2. I explain the development of the DSM diagnostic system and the approaches to the validation of the categories that have been attempted. However, there is now a broad consensus in psychiatry that most of the DSM categories have not been shown to be valid. In particular, they have not been shown to have construct or predictive validity. The DSM-based diagnostic categories do not in general explain the nature of the patient's condition – no causal basis in abnormal neurological processes for these conditions has so far been discovered – and I argue therefore that they do not represent natural kinds. I illustrate this by discussing the example of schizophrenia which, I argue, lacks significant explanatory value. Consequently, there is reason to doubt whether these categories are analogous to disease entities in somatic medicine.

I discuss the biomedical model of mental disorders and the nature of diseases as conceived in biomedicine in Chapter 3. According to this model, a reductionist approach is taken towards explaining diseases, which are typically understood as being caused by abnormalities in specific physiological or biochemical processes. This model is also prevalent in modern psychiatric science and practice, in which an enormous amount of research has been directed towards uncovering neurological pathways underlying mental disorders. It influences the way in which patients are encouraged to view their condition. They are often told when they receive such a diagnosis that they have an illness like any other. I argue that this model reinforces the notion that the patient's condition can be understood as in some way analogous to a biological disease, and specifically one that affects the brain.

In Chapter 4, I argue that diagnosis has a pivotal place in biomedicine and that the statement of a diagnosis to the patient is a perlocutionary speech act in virtue of the consequences it leads to. Referring to John Searle's theory of institutional facts, I also argue that a diagnosis can constitute an institutional fact. This carries implications for the patient's treatment and, in some circumstances, for possible restrictions on her freedom. As such, it helps to explain the power that a diagnosis can have on the individual receiving it, particularly in the case of psychiatric diagnoses which can have a stigmatising effect on the recipient and, in some cases, can provide grounds for her compulsory detention in hospital.

I discuss the narrative consequences of a diagnosis in Chapter 5. I argue that personal narratives are important for self-understanding. I outline some key themes in narrative theory

which have been explored in various ways by several philosophers and to a lesser extent by psychologists. Among other consequences, a psychiatric diagnosis can convey a biomedical narrative about the patient's condition. This may conflict with or supersede her previous self-narrative. I also argue that receiving a psychiatric diagnosis may affect the patient's sense of agency and hopes for recovery. An alternative approach, which can avoid such effects, involves constructing a narrative-based psychological formulation collaboratively with her, such that she can make sense of her condition in other than medicalised terms.

In Chapter 6, I argue that those receiving a psychiatric diagnosis may be vulnerable to experiences of epistemic injustice, as described by Miranda Fricker. I focus particularly on hermeneutical injustice, where individuals lack the ability to understand their experiences or difficulties in ways that make sense to them. I argue that the patient can become hermeneutically marginalised by the diagnostic assessment which privileges a biomedical narrative over the patient's own self-narrative. The medicalisation implicit in psychiatric diagnoses conveys a particular kind of narrative which may conflict with the recipient's previous self-narratives. I quote several testimonies of patients and former service-users which support the view that the diagnosis has had a harmful effect on them. Many of these can plausibly be seen as evidence of hermeneutical injustice. This is particularly likely to be the case when patients are led to believe that their condition is a chronic one. The belief, usually unfounded, that the condition is chronic, and one from which recovery is unlikely, may induce corresponding feelings of hopelessness, which may thereby limit the prospect of a positive outcome for the person concerned. A greater use of psychological case formulation could help to prevent patients being misled by such beliefs.

In Chapter 7, I make some brief concluding remarks about how such injustices might be ameliorated or avoided. Although, as Fricker states, agents do not create hermeneutical injustices on their own, there still remain actions that clinicians in mental health services can undertake to lessen their effect and avoid perpetuating a culture in which such injustices are liable to occur. When a patient is given a diagnosis, the limitations of what the diagnosis means could be explained clearly to the patient to avoid her gaining a misconception about its implications. The time pressures on busy clinicians in under-resourced services need not prevent them taking care to limit the perpetuation of such hermeneutical injustices.

#### Chapter 1

#### The natural kind status of diseases

#### 1.1 Introduction

In this chapter, I argue that diseases can be regarded as natural kinds in medical science. While this might seem a plausible claim, given the regularity with which we talk about common diseases and the apparent predictability of their courses, there is sufficient variation in how disease types are manifested in individual patients to suggest that this is not entirely straightforward. I therefore devote much of this chapter to examining different conceptions of natural kinds, in order to show how this kind of variability can be accommodated within a theory of natural kinds that is applicable across the sciences in general.

In section 1.2, I begin by giving a broad overview of natural kinds and their epistemological function in science, before discussing specific conceptions of them. In section 1.2.1, I discuss the essentialist conception of natural kinds. Since difficulties arise in applying this to the special sciences such as biology, alternative conceptions involving kinds as property clusters have been proposed. I discuss these in section 1.2.2, where I focus on the property cluster accounts of Richard Boyd and Muhammad Ali Khalidi in turn. These accounts emphasise that the clusters of properties constituting natural kinds are characterised by causal relationships among them. I then, in section 1.2.3, discuss two further property cluster accounts, proposed by Anjan Chakravartty and Matthew Slater respectively. These omit the requirement for a causal relationship between the properties comprising the kind, arguing that such a feature is unnecessary. In section 1.2.4, I discuss some objections to the latter view and argue that it's implausible that natural kinds could be explanatorily successful across the sciences without causal relationships of some sort among the property clusters.

In section 1.3, I argue that diseases, or more accurately disease entities, can plausibly be regarded as natural kinds and that a causal process account best explains how disease kinds can ground explanation and induction. I give some examples of diseases that can plausibly seen as natural kinds in medical science. I summarise these claims in section 1.4.

#### 1.2 Conceptions of natural kinds

At a rough first pass, natural kinds can be thought of as the naturally occurring groups of entities which have important properties in common and which underpin our scientific

predictions – what Quine describes as "the functionally relevant groupings in nature... [which] make our inductions tend to come out right" (1969, p126). As I have indicated above, there are various conceptions of natural kinds in the philosophical literature. This immediately presents problems for any attempt to set out a theory of natural kinds, since, as Ian Hacking (2007) has argued, no a priori specification of this can be definitively established. Hacking explores the history of theorising about natural kinds from the 19th century onwards and identifies an increasing proliferation of conflicting accounts. He concludes: "A stipulative definition, that picks out some precise or fuzzy class and defines it as the class of natural kinds, serves no purpose, given that there are so many competing visions of what the natural kinds are" (2007, p239). Consequently, he concludes that the philosophical study of natural kinds can have no further value. In contrast, Miles MacLeod and Thomas Reydon (2013) propose the more optimistic view that natural kinds can still do important work. They take Hacking to be right in his critique of the disunited and conflicting philosophical approaches to defining natural kinds, but overly pessimistic in concluding that no possible account of this notion can have any use in philosophy of science. In particular, MacLeod and Reydon see a continuing agenda for the study of natural kinds in science (and, particularly from their perspective, in the life sciences) to the extent that they can have explanatory and inductive value.

The proliferation of accounts described by Hacking may have arisen from a conflation of two distinct questions: what P D Magnus (2018) identifies as the "taxonomy question" and the "ontology question". According to Magnus the taxonomy question concerns the distinction between a natural kind and an arbitrary category, and the ontology question asks what kind of thing it is that characterises a natural kind. The ontology question typically requires an answer in metaphysical terms – for example Katherine Hawley and Alexander Bird (2011) characterise natural kinds as mereologically complex universals. However, it is not clear that an explanation of the epistemological function of natural kinds requires any strong metaphysical realist account. Ingo Brigandt (2009), for example, argues that a metaphysical account of natural kinds is unnecessary for understanding the epistemic role of kind concepts in scientific reasoning. The task then becomes one of providing an account of how natural kinds in science can fulfil their epistemic function. In this respect, a successful account can provide an answer to Magnus's taxonomy question. My aim here is to address this question. Specifically, I aim to examine how we might distinguish between arbitrary and non-arbitrary categories, and how non-arbitrary categories in science may best be characterised.

Broadly speaking, there is a general consensus that natural kinds in science are expected to ground successful inductions and predictions about the subject matter in the relevant scientific domain. For this to be the case, the properties characterising natural kinds need to be such as to make the kind "projectible". This is the requirement that, given the presence of a number of salient properties in a member of a natural kind, we should be able to predict the existence of such properties in other members – i.e. to extrapolate from examined instances to unexamined ones. Thus, discussing natural kinds of disease in medical science, Neil Williams (2011) illustrates projectibility as follows:

We take treatments to be repeatable, and information gathered from one instance of a disease to be relevant to further instances of that disease, because we take similarity of disease instances within disease types to be a naturally occurring feature of our world. In short, we treat medical information as *projectible*, and we do so on the grounds that disease kinds are natural kinds (p204, italics in original).

So for example, knowing that insulin treatment has worked to save the lives of innumerable patients with Type 1 diabetes, we can be confident that, for the next patient with this diagnosis whom we encounter, their condition can also be managed successfully in the same way. Consequently, whatever account of natural kinds is judged to be most applicable within a given scientific context, a minimum expectation is that the kind should be projectible – i.e. that it should be capable of supporting inductions and predictions. This is a widely agreed upon feature of natural kinds and may represent the primary motivation for articulating any account of them. In the following section, I discuss different accounts and how they set out to explain the projectibility of defining properties of natural kinds.

#### 1.2.1 Essentialist accounts

According to the essentialist conception, objects or entities belong to a natural kind if they share a property (or conjunction of properties) which is essential to membership of that kind. Possession of such a property is a necessary and sufficient condition for membership of the kind which thus constitutes its "essence". An influential account of essences was given by Locke who says: "essence may be taken for the very being of anything, whereby it is, what it is" (1706/1997, Book 3, Ch. 3, Section 15, pp373-374, italics in original). He considers that essences can be conceived in two ways – "nominal" and "real". The nominal essence of a substance contains the perceptible features which we take to be its typical properties, whereas its real essence is the actual constitution of the substance. Hence the nominal essence of gold, using Locke's example, is given by its distinctive colour, weight, malleability—the qualities

by which we know it. By contrast, the real essence of gold, conceived of in microstructural terms, is its invisible composition. This is reflected in the essentialist conception of natural kinds according to which the essence of gold is given by its atomic structure and number.

More recently, essentialist accounts of natural kinds have generally tended to follow from the arguments about the fixing of reference developed by Hilary Putnam (1975) and Saul Kripke (1980). Kripke introduces the term "rigid designator" which applies to any term which designates the same entity in all possible worlds. Proper names are typically rigid designators, since names such as "Napoleon Bonaparte" refer to the same individual in any possible world. Similarly, terms for natural kinds are also rigid designators. This is because they refer to kinds which have an essence which is the same in all possible worlds – the reference of the kind term is fixed by its essence. Kripke illustrates this by explaining how we fix the reference of words like 'gold'. We now know that gold has an atomic number of 79, which fixes its reference, and we determine that an object is a piece of gold by virtue of this feature, its known essence. If some object looks like gold (e.g. a piece of fool's gold), but turns out to have a different molecular structure, we conclude that it is actually not gold, rather than changing the extension of "gold" to accommodate it. We also know that if a piece of gold turned out not to be yellow (perhaps because of some optical illusion), but was in fact blue in its normal state, we would not cease to designate it as 'gold', because its identity as gold is determined by its essence, its microstructure. A similar process applies, Kripke argues, to the fixing of reference for biological kind terms. We normally expect tigers to have four legs, but if we should happen to discover that they only have three legs (again because we have been deceived by a strange optical illusion) we would still conclude that they were tigers, not a different species entirely. This would be so, even without our having adequate knowledge of the underlying microstructure (the essence) of the kind.

A broadly similar type of argument is advanced by Putnam, which I sketch very briefly here. He describes a thought-experiment in which we are asked to imagine a planet identical to ours ("Twin-Earth") in all respects, except that the water-like substance there – the one that is colourless, odourless, safe to drink, etc. – has the chemical composition XYZ, rather than H<sub>2</sub>O as we know it on Earth. Putnam argues that we cannot call this watery substance 'water', even if we would have done so prior to the discovery of the atomic structure of water, because what now determines the reference of 'water' are facts about water as revealed by scientists. That is to say that scientists have now identified its microstructure, such that anything that is water must have the chemical composition H<sub>2</sub>O. This molecular structure,

according to Putnam, is the *essence* of water and anything with a different microstructure must be some other substance.

A common feature of essentialist accounts is the view that the essential properties of natural kinds determine how members of these kinds will behave in response to specified circumstances. Thus according to these accounts, the behaviour of the chemical elements, for example the way they may combine with other elements, is determined by their atomic structure in combination with relevant background conditions. It is the particular structures of hydrogen and oxygen atoms which determine the way in which they combine to form water molecules and which also influences the behaviour of the ensuing substance in various circumstances. The defining essence of a kind can also be a *conjunction* of properties, as is the case for example with water which has a conjunctive essence of its respective hydrogen and oxygen atoms and its specific atomic structure. It is the essence of a natural kind, on this account, which underpins its projectibility – i.e. which grounds the inductions that the kind makes possible.

The essentialist conception is often believed to give a good account of natural kinds in chemistry (although I discuss below Khalidi's account which questions this view). However, it does not seem to provide a satisfactory account of natural kinds in biology and other special sciences. In particular, biological species cannot be understood as being characterised by a single defining essence in the same way as chemical elements. Species adapt and change over time in accordance with evolutionary theory, with new species emerging in response to changing environmental pressures. Wilson, Barker and Brigandt (2007) have argued that biological species kinds are intrinsically heterogeneous in that there is substantial and intrinsic variation amongst individuals within the kind. Thus for example, there is no property which all alligators share and which is not present in other species. The typical dark brown colour on most individuals is not universally present, since there are albino alligators which are white. No other property, such as strong jaws and sharp teeth, is exclusive to the species and defining of it. Moreover this kind of variation forms the basis on which natural selection acts during the process of evolutionary change. Consequently, there is no single defining essence which can play an explanatory role in accounting for the various traits displayed by individual members of the kind. Alternative accounts have therefore been formulated based on the notion of natural kinds as representing clusters of salient properties.

#### 1.2.2 Property cluster accounts incorporating causal processes

I argue in this section that an account of natural kinds as property clusters linked together by causal processes has the particular advantage of explaining how natural kinds function to ground inductions in all the sciences, including the special sciences. In this section, I first review Boyd's account and then go on to discuss Khalidi's alternative account. In section 1.2.3, I review alternative accounts which de-emphasise the role of causal processes.

#### i) Boyd's account

A key source of motivation for developing property cluster theories has been the perceived need to account for natural kinds of biological species. One of the most prominent theories put forward with the aim of doing this is Richard Boyd's (1991, 1999a, 1999b, 2021) account of "homeostatic property cluster" (HPC) kinds. To give a brief summary of this account, Boyd describes natural kinds as characterised by clusters of properties held together by "homeostatic" causal mechanisms. In contrast to the essentialist account, Boyd regards his account as a broadly nominalist one in which the cluster of properties forming the kind constitute something like Locke's nominal essences. These gain their potential for grounding successful inductions in virtue of their ability to reflect the causal structure of relevant phenomena as studied in specific scientific disciplines. I now clarify the key aspects of this account in more detail.

First, the notion of property clusters captures the idea that no one property, or conjunction of properties, can be seen as a necessary and sufficient criterion for kind membership, as in the case of essentialist kinds. Instead, members of the kind will display a significant, though unspecified, number of properties from the cluster through which the kind is identified. The properties defining the kind are correlated, but the correlation is not perfect. This conceptualisation reflects the observation that biological species kinds are not characterised by a single essence which is present in all members of the kinds and absent from non-members, as the example of alligators mentioned above demonstrates.

Second, the co-occurrence of the properties in the cluster is brought about by homeostatic causal mechanisms which account for their co-occurrence. Boyd illustrates this specifically with reference to biological species kinds in which the causal mechanisms maintaining the property cluster take various forms, both intrinsic and historical. These include transfer of genetic material through reproduction, evolutionary history and adaptation in response to environmental pressures. These various types of causal mechanism interact in such a way as to reinforce each other and maintain the species as a sufficiently stable natural

kind to support explanations and induction. However, the stability of such kinds does not entail that their boundaries and extensions are fully determinate – there is some inevitable indeterminacy involved. Boyd explains this as follows:

The *necessary* indeterminacy in extension of species terms is a consequence of evolutionary theory, as Darwin observed: speciation depends on the existence of populations which are intermediate between the parent species and the emerging one. Any 'refinement' of classification which artificially eliminated the resulting indeterminacy in classification would obscure the central fact about heritable variations in phenotype upon which biological evolution depends and would be scientifically inappropriate and misleading (1991, p142, italics in original).

Although Boyd does not say much how such homeostatic causal processes may function to maintain stable kinds in other scientific disciplines, he suggests that the interactions of relevant causal forces can be conceptualised in a similar manner for natural kinds in other sciences.

Third, Boyd emphasises strongly that natural kinds function in such a way as to enable our conceptual practices to accommodate to the causal structure of the world. This element of his account is given particular prominence in his writings on the subject, especially in a later paper. He says:

The naturalness of natural kinds consists in their aptness for induction and explanation.... The thesis I defend here (the *accommodation* thesis) makes the further claim that what is at issue in establishing the reliability of inductive and explanatory practices, and what the representation of phenomena in terms of natural kinds makes possible, is the accommodation of inferential practices to relevant causal structures (1999a, p147, italics in original).

The naturalness of a natural kind is correspondingly reflected in the projectible generalisations to which it can give rise. A putative category which fits past data well may nevertheless fail to generate a sufficient number of projectible generalisations to future cases to meet the explanatory demands of the particular discipline. Where generalisations prove to be projectible on a sufficiently regular basis (as expected for the discipline concerned), the categories grounding these represent the accommodation of our conceptual resources to the relevant causal structures. Boyd says again: "Natural kinds are solutions to problems of disciplinary accommodation: to problems about how to sort things so as to facilitate reliable induction and explanation" (1999b, p72). In other words, the categories and taxonomies we

create are those which best enable us to explain important phenomena and guide us in making reliable predictions based on hypotheses about causal relationships.

However, achieving reliable induction does not guarantee that such generalisations will invariably be borne out by experience in sciences where exceptionless laws are lacking. Thus, Boyd also stresses that the demand for accommodation of conceptual practices to causal structures is particularly pressing in those inexact sciences, such as geology, meteorology, and biology. A key feature of such sciences is that they are characterised by phenomena which are frequently influenced by very large numbers of variables, and which are too numerous to be fully accounted for in the generalisations made about them.

Fourth, HPC natural kinds can only be identified *a posteriori*. It is only by studying nature and checking our inductions against natural phenomena that we can discover whether and how our putative kinds accommodate to the causal structure of the world. In such circumstances, "natural kinds reflect a strategy of deferring to nature in the making of projectability judgments: we define such kinds *a posteriori* in ways which reflect actual causal structure precisely because we are unable to identify or specify projectable generalizations without doing so" (Boyd, 1991, p139). Thus, categories defined in some way *a priori*, which in some cases may reflect purely conventional categorisations of social phenomena, cannot be assumed to be natural kinds.

Fifth, Boyd emphasises that natural kinds are specific to the "disciplinary matrix" in which they are found to generate projectible predicates. In some cases, he envisages that disciplinary matrices may correspond with traditional scientific disciplines, but this is not in general the case. He explains this as follows: "By a *disciplinary matrix* I'll understand a family of inductive and inferential practices united by common conceptual resources, whether or not these correspond to academic or practical disciplines otherwise understood" (1999a, p148, italics in original). This suggests that traditional disciplinary boundaries may to some extent be arbitrarily defined. However, although the functions of natural kinds in grounding induction and explanation may be limited to specified contexts, these do not need to correspond precisely to specific scientific disciplines. Thus for example: "Acids form a natural kind for chemistry, but also for geology, mineralogy, metallurgy, and so on" (op cit., p148). For this to be the case, there must be common inferential practices and conceptual resources operative across these various disciplines.

Boyd's account of HPC natural kinds has been endorsed by many philosophers and used as a basis for understanding how natural kinds in biology can be construed (e.g. Griffiths, 1999; Kornblith, 1993; Rieppel, 2005; Wilson et al., 2007). There have been

various criticisms of this account, centred particularly on questions about the nature of the supposed homeostatic causal mechanisms and whether there need be any such causal mechanisms at all. I review these criticisms in section 1.2.3. Before I do this, however, I first discuss Khalidi's account.

#### ii) Khalidi's account

Muhammad Ali Khalidi (2013, 2018) proposes an alternative account of natural kinds based on the notion of clusters of properties linked by causal processes. He describes natural kinds as "nodes in causal networks" (2018). He argues that this is a unifying account with general applicability across all the sciences. Before discussing this in detail, however, I will describe one criticism of Boyd's account made by Carl Craver (2009), because this is relevant to Khalidi's account and partially motivates it.

Boyd's account relies on the notion of homeostatic causal mechanisms as the means by which natural kinds are individuated. However, Craver argues that this concept fails to provide a fully objective account of natural kinds and instead it unavoidably involves some conventional elements in determining which mechanisms are relevant for kind individuation. Specifically, Craver argues that there is no objective way of determining the boundaries of mechanisms which can suit all epistemic purposes. Instead, what count as the spatial and causal boundaries of mechanisms is dependent on what phenomena we are particularly interested in studying – in other words, on what our choice of property cluster might be. In the case of complex disease processes, for example, there may be no objective way of specifying this. Consequently, he argues, we cannot explain the property cluster by reference to some objective set of mechanisms. What Craver suggests instead is that we can reject the notion of determining mechanisms, but retain a theory of property cluster kinds accounted for by a "simple causal theory" (2009, p579). It is not entirely clear, however, how natural kinds in such a theory as this can reflect the causal structure of the world in the way that Boyd expects.

Khalidi (2013, 2018) endorses Craver's critique of Boyd's notion of causal mechanisms.<sup>2</sup> He also argues that the emphasis on *homeostatic* mechanisms in Boyd's

<sup>&</sup>lt;sup>2</sup> However, it is not clear that Khalidi has accurately reflected Boyd's position on this. Boyd repeatedly refers to causal "mechanisms" (in the plural) as underlying natural kinds, whereas Khalidi, in both his 2013 book and his 2018 paper, reads Boyd as assuming that each HPC kind is underpinned by a homeostatic causal *mechanism* (in the singular). As such, this does not seem to do full justice to Boyd's account. Craver does not seem to fall into this misreading.

account seems very specific to biological species where feedback loops and other counterbalancing influences are in evident in species kinds. Consequently, he says, it appears to be too restrictive in its application, being primarily applicable to natural kinds in special sciences such as biology, geology, and psychology. Nevertheless, he concedes significant similarities between his own and Boyd's account. "The account I am proposing is very similar to Boyd's when one drops the mechanism and the homeostasis" (Khalidi, 2018, p1386, n7).<sup>3</sup>

Khalidi elaborates Craver's notion of a simple causal theory which he develops as a unifying account of natural kinds. He describes natural kinds as sets of properties linked together by more general causal processes which can ground inductions about various derivative properties. This account can be applied to natural kinds in all the sciences, including physics and chemistry. Even in the latter sciences, Khalidi argues, essentialism does not provide a fully adequate account of natural kinds. Essence kinds, as traditionally conceived, are not a distinct kind of natural kind. Rather, those kinds that might be considered as essence kinds are simply positioned at one end of a continuum of property cluster kinds, which vary only in terms of the type of properties comprising the defining cluster and the complexity of the causal relations linking them. Khalidi argues that chemical elements, as well as kinds in all other sciences, are defined as natural kinds by virtue of a cluster of properties which have different causal effects depending upon the particular properties represented in a given sample.

His argument about this is complex and open to possible criticism, and I will therefore examine this in more detail before discussing his conception of how kinds feature in causal explanations more generally. His account of chemical kinds is intended to accommodate the fact that chemical elements typically have isotopes which do not necessarily share many properties with each other. For a given element, the isotopes will all have the same atomic number but different mass numbers. In order to account for this he proposes that such a kind is characterised as a conjunction of one necessary property (the atomic number) and a disjunctive necessary condition comprising one or other of the element's isotopes. It therefore takes the form:  $P_1 \wedge (P_2 \vee P_3 \vee ... P_n)$ . Using Khalidi's example of lithium,  $P_1$  would represent the atomic number of lithium (i.e. 3),  $P_2$  represents the mass number 4,  $P_3$  the mass

<sup>&</sup>lt;sup>3</sup> Khalidi distinguishes between causal "mechanisms" and "processes". However, he does not make clear what is the nature of this distinction, nor why it is important, and it seems to amount just to a question of how the word 'mechanism' should be interpreted. As indicated in the quote above, Khalidi regards his account as similar to Boyd's.

number 5, and so on. On this basis, therefore lithium is a natural kind characterised by a cluster of properties in this manner<sup>4</sup>. Having dismissed other options for kindhood in chemical elements, Khalidi nevertheless concedes the "apparent messiness" of this conception, but claims it is "most consistent with philosophical principles and does the least violence to scientific practice" (2013, p169). This suggests that there is something unsatisfactory about this conception, and the use of the phrase "the least violence to scientific practice" perhaps implies that at least some "violence to scientific practice" might still be discernible. Moreover, this notion of a chemical natural kind still depends upon a *necessary* property for kind membership, even if a sufficient one cannot be specified.

Another possible criticism of Khalidi's account of the nature of chemical element kinds as a property cluster of the kind he describes is that the properties comprising the cluster do not seem to be causally linked, contrary to what his account requires. On his account, the conjunction of the element's atomic number and a disjunctive condition comprising one or other of the mass numbers of its isotopes constitute the defining property cluster. However, there is no obvious causal association between the atomic number of an element and the mass number of any particular isotope of a given substance. Khalidi might reply by saying that for any given element there is a constraint on the number of isotopes that are possible. Lithium, for example, has only two stable isotopes and a small number of highly radioactive ones. Nevertheless, this does not obviously amount to a causal relationship between the property of its atomic number and the respective mass number of each isotope. This seems to be a problem for any account in which property clusters are held together by causal processes of some sort.

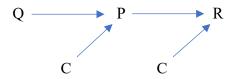
An alternative conception, which would avoid these difficulties, but which Khalidi does not appear to consider, would be to adopt a position closer to traditional essentialism. This would retain the notion of the atomic number as the element's single defining property (the necessary and sufficient condition for kind membership), while recognising the limitations on the other properties that can be directly attributed to this. The different possible atomic masses of the same element would count as background conditions which affect the manner in which other properties, direct or indirect, of the element might be instantiated. Thus, the properties of water cannot be derived directly from the property of the atomic number of hydrogen (along of course with that of oxygen), since different isotopes of hydrogen can

<sup>&</sup>lt;sup>4</sup> Khalidi also argues that there can be many natural kinds of *isobar*, kinds of substance in which the necessary condition is a given mass number and a disjunctive condition representing variable atomic numbers – e.g. there is an isobar kind which contains lithium-8, beryllium-8 and so on.

combine with oxygen to form different types of water. Water formed from the isotope deuterium ("heavy water") is not safe for drinking in the same way as normal H<sub>2</sub>O. In this case, the mass number of deuterium changes some of the resultant properties of the hydrogen atom. Nevertheless, one can still conceive of the kind hydrogen as having an atomic number of 1 for its defining property (its essence), which is sufficient to make some entity be hydrogen if it has this property. Whilst on this revised account a chemical natural kind is no longer defined by a multiplicity of clustered properties, it can still be seen as situated on a continuum of kinds that are property clusters, but lying at the limiting point where it is defined just by a *single* property – a cluster of one. One can plausibly argue that this conception is less "messy" than Khalidi's and fits well with normal scientific practice. It also seems fully consistent with his account of kinds as nodes in causal networks.

In Khalidi's account, natural kinds are embedded in causal networks and the clusters of properties in these kinds are causally linked with other effects in the network. The distinctive clustering of properties that lead to their various effects are what Khalidi terms the "nodes" in the network. The manner in which the properties forming the kind generate their effects is mediated by the effects of background factors and other *ceteris paribus* conditions. This is represented by Khalidi in a "directed causal graph" (2018, p1387) in Figure 1.

Figure 1:



In this schematic figure, core properties of natural kinds ( $Q = Q_1, \ldots, Q_n$ ) cause derivative properties, whether directly ( $P = P_1, \ldots, P_m$ ) or indirectly ( $R = R_1, \ldots, R_k$ ), in conjunction with background conditions, other properties, or *ceteris paribus* conditions ( $C = C_1, \ldots, C_j$ ). This represents a simple linear causal network, but special science kinds will typically be embedded in more complex networks involving many background conditions and also in some cases recurring feedback loops. The resulting network will be a complex web of interacting causal relationships, with natural kinds located at points from which several causal processes converge and diverge. It is this pivotal location in causal networks which supports the inductive value of natural kinds. As Khalidi says: "What enables natural kind

categories to play the role that they do in our inductive, explanatory, and taxonomic practices is that they consist of highly connected nodes in causal networks" (2018, p1387).

What Khalidi shares with Boyd is the view that the properties constituting a natural kind must be causally linked in some way, albeit not necessarily by a homeostatic causal mechanism as described by Boyd. It is not sufficient for these properties to be merely correlated. Khalidi emphasises this: "Mere correlation of properties is not enough, since we are ultimately interested in causation" (2018, p1383). It is the causal structure linking the properties constituting natural kinds which accounts for their ability to feature in causal networks and to support explanations and inductions. Without there being such a causal structure underlying a natural kind, it is not clear how the kind could be individuated as playing a role, *qua* kind, in a causal network, as opposed to some of its individual properties having separate causal functions. There are, however, accounts of natural kinds in which the properties constituting kinds can form stable correlations without necessarily being causally associated. I review some of these below and argue in more detail that stable correlations of properties are insufficient by themselves to constitute the kind of natural kinds that would normally be expected to feature in scientific explanations.

A virtue of Khalidi's account is that it is generally applicable across all the sciences and seems to reflect the practice of scientists in different disciplines. Khalidi provides a few examples of how it can be applied in some of the special sciences. One such example from the field of microbiology is the category of *virus*. To simplify somewhat, the primary properties, those which constitute the natural kind *virus*, consist of a strand of genetic material (DNA or RNA) in a protein casing and a genome which is able to make messenger RNA (mRNA). The secondary properties (the derivative effects) in the ensuing causal network constitute the infectious cycle. These include the attachment of the virus to a host cell in an organism, the translation of the viral mRNA by host ribosomes, the replication of the viral genome, the release of the viral particle to infect other cells, and the consequent repeating process in the host organism. Natural kinds in other special sciences, Khalidi argues, can be understood as functioning in causal networks in broadly similar ways.

-

<sup>&</sup>lt;sup>5</sup> There are various ways in which the constituting properties might be causally linked. For two properties, X and Y, there may be a direct causal link between them – i.e. X causes Y or vice versa. Alternatively, there need be no such direct causal association, but instead both properties might be caused by a third property Z, whilst X and Y remain as constituting properties of the kind. Whether or not Z is part of the defining properties will depend on the particular scientific field and how kinds are identified and understood within that discipline.

#### 1.2.3 Non-causal property cluster accounts

The accounts of kinds to be considered in this section agree with those of Boyd and Khalidi that natural kinds are constituted by clusters of properties which have causal implications in scientific explanations, but they differ in holding that the properties in a kind need not be causally linked with each other (though of course they may be). One such account is given by John Dupré (1993) who holds that natural kinds are groups of properties which cluster together in regular ways. What on this account primarily constitutes a natural kind is a reliable correlation of properties which enables predictions to be made about other properties and behaviour of kind members. His account is particularly geared towards philosophy of biology and the need to explain how the different types of taxonomy used in biology can be equally valid. Species concepts and the resulting taxonomies of categories can be organised according to phenetic, interbreeding, ecological, and phylogenetic theoretical frameworks, depending upon the particular explanatory requirements at issue. These are in effect valid, but cross-cutting, taxonomies with different explanatory aims, and on Dupré's account none is in any way ontologically prior to the others<sup>6</sup>. This type of account in which different crosscutting taxonomies can be equally valid or "real" reflects a stance which Dupré calls "promiscuous realism".

More recent accounts of non-causal property cluster accounts have been proposed by Anjan Chakravartty (2007) and Matthew Slater (2015). I will focus on these, rather than Dupré's, as they explicitly address Boyd's account of HPC kinds and offer what they see as an alternative to it. Whilst they do not explicitly address Khalidi's causal account, their conception of natural kinds can also be seen as a challenge to the latter.

Chakravartty places his conception of natural kinds within his broader account of scientific realism, which he labels variously as "semi-realism" (2007) and "dispositional realism" (2013). These terms seem to be effectively equivalent, though the latter seems to reflect his account more clearly. Dispositional realism on this account is the view that the dispositions of entities equate to their causal properties. Thus Chakravartty states: "The properties of entities... are *causal* properties: they confer dispositions for behaviour on the entities that have them" (2013, p117, italics in original). The relations between causal

<sup>&</sup>lt;sup>6</sup> Khalidi also argues that cross-cutting, and superficially incompatible, taxonomies are possible, and the choice of one's taxonomy at any point in time will depend upon the specific epistemic purposes of the relevant scientific community at that point.

properties are causal laws which constitute the laws of nature. Natural kinds are constituted by entities with causal properties which give rise to law-like behaviours of the entities which possess them.

Chakravartty identifies two types of natural kind and appears to want to distinguish them. Thus, he says: "I will refer to kinds with essences and those without as *essence* kinds and *cluster* kinds, respectively" (2007, p157, italics in original). The former have a definable essence which account for kinds in physics and chemistry. In contrast, cluster kinds, which are needed to account for kinds in biology and other special sciences (where essence kinds are inadequate), are similar to Dupré's conception of property clusters. Importantly, Chakravartty regards the causal mechanism element in Boyd's HPC account as too restrictive and argues that clusters of properties in natural kinds need not be causally linked. This raises the question of how property cluster kinds can be individuated, if not through causal associations of some kind. Chakravartty answers this by saying that such properties are "sociable", which he describes as follows:

Properties, or property-instances, are not the sorts of things that come randomly distributed across space-time. They are systematically 'sociable' in various ways. They 'like' each other's company. The highest degree of sociability is evidenced by essence kinds, where specific sets of properties are always found together. In other cases, lesser degrees of sociability are evidenced by the somewhat looser associations that make up cluster kinds. In either case, it is the fact that members of kinds share properties, to whatever degree, that underwrites the inductive generalizations and predictions to which these categories lend themselves. This is a reflection of the striking, poetic fact that some collections of property instances like each other's company and others do not (op cit, p170).

This notion is puzzling and Chakravartty concedes that "sociability is just a metaphor, intended to describe the *metaphysical* fact that in cases referred to as examples of kinds, property instances tend to cluster" (op cit, p170, italics added). Moreover, the statements about degrees of sociability seem to suggest that he may not after all view essence kinds and cluster kinds as distinct kinds of natural kind. His position here seems a little ambiguous. Nevertheless, he does not elaborate on this any further. To support the argument that the defining properties of a cluster kind need not be causally linked, he uses only one example, namely that of the electron as defined by specific values for charge, mass, and spin which are causally unrelated to each other. He argues that the defining properties of the electron, as with other essence kinds, are just a "brute fact". I shall examine this argument in more detail

below, but before doing so I shall briefly discuss Matthew Slater's similar account of natural kinds.

Slater (2015) largely endorses Chakravartty's account and in particular interprets the latter as meaning that there is a "spectrum of property 'sociability' phenomena" (2015, p380) – i.e. a continuum of cluster kinds, not a clear distinction between different kinds of kind. The kinds of property clusters thus produced are regarded by Slater as "stable property clusters" (SPCs). As a way of clarifying this conception of stability and drawing on Chakravartty's metaphor of sociability, he describes this as "cliquish stability" – the idea that certain properties like to gather together, in a metaphorical sense, in the same places at the same times. He attempts to explain this by reference to "sub-clusters" in which some, but not all, defining properties of the kind are known to be present in an entity. He says:

Call this conception of stability 'cliquish stability'....The idea is to capture the fact that some properties are clustered in such a way that possession of some of them reliably (if imperfectly) indicates the possession of the whole cluster (if not each property in the cluster) at that time (op. cit., p397).<sup>7</sup>

What this seems to mean is that the stability of a property cluster kind, and hence its status as a natural kind, can be indicated by the reliability with which the occurrence of various subclusters is likely to imply the occurrence of the whole set. Where the levels of such reliability are high, one can regard the property cluster as having sufficient stability to constitute an SPC natural kind.

In arguing that natural kinds do not need to be characterised by any causal maintaining mechanisms or processes, one might wonder whether Chakravartty and Slater need to find some other way of accounting for the stability of property cluster kinds. One possible response would be to reject such a demand on the basis that a brute fact is just that and, as such, it stands in need of no further explanation. Nevertheless, Chakravartty's use of the metaphorical notion of "sociability" and Slater's of "cliquish stability" suggests a recognition of a need for some sort of account of property clusters. Staying with this metaphor for a moment, the sociability of people who "like each other's company" would seem to suggest the existence of some set of relationships linking them together. More generally, any attempt at framing such an account seems to depend in some way on metaphorical notions explicitly

<sup>&</sup>lt;sup>7</sup> This statement of Slater's is a little confusing in suggesting that "the whole cluster" might be possessed at some instant without necessarily including "each property in the cluster". This might suggest a significant flaw in his account, but presumably Slater could re-phrase the statement to remove the apparent contradiction. I will therefore not dwell further on this particular point.

or implicitly involving ideas similar to that of sociability. Without any further elaboration, this would seem to be a weakness of any such account which seeks to dispense with causal processes as a fundamental feature constituting a property cluster natural kind. I argue here that natural kinds conceived as property clusters linked by causal processes, in the manner described by Boyd and Khalidi, is consistent with the ways in which natural kinds in particular sciences are studied for their explanatory utility. In the next section, I examine some objections to the Chakravartty/Slater conception to support the view that causal processes are associated with properties which constitute natural kinds.

#### 1.2.4 Objections to non-causal accounts of cluster kinds

There are a number of possible objections to the non-causal accounts of natural kinds described above. Thus, Boyd argues that the sociability of properties in natural kinds proposed by Chakravartty and Slater would require some form of "homeostatic 'buffering' processes that underwrite their sociability in the face of destabilizing forces" (2021, p.S2899) if the kind is to be an explanatorily important cluster kind. While he doesn't regard this as a conceptual truth, he thinks it's scientifically plausible.

More specifically, other objections can be made to such non-causal accounts. First, there is a paucity of clear examples in both Chakravartty's and Slater's accounts of natural kinds constituted by causally unconnected properties. Chakravartty gives only one – the electron, with specific values for mass, charge and spin. Slater gives a similar one – the up quark with a small number of fixed defining properties (spin, charge, mass, and baryon number). However, it does not follow from the fact that we don't currently know what causal relationship, if any, governs these properties that there is no such association. Although major theoretical advances in sub-atomic physics occur infrequently, when they do they can reveal causal explanations for phenomena previously regarded as just brute facts. Moreover, sub-atomic physics is a basic science which constitutes a bottom level of explanation in science. On its own, this objection is not a conclusive refutation of non-causal property cluster accounts, since it could still be maintained that electrons and up quarks fit this kind of

<sup>&</sup>lt;sup>8</sup> It can also be pointed out, as Slater does, that Boyd uses the metaphorical notion of "homeostasis" in his HPC account. Referring to this, Slater says: "I believe that an account of natural kinds ought to rest on firmer theoretical foundation than a metaphorical similarity with other known entities and processes" (2015, p394). Of course, metaphors always allude to something else. In Boyd's case, it is fairly clear what type of allusion he has in mind (e.g. stable self-regulating systems in natural entities). In Slater's case the allusion is to something, i.e. people in social groups, not obviously connected with the concept to be explained.

account and that this is sufficient to support the argument. However, it is notable that neither Chakravartty nor Slater provide a fully worked example of a non-causal cluster kind in any of the special sciences, although Slater hints, without going into detail, that there might be some in cell biology.

Second, reliance on brute facts as an adequate level of explanation does not fit well with normal scientific expectations. In all areas of science, we typically expect there to be some sort of explanation for important phenomena even where this is currently lacking. For practical reasons, we may temporarily accept brute facts in certain circumstances, particularly if we have other interests to pursue, but it does not follow that we accept that causal explanations are necessarily lacking in such cases. Whilst this is not a conclusive refutation, what is at issue here is how best to understand natural kinds *in science*, and as such we need an account that maximally reflects the way kinds in science function in grounding explanations and induction.

Third, Chakravartty seems to accept (though not always consistently) a distinction between "essence" kinds and "cluster" kinds. Assuming this distinction, the electron is clearly an essence kind, not a cluster kind. As such, the specific values for mass, charge and spin which define the electron clearly constitute the essence of the electron. More precisely, it is the conjunction of these values which is the essence and which is the necessary and sufficient condition for something to be an electron. Chakravartty's choice of this example (his only one) does not therefore have any bearing on whether there need to be any causal associations between the properties constituting a *cluster* kind. Slater does not place much emphasis on an essence/cluster kind distinction, but for the same reason it is not clear how his main example of the up quark (which also seems to have an essence in the conjunction of its defining values) can apply to cluster kinds. I have argued above that Khalidi's account of causal property clusters can be modified, such that chemical elements (and by extension subatomic particles) can be seen as the simplest end of a property cluster continuum where the constituting cluster is a single property, which may be a conjunction of specific properties. In such cases, the question of a causal link between defining properties need not arise, and the causal property cluster account is not thereby threatened.

The fourth objection requires lengthier consideration. All accounts of non-causal cluster kinds (Dupré, Chakravartty, and Slater) seem to assume that when salient properties are clustered in a regular sort of way (however that might be understood), then the individuation of a natural kind which encompasses these properties will be a relatively straightforward matter. However, it is not clear how such a kind would in fact be individuated. What these

accounts emphasise is that the defining properties need have no causal association with each other. The individual properties may well have causal implications, but what is supposed to be important about the kind is that *it*, *qua* natural kind, should be able to ground explanations and induction over and above the effects of the individual properties. But it is difficult to see how a *cluster* of properties with no causal connection to each other could have epistemic significance beyond that of the individual properties. A cluster of such properties is nothing more than a correlation. Without any causal associations between the properties, the individuation of a particular kind would depend solely on the strength and statistical significance of the observed correlation, and this would seem to introduce an unavoidable element of conventionalism, and potentially an arbitrary one, into the manner in which kinds are identified – that is according to the conventions employed within a given epistemic community.

A possible response to this worry is given by Slater who argues that SPC kinds can be "domain-relative", which would mean that any element of conventionalism in individuating kinds could be determined according to accepted practice in the discipline concerned. He says: "... the norms and aims of certain domains may require different levels of cluster cohesiveness – that is, different disciplines may tolerate different degrees of flexibility in the clustering required by their respective kinds" (2015, p403). What Slater means by kinds being "domain-relative" appears very similar to Boyd's account of natural kinds as being specific to a given disciplinary matrix. Nevertheless, where Boyd insists that HPC kinds must be held together by causal mechanisms of some sort, Slater seems to allow that some domains or scientific disciplines may tolerate quite loosely correlated property clusters without evidence of causal linkages. In such cases, it is not clear how natural kinds could actually be individuated without making prior assumptions about what kinds one ought to find.

To see how this might be difficult, one should consider how properties manifest themselves. They do not just float around freely detached from the individuals or entities which are supposed to possess them. Instead they are discernible as attached unavoidably to particular entities and, where the cluster of properties constitute natural kinds, they attach to the entities that are members of the kind. Thus, the defining property of hydrogen, its atomic number, is attached to every atom of hydrogen in the universe. Similarly, the properties which clustered together constitute the kind "grey squirrel" attach to every actual grey

-

<sup>&</sup>lt;sup>9</sup> The ways in which clusters of properties are discussed in some of the literature, as though they are somehow independently existing entities, may sometimes encourage this impression.

squirrel (living, dead, and not yet born) in the world, although as a property cluster kind there will be some imperfection in how these properties are manifested across the individual kind members. What is clear, however, is that the properties constituting the kind "grey squirrel" are highly correlated with each other – actual grey squirrels are all very similar, despite whatever individual variations there might be. These properties are also causally linked (as Boyd argues they are for all species), even though the nature of the causal relations are complex.

I argue that similar issues are in play for natural kinds in other sciences, and in medical science in particular. This can be seen in the way in which recent scientific progress has led to strengthening the case that Alzheimer's disease can be seen as a medical natural kind. In the past, when evidence of salient biomarkers was generally absent, diagnosis of the condition was very difficult. The patient's symptoms, with which clinicians were typically presented, were memory loss and changes in personality which could be quite subtle. Problems of differential diagnosis could arise between dementia and depression (which could appear quite similar on initial presentation), as well as between different types of dementia (e.g. Alzheimer's disease, Pick's disease, etc.). Consequently, patients with Alzheimer's disease could display various symptoms, all of which might also be indicative of other conditions. In other words, Alzheimer's disease was characterised by a cluster of symptoms (properties) which overlapped with other symptom clusters indicative of other conditions and, although it was believed likely to be a distinct disease, the aetiology was not understood. Consequently, the individuation in this case of a natural kind *purely* on the basis of clusters of properties would have been very difficult. That it was considered a distinct disease kind is attributable to the physician, Dr. Alois Alzheimer, who in 1906 discovered in an autopsy on a patient a dramatic shrinkage in her brain and presumed this to be causally related somehow to her declining psychological state prior to her death. In modern times, with the availability of CT and MRI scans, it has become possible to improve diagnostic reliability. Even though the causal processes are not yet fully understood, it now seems clear that the aetiology is connected in some way with the accumulation of amyloid proteins and neurofibrillary tangles in the brain. As such, this disease can now plausibly be understood as a natural kind conforming to both Boyd's and Khalidi's causal accounts.

If causal processes are not considered necessary to the individuation of natural kinds, we must then ask how else they can be identified, given the difficulties just discussed. If this is done just on the basis of property clusters, one might suppose that some standard of similarity between entities in a putative kind, perhaps in terms of significant shared

properties, would be sufficient to indicate a cluster kind. It is difficult however to articulate what such a notion of similarity could amount to and, by extension, how significance could be evaluated without a prior notion of kind to ground this. Quine (1969) argues that notions of "kind" and "similarity" are inter-dependent in any particular case, and that consequently there is no primitive notion of "similarity" which can do the required work of identifying a kind. In the absence of such a notion of similarity, we would therefore seem to be left with clusters which could just be brute facts (even if some happen to have a causal basis) and thus unexplainable. However, as I have argued above and as Khalidi also argues, this seems inconsistent with the epistemic aims of the sciences which are driven by the search for *causal* explanations of salient phenomena.

To summarise, if natural kinds themselves, and not just the individual properties comprising them, are important in grounding explanation and induction, it is hard to see how they can function in this way if the constituting properties are not connected causally in some way. Where there happens to be a group of entities appearing to display a significantly correlated cluster of properties, without any clear causal associations amongst them, these may well be taken as *evidence* for a possible natural kind and therefore guide further epistemic efforts. However, if we expect an account of natural kinds which optimally reflects how explanatory categories are formulated and how their causal implications are studied in all the sciences, then a non-causal account of kinds seems to lack the resources to do this. By contrast, Khalidi offers a unifying causal account which is applicable across the sciences, including the special sciences and medical science in particular. In the next section, I argue that diseases can in general be regarded as natural kinds in medicine and that Khalidi provides a plausible account of how disease kinds can function as explanatory concepts in medical science.

#### 1.3 Diseases as natural kinds

Having indicated above my claim that diseases can be regarded as natural kinds, it is probably more strictly accurate to say that *disease entities* can be so regarded. However, for most of this discussion and for the sake of simplicity, I continue using the term "diseases" where the context makes clear that disease entities are being considered. The notion of "disease entity" is of central explanatory importance in medical science where epistemic and diagnostic efforts are directed towards understanding the causes of ill health in patients (Hucklenbroich, 2014; Simon, 2010; Whitbeck, 1977). The manner in which disease entities

are understood in terms of their particular aetiologies – i.e. the sets of factors which in combination cause diseases – shows that these entities can plausibly be seen as natural kinds in accordance with the causal account of kinds discussed above. I start by describing the notion of "disease entity" in more detail. I then discuss the kinds of explanations that medical scientists seek in explaining diseases. Following from this, I then make the case that diseases can plausibly be seen as natural kinds. Having said that, it is necessary to distinguish the thesis that diseases (in the plural) can be natural kinds from the view that *disease* (as a singular and general concept) is a natural kind. I do not argue for the latter view. I end by emphasising that natural disease kinds can be explained effectively by Khalidi's account, but that non-causal accounts of kinds are too liberal in their scope to usefully explain the epistemic aims of medical scientists.

The understanding of disease entities as defined by aetiology can be traced back to the success of the germ theory of disease in the 19<sup>th</sup> century (Whitbeck,1977). This initially conceived of disease entities as deriving from the disease process initiated by the relevant infective agent. The disease entity therefore became associated with a specific aetiological agent. However, as Whitbeck argues, this model is inadequate for understanding more complex disease processes which cannot be attributed to a unique aetiological agent in this way. Instead, diseases such as cancers and those deriving from genetic abnormalities are better understood as a complex of processes interacting with background conditions in the organism and the environment. Nevertheless, a disease entity need not apply solely to those conditions characterised by unique aetiological agents, but can equally embrace complex diseases of the sort indicated by Whitbeck.

A more detailed account of the concept of "disease entity" is given by Peter Hucklenbroich (2014). According to Hucklenbroich, this is a theoretical concept which is used to explain particular disease instances and guide diagnosis in individual cases. As he conceives it, the distinction between "disease" and "disease entity" is that the former refers to the whole course of the illness from cause to outcome in the individual case, and the latter refers to the *type* to which the individual disease instance belongs. In other words, this reflects a token-type distinction – "That is to say, every disease is a *case* or *instance* of a disease entity" (Hucklenbroich, 2014, p613, italics in original). Jeremy Simon also endorses this view of disease entities as types of disease and offers an example of this kind token-type distinction:

Medical science and practice make constant, essential reference to various conditions – cystic fibrosis, tuberculosis.... – both as types, when, e.g. looking for a cure for cystic fibrosis, and as tokens, as when we say that someone has cystic fibrosis (2010, p333). It follows from this that there cannot be any disease instance which is not a token of a specific disease entity (although in complex cases it may not be possible to diagnose which disease entity the patient's illness reflects, or even to be sure that the patient suffers from a currently known disease). It is also possible for a patient to be suffering from more than one disease at the same time in cases of co-morbidity.<sup>10</sup>

Hucklenbroich makes clear that he regards disease entities as natural kinds in the following statement.

The fundamental idea lying behind the concept of disease entities is this: disease is a phenomenon neither totally uniform nor totally variable concerning its form of appearance; rather, it presents itself in the form of certain definite, natural *kinds* or *classes* of disease... Just as there are *species* of animals and plants, and *classes* of chemical substances in nature, there are natural kinds of diseases, called *species morborum* or *disease entities*... Therefore, any disease entity – like *influenza* – is a theoretical entity of medicine, like any element – for example, *oxygen* – is a theoretical entity of chemistry (2014, pp611-2, italics in original).

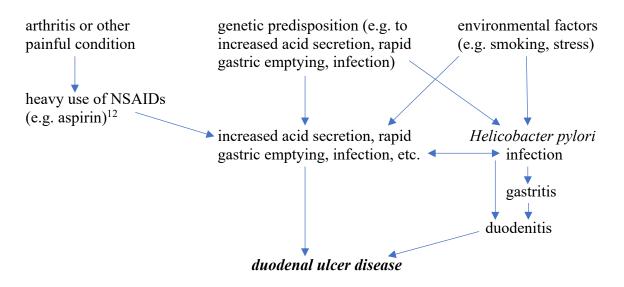
He does not in any way elaborate on what conception of natural kinds he has in mind, although his reference to disease entities being "neither totally uniform nor totally variable" suggests that they are likely to conform to a property cluster conception. He also makes clear that disease entities have distinct causal structures underlying them: "disease entities are able to form a basis for causal (etiopathogenic) *explanation* of symptoms and other manifestations or findings" (op. cit., p616, italics in original). This is consistent with approaches typically used by medical scientists in their search for disease explanations in terms of distinct causal structures (e.g. Thagard, 1999). Therefore, the conception of natural kinds implied here is one of causally linked property clusters. These seem to be more easily conceived in terms of Khalidi's account, rather than Boyd's, in that the structures of the relevant causal processes

<sup>&</sup>lt;sup>10</sup> I discuss concepts of disease in more detail in Chapter 3, where I consider normative accounts of disease in which there is no naturalistic element specifying that there must be some kind identifiable aetiology in order to individuate diseases. While I shall argue that a plausible account of disease requires a normative element, I shall argue for a hybrid account which should also include a naturalistic element. The discussion of disease entities here assumes a naturalistic or hybrid account of disease.

seem to be very varied depending upon the particular disease under consideration. Whilst it is plausible to conceive of aetiological causal processes in diseases as "mechanisms" in Boyd's sense, in many diseases it is hard to see how these can be "homeostatic" causal mechanisms, given that many diseases progress in their course, sometimes very rapidly, such that particular disease instances are fundamentally unstable. This seems at odds with the notion of homeostasis implied in Boyd's conception of kinds of biological species, which remain stable over very long time periods.

As indicated, medical scientists seek causal explanations for disease entities.<sup>11</sup> Disease explanation is what Thagard calls "causal network instantiation" (1999, p113). Such explanations typically involve complex causal networks identifying how the disease entity might be manifested in individual cases. These networks demonstrate the causal relationships between the defining symptoms of the disease entity and various factors internal and external to the patient which affect the actual course of the disease in the patient. This can be illustrated in the case of duodenal ulcer disease in Figure 2, taken from Thagard (1999, p115).

Figure 2:



In this figure, duodenal ulcer disease is the disease entity, and the causal network indicates the various causal factors influencing how the disease can be instantiated. Different disease instances in different patients will depend in part upon the extent to which the various

35

<sup>&</sup>lt;sup>11</sup> Correspondingly, medical practitioners seek explanations for disease instances in individual patients. That is, they attempt to make a *diagnosis* of the patient's illness, which will usually, but not always, make reference to the relevant disease *entity*.

<sup>&</sup>lt;sup>12</sup> NSAIDs are non-steroidal anti-inflammatory drugs.

background conditions – e.g. genetic predisposition, history of smoking or stress, heavy use of NSAIDs, etc. – are present in individual cases. Thagard makes a number of points about this kind of explanation. First, this explanation is not deductive in the sense that there is no universal law from which the occurrence of an ulcer can be deduced. Thus not everyone with *H. pylori* bacterial infection will develop an ulcer. Second, explanation is not statistical. Whilst correlations may indicate evidence of a causal relationship, they are not themselves explanatory. Third, explanations do not generally consist of single causes. As the figure above indicates, although *H. pylori* is a major cause of ulcers, a range of other causal factors influence whether and how it actually results in the manifestation of disease.

Thagard argues that the causal network instantiation model fits well with the kind of explanations sought by medical researchers. It is applicable to other complex diseases such as diabetes, cancer, and so on. The causal reasoning involved is characterised by abductive inference that certain factors have the power to produce effects. Although Khalidi does not refer to Thagard's explanatory model, it is clearly consistent with his unifying causal account of natural kinds. A further feature of the model is that the causal relationships can be understood in terms of James Woodward's (2003) manipulability model of causal explanation, with which Khalidi's account of kinds is also consistent. In the example of duodenal ulcers above, interventions or manipulations in some of the causal processes can produce different outcomes. Thus, an individual who limits their consumption of NSAIDs, stops smoking, or reduces their life stressors can be expected to avoid the kinds of processes that lead to ulcers. As Woodward argues, the causal factors need not be immediately amenable to human intervention to be shown to be causally relevant. It is enough if they can be shown to be causally effective in clinical trials where, for example, patient groups may be differentiated on the basis of different levels of measured stress which then demonstrate different incidences of the disease.

If, as Hucklenbroich claims, disease entities are natural kinds, which actual diseases (in the sense of "disease" used in common parlance) are natural kinds? In other words, which can be seen as disease entity natural kinds? A difficulty in answering this question arises from the wide and rather imprecise usage of the word "disease". Indeed one interpretation of Hucklenbroich's arguments is that he is attempting to suggest a more precise understanding of "disease" which corresponds with its usage by medical researchers. There is controversy within medicine, however, over whether some conditions, such as chronic fatigue syndrome (CFS) and irritable bowel syndrome (IBS), actually constitute real diseases (Reiss & Ankeny, 2016). In the case of CFS, the main symptom (as the name suggests) is fatigue after exertion

with chronic duration. Other symptoms are very varied, as is the severity of these. There is no known actiology for the condition and the prognosis is very varied. It is nevertheless possible that the various symptoms characterising it may be significantly correlated. If so, it might constitute a non-causal property cluster kind, but not one that would fit a causal account of natural kinds. Whilst such a non-causal account might be applicable here, there is insufficient reason to regard CFS as a distinct *disease entity* of the kind which medical scientists expect to be able identify with a known causal basis. Similar considerations would apply to IBS, which also has no known actiology and uncertain prognosis, and to other conditions where the clustering of symptoms appears to be loose without any evidence of causal connections. Clearly however, in the event that future research reveals a distinct causal history for one or other of these conditions, there would then be grounds for recognising that such a condition does represent an identifiable disease entity and thus a natural kind with causally connected symptoms.

The purpose in identifying distinct disease entities with clear aetiologies is to support clinicians' efforts at prevention and cure of diseases. Where such a disease entity is identified, it then becomes possible, at least in principle, to target interventions at pivotal points in the causal network to prevent or cure the disease. Similarly, interventions can be targeted at the population level (e.g. campaigns to reduce smoking) to reduce the incidence of a given disease. Without any clear indication of a causal network in syndromes such as CFS, no such interventions are possible. Disease entities therefore indicate possible causes, interventions, and prognosis. Or, in the language of natural kinds, disease kinds ground explanation and induction – they are projectible – and the corresponding categories constitute a medical taxonomy. Thus the role of natural kinds in medicine is described by Hilary Kornblith as follows:

.. [medical researchers'] interest in treatment can only be served by allowing their taxonomies to be shaped by, and successively approximate, the real kinds in nature. The successive transformation of medical terminology to reflect the causal structure of the world thus mirrors the history of chemical terminology (1993, p50).

<sup>&</sup>lt;sup>13</sup> As noted in footnote 10, I discuss concepts of disease in Chapter 3. If one accepts a purely normative account of disease, there might be no reason to judge that CFS is not a real disease. However, the fact that medical practitioners debate this question suggests that, at least implicitly, they employ a naturalistic or hybrid conception in their thinking.

A number of other philosophers have argued that diseases can be natural kinds.<sup>14</sup> Various examples are discussed in this context and taken together they represent a wide range of types of disease (e.g. auto-immune, genetic, etc.) in addition to those caused by a single infective agent, all of which have explanatory value in explaining particular sets of symptoms. Thus Marc Lange argues:

Medicine aims to identify the disease(s) afflicting a patient. Such a diagnosis is intended to *explain* the patient's signs and symptoms. Therefore, a disease category must have 'validity' (in the medical sense), which means that the disease must be a natural kind rather than an arbitrary category (2007, p266, italics in original).

It is clear from this that Lange, in this passage, intends "disease" to equate to "disease entity" in the sense used by Hucklenbroich and Simon. This is evident in the following passages:

For a patient's disease to explain her signs and symptoms, the disease must be distinct from its clinical picture, since otherwise physicians would be calling upon the picture to explain itself. Likewise, if having the disease were nothing but exhibiting enough signs and symptoms from a certain category, then the disease would not be explanatory; that Jones exhibits two or more of symptoms A, B, and C fails to explain why Jones exhibits symptom A (op. cit. p268).<sup>15</sup>

.. one goal of medicine has been to identify the real diseases. Just as two geological samples tend to share certain properties because they are samples of the same mineral, so two patients tend to share certain properties because they have the same disease (op. cit. p269).

To illustrate how this works, Lange uses the example of phenylketonuria (PKU), an inherited genetic disease, as an example of a natural disease kind. PKU is characterised by a deficiency in a certain enzyme necessary for metabolizing phenylalanine (an amino acid). When diagnosed at birth, the affected infants have to be fed on a diet free of the amino acid and this dietary restriction has to remain in place for life. Failure to do so will cause microcephaly and intellectual disability in the child.

<sup>15</sup> This is an important issue in the context of psychiatric diagnostic categories which are frequently characterised by just such a polythetic diagnostic approach.

<sup>&</sup>lt;sup>14</sup> The philosophers who argue for this generally use the word "disease" to represent what Hucklenbroich understands by "disease entity". This is normally clear from the context of their respective discussions.

Another example of a natural disease kind, this time of Graves disease (an autoimmune disease), is discussed by Stefan Dragulinescu (2010). This disease is defined by the presence in the blood of thyroid receptor antibodies and high levels of thyroxine. This is caused by a complex set of biochemical processes interacting with certain background conditions (e.g. genetic predisposition, history of smoking, etc.) to produce a range of symptoms in the patient, including goitre, exophthalmia, weight loss, and various others, not all of which are evident in each patient with the disease. As such, this is a clear example of a property cluster natural kind where the actiology and symptoms are linked together within a complex causal network. Another philosopher who argues that diseases may be natural kinds is Neil Williams (2011), who cites rheumatoid arthritis as an example. All these philosophers conceive of diseases as property cluster kinds of varying degrees of causal complexity and Williams explicitly draws on Boyd's HPC account as a basis for his conception of disease kinds. Due to the recent publication of Khalidi's account of natural kinds, none of these philosophers refer explicitly to it. Nevertheless, Khalidi argues that his account is as applicable within medical science as it is in other special sciences.

There does not appear much criticism in the literature of the view that many diseases are natural kinds. Reznek (1995) argues that diseases as a group do not constitute a natural kind, but he makes clear that it is the general concept 'disease' which is his focus and not individual diseases. He leaves open the question of whether individual diseases are natural kinds. Sulmasy (2005) does argue that diseases are not natural kinds, but his argument is based on an essentialist account of kinds and makes no reference to Boyd or other property cluster theorists. This therefore does not bear directly on the accounts of disease kinds considered above. Moreover, he seems to treat "disease" as meaning "disease instance" rather than "disease entity". In the former sense, one can accept that they are not natural kinds, precisely because they are just individual instances of a particular category and it is the category which is expected to have explanatory significance. However, he hints at the notion that disease explanation may rest on natural kinds of some sort, without developing this clearly: "The purpose of disease classification (nosology) is, in the first place, explanatory" (2005, p496). It seems that the natural kinds he has in mind are the organisms (particularly humans) which succumb to diseases, but his reference to the explanatory function of diseases suggests there is some kind of causal role played by diseases themselves. It seems therefore that his argument does not threaten the view of disease kinds as property clusters in causal networks.

In summary, there is good reason to think that diseases, or more accurately disease entities, can count as natural kinds. The distinction between disease entities and disease instances is a type-token distinction and it is these *entities* which correspond to natural kinds. There is little support amongst philosophers for the view of diseases as essentialist kinds, but rather for a property cluster account which reflects the complexity and variable presentation of many diseases.

#### 1.4 Conclusion

I have argued here that many disease entities can be regarded as natural kinds in medical science. In order to demonstrate this, I have discussed different accounts of natural kinds here and have argued that property cluster accounts in which the clustering of properties is a function of causal processes, such as those described by Khalidi and Boyd, are applicable across the sciences generally and consequently also in medical science. While Khalidi emphasises the difference between his account and Boyd's, both of them describe natural kinds as being constituted by causally linked property clusters. In this respect, they both offer accounts of natural kinds as being non-arbitrary categories in scientific disciplines. Khalidi also argues that his provides a unifying account of kinds across the sciences. I have argued that both Boyd and Khalidi offer a more powerful account of natural kinds than non-causal property cluster accounts. In particular, this is the case in the context of medical science, in that it has the resources to distinguish between clearly definable disease entities such as the examples discussed above and mere clusters of symptoms which might constitute a conventional syndrome (such as CFS) without any evidence of a corresponding disease entity.

In the next chapter, I discuss whether psychiatric diagnostic categories as defined in the Diagnostic and Statistical Manual of the American Psychiatric Association (2013) can also constitute natural kinds, on the basis of the accounts given by Boyd and Khalidi.

### Chapter 2

#### The natural kind status of psychiatric diagnostic categories

#### 2.1 Introduction

The theme of this thesis is the question of what effect the assignment of a psychiatric diagnosis may have on the individual receiving the diagnosis. Such diagnoses are given in a medicalised context and thus are likely to convey the message that the individual concerned has an illness of some kind, however broadly conceived. A key question to be examined therefore is the extent to which a psychiatric diagnosis represents a disorder that can be regarded as a disease of some kind or other biomedical condition.

The arguments developed in Chapter 1 support the view that disease entities in general (although not necessarily in all cases) can be seen as natural kinds in medical science. This is based on a conception of natural kinds according to which they are delineated by clusters of properties with causal linkages amongst them, as described by Richard Boyd and Muhammad Ali Khalidi. An important aspect of these accounts is that natural kinds are a means of reflecting the causal structure of the world within the relevant scientific discipline. In Boyd's terms, this means that natural kinds are the concepts through which we accommodate our thinking to the causal structure of the world. For Khalidi, they are the concepts forming the nodes in the causal networks that constitute the basis of our explanations in science. In other words, natural kinds on these accounts are expected to feature in scientific explanations for phenomena of interest in the relevant sciences and to have some explanatory value.

In this chapter I discuss the question of whether mental disorders can also be seen as natural kinds on the same basis. The answer to this question is not necessarily the same for each psychiatric diagnostic category, since the role these concepts play in explaining the patterns of symptoms displayed by individuals is likely to differ, depending on the category in question. To be more precise, the issue is not about whether mental disorders generally are natural kinds, but rather whether those disorders that have been explicitly labelled as disorders and given specific names in the recognised diagnostic manuals can be seen as kinds on the above basis.<sup>16</sup>

<sup>&</sup>lt;sup>16</sup> There has been some criticism of the widespread use of the term 'disorder' rather than 'distress' on the basis that this begs the question of how the experiences of people suffering severe psychological

In section 2.2, I start with a brief overview of the debate about whether the disorders as classified in the Diagnostic and Statistical Manuals (DSM) of the American Psychiatric Association can be so viewed and what objections have been raised against such a view. Another diagnostic manual in regular use is the International Classification of Diseases, 11<sup>th</sup> edition (ICD-11), published by the World Health Organisation (2019/2021). This is a multichapter classification of all recognised diseases and pathologies, with mental and behavioural disorders being listed in its fifth chapter, and is regularly used by clinicians in the UK. The ICD and DSM categories closely parallel each other, although the ICD categories are not as closely defined in operational terms as those in the DSM. Since the two manuals are very similar and much of the debate about diagnostic categories has focussed on the DSM, I therefore discuss the DSM, rather than the ICD, from now on. The conclusions I reach about the natural kind status of the diagnostic categories are applicable to both.

In section 2.3, I argue that validity in some form needs to be demonstrated if the diagnostic categories are to be understood as representing natural kinds. I discuss what validity in this context might mean and how attempts have been made to demonstrate validity for the DSM categories. However, these attempts have not been successful. It is now largely recognised that most of the categories have not been shown to be valid, although they are still considered to have utility (e.g. Jablensky, 2016).

In section 2.4, I discuss how the traditional approach to validation of each category has depended on the assumption that the condition is fully represented by the criteria by which the condition is diagnosed. However, this assumption does not address the idea that there might be an underlying condition for which the diagnostic criteria are simply signs of its presence, which would suggest that any approach to validation based solely on the diagnostic criteria could miss an important aspect of validity. Therefore, there is a question, as John Campbell (2017a) argues, of whether we need to entertain some kind of hypothesised governing conception of each disorder, such as a possible disease process, if the categories are to be validated. Without such a conception, the DSM categories would seem to be nothing more than clusters of symptoms forming syndromes with no clear causal mechanisms.

Following from this, in section 2.5, I discuss the ways in which these categories have frequently been understood in clinical practice, which has often involved an assumption of

distress should be understood (e.g. Kinderman et al., 2013). Whilst this is an issue that can be linked with the debate about a disease model of mental distress, it is most convenient for my purposes here to continue using the term 'disorder', since this relates clearly to the language used in the diagnostic manuals and other literature.

some underlying condition beyond what is described by the specific diagnostic criteria. I discuss the tendency to reify the diagnoses which can lead to the expectation, despite their being solely defined by the diagnostic criteria, that they somehow represent an underlying reality of some kind. Such an underlying reality may be believed to be in some way causally relevant to the presentation of symptoms in the individual patient, even when, as is generally the case, there is no consistent evidence of such an underlying cause.

I next discuss, in section 2.6, whether DSM diagnoses can have explanatory value and I claim that they do not. In sections 2.6.1 and 2.6.2, I consider two arguments to the effect that they may be explanatory and conclude that neither presents a strong counter-argument to the view that the DSM diagnoses are not explanatory. In section, 2.7, I examine the question of whether the category of schizophrenia can constitute a natural kind. I conclude that it does not, on account of the lack of explanatory value associated with it. I conclude in section 2.8 that the DSM categories are unlikely to represent natural kinds.

#### 2.2 The natural kind status of mental disorders

The question about whether mental disorders constitute natural kinds is typically addressed in terms of whether the specific diagnostic categories listed in the diagnostic manuals refer to distinct disorders which might be natural kinds. In other words, do specific categories, such as major depressive disorder, generalised anxiety disorder, schizoaffective disorder, etc., refer to natural kinds? Different philosophers have taken different views on this. Some (e.g. Haslam, 2014; Zachar, 2000) have argued that in most cases they cannot represent natural kinds. Typically, the arguments in support of this position are based on an essentialist conception of natural kinds, according to which a kind is characterised by a defining property, or set of properties, which constitute a necessary and sufficient condition for kind membership.

However, other philosophers have argued that such an essentialist conception of kinds is too restrictive for assessing the natural kind status of mental disorders, since in most cases no such necessary and sufficient conditions can be identified. Instead, they argue that mental disorders can be better understood as property clusters, no single property of which constitutes a defining essence of the disorder. Thus, Rachel Cooper (2007) suggests that diagnostic categories representing disorders may be natural kinds on the basis of Dupré's conception of kinds, in which the characteristic symptoms seem to correlate together, if not all of them are present in any given case, and the examples of the disorder are similar to each

other in various ways. Others (e.g. Beebee & Sabbarton-Leary, 2010; Tsou, 2013) argue that Boyd's homeostatic property cluster (HPC) conception of natural kinds is applicable to mental disorders. Beebee and Sabbarton-Leary give schizophrenia and Tourette's syndrome as examples of conditions that can be natural kinds on Boyd's account.<sup>17</sup> Tsou proposes that depression and suicide are natural kinds on the same basis.<sup>18</sup> Similarly Samuels (2009) suggests that delusion is a natural kind in Boyd's terms. Whilst these proposals for natural kind status of the disorders mentioned are debatable, they highlight the tendency for Boyd's account to be the preferred model in recent literature for application in this context.

Discussing the question of whether delusions constitute a natural kind, Dominic Murphy (2014) argues that Boyd's account has the advantage of focussing on causal processes as the determining factor for establishing the natural kind status of any given disorder. It is not enough to assume that a putative disorder, such as delusions, is a natural kind purely on the basis that certain clusters of features of those presenting with the disorder strike us as similar in some way. Instead Murphy argues "There must also be shared causal processes" (2014, p120) if the concept representing the presumed disorder is expected to reflect a scientifically significant condition. Such causal processes must be relevant to explaining how the disorder has arisen in the particular cases where it is identified, and also be capable of grounding inductions about the course and outcome of the condition. In these respects, the conception of natural kinds against which mental disorders should appropriately be judged is in effect the same as the Boyd/Khalidi conceptions discussed in Chapter 1, where I have argued that this approach allows us to determine natural kind status for medical diseases.

### 2.3 Traditional approach to the validation of DSM categories

The views outlined in the previous section suggest that many of the DSM categories, or some future refinement of them, could potentially be understood as natural kinds. For this to be the case, the categories would need to be shown to be valid in some way. Without any such demonstration, the claim that they may be natural kinds with significant explanatory value

\_

<sup>&</sup>lt;sup>17</sup> I argue below that there are serious difficulties in seeing schizophrenia as a natural kind on Boyd's account.

<sup>&</sup>lt;sup>18</sup> Suicide is a doubtful candidate for consideration as a natural kind. It is not strictly speaking a mental disorder at all, but an action committed by an individual who, in many cases, may be suffering from such a disorder, though not in all cases. Sometimes suicide may be a deliberately chosen action by someone with no diagnosable disorder.

has little other argument to support it. Consequently I examine here the manner in which attempts have been made to establish validity for the DSM categories and the limitations that such attempts have encountered.

This raises the question about what is meant by *validity* in this context. This, it should be noted, is quite different from the concept of validity in formal logic. In the context of psychiatry, there are difficulties in defining it precisely. Kendell and Jablensky explain that: "There is no single, agreed upon meaning of validity in science, although it is generally accepted that the concept addresses 'the nature of reality'" (2003, p5). Loosely put, validity is often taken to mean that the relevant concept represents reality in some way – i.e. that if there is a concept X which is valid, then there is some entity in reality that X represents. This, of course, leaves open the question of how this is to be determined in any given situation. Since typically this cannot be directly demonstrated, alternative approaches to assessing validity have to be used.

The concept of validity is regularly employed in psychology and psychometrics, such as intellectual and personality assessments, in which a test is considered to be valid if there is evidence that it measures what it is intended to measure. For example, an intellectual assessment tool is valid if it can be shown to measure intelligence to an acceptable standard of accuracy. Factor analysis of the results of standardised intellectual assessments carried out on a large number of individuals reveals a general factor, typically referred to as 'g', which accounts for a substantial proportion of the variance on the performance of people on these assessments. This is represented by IQ scores on individual assessments and is taken to measure a stable faculty of 'general intelligence' in the person concerned. There are several types of validity which can be used to establish validity for such measures. Two frequently used types in psychometrics are construct validity and predictive validity. Construct validity refers to the extent to which the measure in question reflects the underlying theory on which the measure is based. Predictive validity is said to be established if the measure successfully predicts future outcomes or properties. In the case of IQ assessments, some level of predictive

-

<sup>&</sup>lt;sup>19</sup> There has been some contention over many years about whether 'g' can be interpreted in this way, or even whether statistical analysis actually supports such a factor. I am not taking a position on this debate. My point here is just that 'g' is frequently interpreted in the manner I have described.
<sup>20</sup> There is some potential for confusion in talking about types of validity in this way. It is common in psychometrics to talk about validity in general and specific types of validity as contributing to establishing validity for the measure in question. A measure may be said to be valid if it is shown to be valid on one or more types of validity.

validity can be assumed if positive correlations between the test scores and subsequent educational attainments can be demonstrated.

In the context of the DSM classifications, it is doubtful that many of them have construct validity, because of the way they were explicitly framed without any intention that they would reflect any underlying theory about the aetiology for the conditions to which they refer (I will discuss the development of the DSM further below). Predictive validity would seem to be important, in that one would expect disease or illness categories to have some implication for the outcome for the persons affected. However, I will argue below that for many DSM categories the outcomes are actually very variable.

There has been much debate about the validity of the DSM categories and how this should be understood and assessed. Much of this arose following the publication of the third edition (DSM-III) in 1980 (American Psychiatric Association, 1980), which marked an important reconceptualization of the aims of the manual and the manner in which its categories were developed. The previous versions of the manual (DSM-I and DSM-II) made use of rather vague and ambiguous criteria for determining diagnoses, and the manner of their formulation was based on theoretical views about their presumed aetiology (First, 2012). However, with the increasing use of DSM-II categories as a basis for empirical research, it became clear that these categories were unreliable, in that research carried out in one centre could not be replicated elsewhere, because the category descriptions could permit significantly different interpretations. This is an issue of inter-rater reliability, according to which one would expect there to be a strong probability that two raters independently assessing the same individual or situation would arrive at the same conclusion. Without first establishing that the categories are reliable in this way, there would be little prospect of demonstrating that they had significant predictive validity. Reliability is a necessary condition for validity, but not a sufficient one.

For example, in a study by Kendell and colleagues (1971), groups of British and American psychiatrists were shown videotapes of psychiatric cases. It was found that the American psychiatrists assigned a diagnosis of schizophrenia much more frequently than their British counterparts, who made greater use of other diagnoses, such as manic-depressive psychosis or personality disorder, for the same cases. It was also clear from comparative hospital statistics that psychiatrists in the USA were diagnosing schizophrenia much more frequently than those in the UK, where the latter were making more use of the diagnosis of manic-depressive psychosis (First, 2012). These findings were very concerning, since they were felt to put the credibility of the psychiatric profession in jeopardy (Aragona, 2015).

The intention, therefore, in creating DSM-III was to eliminate as far as possible the unreliability inherent in the earlier editions. In order to achieve this, the DSM committees adopted two changes to the way the categories were constructed. Firstly, they decided that DSM-III would be constructed around the aim of maximising the reliability of the defining criteria of the diagnoses, such that different users of the manual would be likely to assign diagnoses to their patients in a consistent manner. Whilst previous editions had aimed for reliability to some extent, the identification of diagnosable disorders was based on glossary definitions, rather than anything more precise. Hence, in order to achieve enhanced reliability, every disorder was defined as far as possible in terms of operationalised criteria which could be assessed directly by clinicians without making assumptions about any presumed and unproven aetiology (Aragona, 2015). Thus, the second change followed from this, in that no assumptions were made in defining the categories about the aetiology or causal processes of the disorders.<sup>21</sup> The new categories were therefore stated to be atheoretical, in the sense that their use in practice would not depend on any theoretical assumptions about their aetiology. However, it is not clear that it was invariably understood to be atheoretical and some critics argued that this would be conceptually impossible. Where aetiological processes were relatively well understood, these were incorporated into the definition of the disorder itself (Aragona, 2015). Nevertheless, it was anticipated that the changes introduced into the new manual would enable it to prove useful to researchers and clinicians alike. These changes were carried forward into the fourth edition, DSM-IV (American Psychiatric Association, 1994).<sup>22</sup>

Whilst these changes made for some degree of improved reliability in diagnoses made by clinicians, the validity of the diagnostic categories was not automatically guaranteed by that. Reliability does not entail validity. Therefore, there remained the need to establish appropriate means by which the validity of the categories could be established.

Since the DSM categories were purposely created in such a way as to avoid assuming the existence of a corresponding underlying entity with an aetiology which might be a common feature of the disorder, the validity of the categories needs to be understood and established in a different way. A method for achieving validity was outlined by Robins and

\_

<sup>&</sup>lt;sup>21</sup> There were a small number of exceptions to this where a diagnosis would be dependent on information about aetiology – e.g. post-traumatic stress disorder, where the diagnosis is only given if it is known that the disorder can be attributed to an earlier traumatic experience.

<sup>&</sup>lt;sup>22</sup> However, the criterion that the diagnostic categories should be atheoretical has been dropped from the fifth edition, DSM-5 (American Psychiatric Association, 2013). I discuss this change further below. The discussion in this section therefore refers primarily to DSM-III and DSM-IV.

Guze (1970) who proposed that diagnostic validation can be established by means of correlational analyses with other measurable variables, which can be understood as validators of the categories. They outlined five "phases" in the process of achieving diagnostic validity. Briefly, these were: clinical description, which includes personal information as well as symptoms; laboratory studies, such as biochemical, anatomical and radiological findings; delimitation from other disorders, such that one disorder can be delineated from another in such a way that each disorder would feature as a discrete entity; follow-up studies investigating response to treatment and long-term outcome; and family studies examining the prevalence of the same disorders in close relatives. It is not made clear by Robins and Guze whether or not each of these phases would contribute equally to the establishment of diagnostic validity. For example, the phase of clinical description, whilst obviously important in individual clinical encounters, seems to be of little relevance to the establishment of validity for the diagnostic category in general application to a large range of patients whose clinical descriptions will vary widely. Nevertheless, their expectation was that such analyses would demonstrate the validity of the categories and establish that they are separate categories discontinuous from each other. Following the publication of their paper, this approach has been widely adopted by researchers in the field concerned with studying the validity of the DSM-III categories.

One way of attempting to establish that the diagnostic categories represent discrete conditions is through using the validators listed in guidelines produced for DSM-5 (Kendler et al., 2009) which are grouped under three headings. These are: (1) antecedent validators including familial aggregation (i.e. the extent to which the disorder has been diagnosed among close relatives), socio-demographic factors, environmental risk factors, and previous psychiatric history; (2) concurrent validators such as psychological variables (which are independent of the diagnostic criteria) and genetic or neurological markers; and (3) predictive validators obtained from follow-up studies, in which diagnostic stability, response to treatment, and long-term outcomes are measured. The latter two groups, together with familial aggregation, have traditionally been considered in these guidelines to be the most important. Assen Jablensky, for example, notes that "psychiatrists have mainly been concerned with concurrent and predictive validity, partly because of their relevance to the issue of the validity of diagnoses" (2016, p29). To the extent that measures on one or more of these validators can be shown to correlate with a given diagnostic category, that category will be judged to have validity. Typically, for example, studies which show that a diagnostic category reliably predicts long-term outcome can be taken as showing support for the validity

of that category. Similarly, one which can be shown to have a higher probability of occurrence in close relatives of an affected individual is likely to be thought of as valid to some extent. Hence, the greater number of validators with which it correlates, the more valid it will be assumed to be.

However, despite the view that the method advocated by Robins and Guze can establish validity for the DSM categories, there have been continuing and widespread concerns about their lack of validity. For example, Kendell and Jablensky (2003) note that the validity of these has not been well established and argue that diagnostic categories should only be judged to be valid if they are seen to represent distinct entities with clear boundaries, or "zones of rarity", between them. This has been due in part to the significant degree of overlap of symptoms between different diagnoses and a high degree of comorbidity (the assignment of two or more diagnoses to the same patient) that the classification system has led to. Peter Tyrer (2014) illustrates how psychiatric diagnoses differ from other medical ones in this respect. He cites as an example three different diseases that all cause anaemia and consequently have roughly similar presenting symptoms. Nevertheless, the diagnoses of pernicious anaemia, iron-deficiency anaemia, and lymphatic leukaemia identify quite separate diseases which are identified by blood analysis. They have distinct aetiologies which allow one to conclude that there are zones of rarity between them. By contrast, the diagnosis of most psychiatric disorders is based entirely on clinical examination and the absence of any zone of rarity at the boundaries entails that the threshold for a diagnosis may be arbitrary.

The consequence of this feature of psychiatric diagnoses has been that the DSM classifications have not resulted in the discovery of the aetiology or clear causal processes that had initially been hoped for (Hyman, 2010). Similarly, Thomas Insel, former director of the US National Institute of Mental Health which has a history of supporting the DSM, has stated that the weakness of the DSM lies in its lack of validity. He states: "Unlike our definitions of ischemic heart disease, lymphoma, or AIDS, the DSM diagnoses are based on a consensus about clusters of clinical symptoms, not any objective laboratory measure" (Insel, 2013). Such diseases as these are typically diagnosed after confirmatory biomedical tests are carried out, but such tests are not relevant in decisions made about psychiatric diagnoses. For most of the diagnostic categories, therefore, validity has not been demonstrated. There is now a broad consensus that the expected validity of the DSM diagnostic categories has not been established. This indicates that the conceptual basis of the Robins and Guze approach to validation has significant limitations, with no obvious prospect that these could be easily overcome within the framework of the DSM.

# 2.4 Critique of the traditional approach to validation

The approach to validation described above works, if it works at all, on the basis that the conception of the disorder to which the category refers is fully specified by the diagnostic criteria listed for it. However, as John Campbell (2017a) argues, this approach misses out an important dimension of validity and in consequence it fails to dispel the uncertainty surrounding the validity of the diagnostic categories. The ultimate aim of establishing validity is to show as far as possible that the concept specified by the diagnostic term represents some kind of real entity in the world which is separate from the method used to identify it.

Specifically, the Robins and Guze approach assumes that there is nothing else in terms of a distinctive causal structure to the disorder beyond what the criteria specify for it. Campbell argues, however, that if we are to make progress in establishing that these categories have validity, we need to have some kind of 'governing conception' of what the disorder might be which is not described by the diagnostic criteria. Without such a conception, there can be no way of knowing whether the methods we have for identifying the existence of a disorder are actually succeeding in doing so.

Campbell's argument, concerning the lack of any kind of governing conception about what the disorder under consideration might be, means that it is difficult to establish when the correct answer about convergence on a stable category has actually been achieved, or is even close to that goal. Campbell argues that in order to know what progress we are making in doing that, we need to have a conception of the kind of entity we are trying to identify, and that this conception will make some reference, however imprecise, to the causal processes hypothesised to underlie it. In other words, the reference of the kind term needs to be fixed in some way that would allow us to assess to what extent the criteria for identifying it (the criteria specified in the manual) are successful in actually doing so. This raises the question of how fixing the reference of the term might be relevant to allowing us to assess this. The important distinction for Campbell is that between the diagnostic criteria and the condition that applying the criteria in a given case is expected to reveal. Campbell describes it thus:

In order for there to be such a thing as schizophrenia, for example, there must be the external phenomenon to which we are causally responding in using the term. There must be something "out there" we are responding to. That is the condition we are talking about. (2017a, p263).

He considers this to be a statement of "minimal realism" about the condition. The governing conception, which he refers to, of the condition concerned is "a picture or model of the kind of phenomenon that we are causally responding to when we use the term" (p263). Such a conception might be quite a sketchy model – what might be thought of as a metaphor in Richard Boyd's (1979) sense of a "theory-constitutive" metaphor – but it will suggest some kind of working model of the disorder in question. This will involve some theoretical picture or hypothesis of the causal structure underlying the disorder, which is not indicated by the criteria listed in the manual. It is on this model or conception that the reference of the diagnostic term will be fixed. In order to fix the reference of the diagnostic term, some description of the known or hypothesised causal structure is necessary. Without such a conception, it will not be clear whether the term refers to anything at all.

The distinction between the diagnostic criteria and the guiding conception of the disorder is therefore important in the fixing of reference of the terms. Referring to the DSM criteria and the Robins and Guze approach to validation, Campbell says:

If that is all we have to go on as fixing the reference of a term like "major depression", for example, then what does it take to be depressed? Is it enough if one merely meets the diagnostic criteria? Or is there some further condition that one has to have, so that one could in principle meet the diagnostic criteria and yet not in fact be depressed (just as one could have the symptoms of a viral illness yet not have the virus)? (2017a, p269).

This can be illustrated by considering the example of Alzheimer's Disease. Whilst the precise causal mechanisms constituting the disease are not yet well understood, it is generally believed that the symptoms are caused in some way by the accumulation of amyloid proteins and neurofibrillary tangles in the brain. This in broad terms represents the governing conception of Alzheimer's Disease and is what fixes the reference of the term. The diagnosis of the disease in an individual patient is based on the typical symptoms of memory impairment and personality change, together with the results of neuroimaging tests, but the reference of the term is not fixed on this cluster of diagnostic markers. Indeed, we can talk meaningfully about the disease developing in a person's brain before any of the classic symptoms become evident and before any diagnosis is actually possible. Therefore, when we talk about a disease of which we have good reason to believe in its reality, we are referring to the governing conception we have of the causal structure of that disease, not the methods we have for assessing its instantiation in any given case. However, the Robins and Guze approach to validation, which concentrates on external validators, does not deal with this

aspect of the entity which the diagnostic criteria are expected to refer to, if they refer to it at all.

One possible objection to this view is that successive editions of the DSM could be seen as representing an epistemic iteration, with the diagnostic criteria in the newer editions offering more precision thanks to epistemic progress achieved since the previous ones were published. Such an argument is offered by Hasok Chang (2017) who suggests that the periodic developments in the DSM criteria may lead to a gradual convergence on to stable diagnostic categories. An epistemic iteration of this nature would then, if successful, converge on a conception of the disorder, such as Campbell might expect, and allow us to say more about what, in causal terms, actually constitutes the disorder. Indeed Campbell (2017b) argues that it is hard to know how successful epistemic iteration might be understood, if not in terms of an increasingly accurate convergence on a particular disorder by the refinement of the diagnostic procedures. Nevertheless, Chang observes that convergence, however stable, does not guarantee that a correct answer has been obtained, or is even close to being obtained. He says:

For realists, the kind of stability afforded by an iterative process may not be considered good enough, as stability does not mean truth. This is a very serious problem that needs to be overcome, if we are to regard epistemic iteration as the model of cumulative progress in science (2017, p233).

Despite this qualification, Chang seems to be optimistic that the method of epistemic iteration that he sees in the periodic revisions of the DSM is leading to a mature and stable model for representing a valid taxonomy of mental disorders. However, Chang does not make clear what grounds, if any, there are for optimism here.

Moreover, there can be other possible explanations of the apparent stability of the DSM categories. For example, Rachel Cooper (2015) argues that these categories have become stabilised due to processes of what she calls 'path-dependence' and 'lock-in', which can occur when a previously useful technology becomes fixed despite the evidence that it has ceased to be optimal in new circumstances. Using the example of the QWERTY keyboard, which was designed for mechanical typewriters but has since been carried over to computer keyboards where it is no longer necessary for efficient functioning, Cooper argues that an analogous process has occurred with the DSM categories. The earlier editions of the DSM have established a path-dependence in which it is often easier to carry over pre-existing categories with no, or only minimal, changes into new editions, rather than creating new categories. In this way, categories can become locked in across successive editions of the

manual. In other words, the stabilisation of diagnostic categories may be equally well explained by contingent historical factors which have no necessary implications for their truth or validity.

Another, more explicit, objection to Campbell's argument is advanced by Georg Repnikov and Dominic Murphy (2017) who argue that Campbell is too restrictive in his explanation of the reference of mental disorder terms. They argue that he fails to recognise that the names of the DSM disorders refer to syndromes, understood just as the cluster of symptoms as defined in the diagnostic manual. In other words, there need be no governing conception of the kind Campbell emphasises for the reference of the term to be fixed. Repnikov and Murphy agree that psychiatrists and scientists may indeed be interested in studying the causal processes that may underlie the presenting symptoms, but argue that such conceptions need not be what fix the reference of the terms. In this respect they seem to be correct, in that the terms for medical syndromes where there is no known, or even hypothesised, explanation have as their reference the cluster of symptoms by which they are identified. Consequently, the Robins and Guze approach to validation would seem to be the only applicable method for diagnostic categories for which no common causal processes can be hypothesised. In general medicine such disorders are typically known as 'medically unexplained physical symptoms' and include conditions such as chronic fatigue syndrome and irritable bowel syndrome. Indeed, such cases as these, where there is no known unifying causal explanation, are what are generally described in medical practice as "syndromes" rather than "diseases". Even in these cases, however, a search for an explanatory causal mechanism is often felt to be desirable (Cournoyea & Kennedy, 2014), which if identified would then fix the reference of the relevant term, as Campbell would expect.

In summary, the question remains about whether the mental disorder terms as listed in the DSM are to be thought of as representing syndromes analogous to chronic fatigue syndrome or as some kind of disease entity with a causal structure which generates the reported symptoms. Insofar as the DSM categories are taken to be referring just to the clusters of symptoms defining them, any attempt at validation will be unable to answer the question of whether there is any distinct underlying condition that they represent. However, in the following section, I discuss some reasons why conceptualising DSM categories as merely representing syndromes does not seem to be consistent with the way that clinicians frequently reach their diagnoses in actual clinical practice.

# 2.5 The understanding of DSM categories in clinical practice

As described above, the nosologies in DSM-III and DSM-IV were designed in accordance with the principle that the diagnostic categories were not expected to reflect any hypothesised aetiology or causal structure. Instead, they were intended to represent the clusters of reliably correlated symptoms as constituting the disorders in question, broadly in the manner described by Repnikov and Murphy. However, although this was the intention of the manuals, the usage of the diagnostic categories in practice appears not to have conformed entirely to this conception. Instead, there is evidence that clinicians actually do make judgments or hypotheses about causation when they make diagnostic assessments, or alternatively they reify the diagnostic terms they use. In this respect, they may actually form a governing conception, in Campbell's terms, of the disorders they are dealing with.

As an illustration of this, Campbell cites evidence reported by Ahn and Kim (2008) of how clinicians actually reach their diagnostic judgments. In their studies of practising clinicians, they found that clinicians would typically invoke some hypothesised causal structure, however unrefined, in reaching their diagnostic conclusions. In other words, their diagnostic assessments were based either on additional considerations to those listed as the criteria in the DSM or on selective weighting of some specific criteria hypothesised to have some causal relevance. Thus, for example, in DSM-IV anorexia nervosa should be diagnosed when all four listed criteria are met, these being in brief: 1) fear of being fat, 2) distorted body image, 3) refusal to maintain body weight, and 4) absence (in females) of a period for three or more months. These criteria are all given equal weight in the DSM. However, Ahn and Kim report that the feature of distorted body image was treated as a central causal element in the clinicians' theories about the disorder, whereas the absence of a period for three months was treated as causally peripheral. Similarly, distorted body image was considered to be of most diagnostic importance and absence of a period of least importance. Yet this kind of weighting of diagnostic criteria is not consistent with strict adherence to the manual's guidelines.

The use of some kind of governing conception of a disorder, whether conscious or unconscious, such as that of distorted body image in anorexia nervosa, can lead to such a conception becoming reified as representing an underlying pathological condition. This is described by Steven Hyman (2010) as the "problem of reification" of the DSM categories. Hyman refers to the tendency described by John Stuart Mill as the "personification of abstracts" which Mill explains as follows: "Mankind in all ages have had a strong propensity to conclude that wherever there is a name, there must be a distinguishable separate entity

corresponding to the name" (1843, p756). Specifically, Hyman analyses how the nature of classification in the DSM can lead to the disorder categories being treated in clinical practice, both by clinicians and by patients, as representing some kind of underlying pathology, such as might be understood as constituting a disease of some sort. Hyman notes that, whereas the disorders in the DSM have proved to be heuristics which are useful in various ways, especially in improving inter-rater reliability, they have become reified in frequent practice. He states: "Disorders within the DSM-IV... are often treated as if they were natural kinds, real entities that exist independently of any particular rater" (2010, p156). This stems from the manner in which disorders are categorised in the manuals, which explicitly eschew any kind of dimensional approach to conceptualising mental disorder. The treatment of the diagnostic terms as representing discrete categories, in the way in which they are explicitly described, encourages clinicians to think of them as constituting entities of some kind which have an existence beyond the defining clusters of symptoms. As Hyman also comments: "... cautionary statements within the DSM-IV, if read at all, provide little protection among many communities of users against reification of the disorders listed within" (2010, p158). In this way, the categories have tended to become reified in much clinical use.

It might be objected that clinicians who think of the categories in DSM-III and DSM-IV in this way are simply misunderstanding what their purpose is. Since the categories are not intended to reflect any kind of underlying entity or aetiology, any clinician who consciously or automatically construes them in this way could just be said to be making a mistake. Firstly, in answer to this, one could note that if well trained clinicians are regularly making such mistakes, then this would suggest that there is some design flaw in the DSM which predisposes them to err in such a way. The manuals are designed to be, amongst other things, tools for clinicians to make usable diagnostic judgments. If it is the case that clinicians have difficulty doing this in the way that the manual specifies, then it would seem that the manuals are not doing the job they were designed for. Secondly and more specifically, the Robins and Guze approach to validation assumes that there are discontinuities between mental disorder categories and this assumption is reflected in the construction of the DSM manuals. In other words, there are expected to be zones of rarity forming boundaries between the disorders (Kendell & Jablensky, 2003) – the conditions described by the categories appear to be discrete. Therefore, there seems to be some ambiguity in how the categories are to be interpreted. An alternative approach to describing mental disorders would involve conceptualising them in dimensional or quantitative terms, in a manner similar to the psychometric measurement of personality. Such an approach might reduce the tendency

towards reification of mental disorders, but this is a quite different conceptualisation of mental disorders from that contained in the DSM.

Furthermore, the objection above does not seem to be applicable to the latest edition of the manual, DSM-5 (APA, 2013). This edition explicitly changes the conceptual framework of mental disorder on which the listed diagnoses are said to be based. Specifically, the descriptive and atheoretical approach which guided the formulation of the diagnostic categories in DSM-III and DSM-IV has been abandoned in DSM-5.<sup>23</sup> The intention in the new edition is that the classification be shaped by theories about aetiology as far as is possible given the current state of knowledge about them. The original aim was to create a 'meta-structure' in which the various disorders would be organised into five clusters representing different aetiological risk factors. Whilst this could not be achieved due to insufficient empirical support, the manner in which the categories were ordered is nevertheless intended to indicate some information about hypothesised aetiology (Cooper, 2018). This is made explicit in the introduction to the manual.

The proposed organization of the chapters of DSM-5, after the neurodevelopment disorders, is based on groups of internalizing (emotional and somatic) disorders, externalizing disorders, neurocognitive disorders, and other disorders. It is hoped that this organization will encourage further study of underlying pathophysiological processes (APA, 2013, p13).

Cooper notes that, while many of the changes from DSM-IV are subtle and may be easily overlooked by clinicians, the result nevertheless represents a significant reconceptualization of the DSM: "Once, the classification set out to be descriptive and atheoretical; now, it seeks to reflect theoretical knowledge" (2018, p59).

Given this reconceptualization in DSM-5, the reification of categories by clinicians would no longer seem to be a result of a misunderstanding of what the categories are supposed to represent. If clinicians do tend to reify them in their thinking, this would not seem to be inconsistent with the latest conceptualisation in the DSM. These categories, it would seem, are explicitly intended to represent as far as possible some kind of underlying entity, however imperfectly this may be achievable in the current state of knowledge. Of course, the hope that the DSM classifications might turn out to have distinct underlying

<sup>&</sup>lt;sup>23</sup> To be precise, the descriptive approach was explicitly stated as a basis for DSM-III, but this was not made explicit in DSM-IV. However, the structure of the taxonomy and the contents were very similar to those in DSM-III and consequently it could be assumed that the descriptive approach still underpinned these (Cooper, 2018).

pathophysiological processes showing them to be discrete entities of some kind does not necessarily entail that they must be natural kinds.

In summary, many clinicians seem to have been treating the DSM categories, either consciously or unconsciously, as if they do represent some underlying condition beyond what the diagnostic criteria explicitly specify, and the reconceptualization in DSM-5 appears to support this to some extent. Nevertheless and despite this, there has been no significant progress in validating the categories.

# 2.6 Explanatory deficiencies of DSM categories

The issues discussed above concerning the approach taken to validation of the DSM categories and the tendency towards reifying them, reinforced apparently by the reconceptualization of mental disorders in DSM-5, have implications for whether they can be regarded as natural kinds. If the categories, or at least some of them, had been shown to be valid according to the Robins and Guze methods, this would have strengthened the argument that they could be natural kinds. Specifically, if they were shown to be predictive of various outcomes, such as treatment response and course of the disorder, there would then be some grounds to infer that they might constitute, or at least approximate to, natural kinds. However, this would still leave open the question of whether they performed any explanatory function, which would be in accord with the conceptualisation of natural kinds proposed by Boyd and Khalidi, as described in Chapter 1. In order to do this, there would need to be some information about the aetiology of the disorders represented by the diagnostic categories, such that the various disorders could be distinguished from each other on this basis. It is only on this basis that we can expect, as Campbell (2017a) argues, that the categories represent "something out there" and that the entities represented are likely to be discrete ones. For somatic diseases in general, this condition is met – e.g. the difference in aetiology between viral and bacterial infections serves to discriminate between these types of diseases, even when the presenting symptoms appear very similar. However, for most of the disorders listed in the DSM there is no such aetiological information available. Against this, there seem to be two alternative perspectives, proposed by Maung and Murphy respectively, which argue that diagnoses can make an explanatory contribution. I now discuss these in turn.

## 2.6.1 Maung's arguments

It can be objected that, despite the lack of aetiological information available, the DSM categories can nevertheless have some explanatory value. This claim is made by Hane Maung (2016) who argues that psychiatric diagnoses may yield some clinically relevant causal information. Maung agrees that many diagnostic categories are causally heterogeneous, in a manner that does not reflect the clearer causal information typically indicated by diagnoses in general medicine. Despite this, he argues that psychiatric diagnoses are not entirely devoid of causal information. He describes three ways in which such information may be revealed, and I will discuss each of these in turn.

Firstly, Maung argues that diagnoses may provide some *negative* causal information. Specifically, he argues that a psychiatric diagnosis can, and often does, exclude other sorts of diagnosis, particularly somatic ones, and in this way does therefore provide some explanatory relevant information. Thus, for example, a diagnosis of major depressive disorder (MDD) can exclude other diagnoses, such as thyroid or adrenal disorders, for which the causal processes are clear, but which can produce symptoms similar to those of MDD. However, these other diagnoses can only be excluded when the relevant biomedical or biochemical tests are carried out. Maung cites psychiatric textbook sources to the effect that such tests should be carried out, but in busy and over-stretched mental health services this does not always happen before the diagnosis is made. Even when they are, the diagnosis of MDD itself does not exclude such biochemical aetiologies as causes of the patient's symptoms, since it is actually the tests carried out which do this. Therefore, it is not clear what extra explanatory information can be provided by a diagnosis which does not itself cite a specific cause, beyond that provided by any biomedical tests. Maung refers to David Lewis's account of causal explanation, which he sees as supporting his argument, and quotes Lewis who says: "to explain an event is to provide some information about its causal history" (1986, p217). However, the diagnosis of MDD on its own does not seem to provide any such information, and this conclusion can be extended to other psychiatric diagnoses to the extent that they also fail to do the same.

Moreover, it is not clear how useful any such information is in the absence of a clear causal explanation. To be informed in any situation, where there is a problem that needs diagnosing, that the person attempting the diagnosis can only say what is not causing the problem is not normally seen as providing any kind of explanation. This does not seem to change no matter how many potential explanations are excluded, as long as no actual positive explanation is available. It would seem the best that could be said is that the correct explanation is one among a finite number of specified possible ones. This could perhaps be

useful in some circumstances, but only as an intermediate step in leading to the actual explanation.

Secondly, Maung argues that there can be disjunctive causal explanations. Thus, to take the example of MDD again, it may be the case in a given patient with depressive symptoms that a disjunction of several biological or psychological states,  $P_1$ ,  $P_2$ ,  $P_3$ ...  $P_n$ , resulting from different combinations of variables can be under consideration as possible causes. Maung argues that such a disjunction can provide causally relevant information, in the absence of knowledge of any more specific causal process. Again however, this raises the question about whether this is always the case and, when it is, what part of the disjunction is actually providing the causal information. Maung cites the example, taken from Kim (1998), of a patient with joint pain which could be caused by either rheumatoid arthritis (RA) or systemic lupus (SL). Kim argues that this disjunction of possible causes fails to give us an explanation of the patient's symptoms, because as a disjunction no specific cause is identified. Maung, however, claims that such a disjunction does give us causal information. In this example, both conditions are types of multisystem autoimmune disease and the disjunction does, therefore, give us the information that the patient suffers from this kind of disease. However, this argument overlooks the question of how it could have been established that the patient suffers from one or other type of autoimmune disease. The process by which this has been established, e.g. by thyroid function test or past medical history, is relevant here. It is not the disjunction itself which reveals this information. In this respect, the answer to Maung is similar to that given to his first argument above. In another case when we only know that patient's symptoms are caused by disease  $P_1$  or disease  $P_2$ , we have no other information available, unless some other information is given to us. It is difficult therefore to see how such a disjunction can yield any explanatory information. Maung's argument here seems to be the converse of his argument above that diagnoses can provide negative causal explanation and is open to the same kind of objection.

Thirdly, Maung argues that psychiatric diagnoses can provide explanatory information through the *causal networks of symptoms* with which they might be associated. He refers specifically to the work of Denny Borsboom and colleagues (e.g. Borsboom, 2017; Borsboom & Cramer, 2013) who have developed a network model of mental disorders. According to this approach, which Borsboom explains is an alternative to the nosological system in the diagnostic manuals, causal relationships can be hypothesised to exist between different symptoms of mental disorder, such that symptom clusters can be identified in which the different symptoms reinforce each other. In the case of MDD for example, it may be the

case that fatigue may lead to other symptoms, such as low mood and anxiety. These may then lead to sleepless nights, which can then reinforce the experience of fatigue. However, the clusters of symptoms and the causal networks in which they are embedded may not necessarily map directly on to existing categories in the DSM, and indeed Borsboom (2017) makes clear that the network model is a radically different approach. Moreover, there may be no more information about the causal history or aetiology of the symptom clusters than there is for DSM diagnoses, and hence they may be equally limited in terms of any causal information they may offer.

Despite the lack of any aetiological information it may provide, the network model nevertheless does offer other useful information, as Maung describes. Specifically, the networks in which individual symptoms are located do seem to offer some predictive value in terms of predicting what other symptoms the patient may be likely to experience, either concurrently or in the future. However, this does not in itself add much by way of explaining the causal processes leading to the mental disorder in the first place.

# 2.6.2 Murphy's "exemplar" argument

Another way in which correlated clusters of symptoms might have some explanatory value is proposed by Dominic Murphy (2010, 2014). Murphy agrees that most mental disorders do not have clear causal histories, in the way that Huntington's disease, for example, has a specific genetic cause. Instead, he proposes that mental disorders can be understood as groups of symptoms that cluster together reliably in such a way that causal connections can be assumed to hold between them.<sup>24</sup> He takes these, or at least some of them, to be instances of homeostatic property clusters, as described by Richard Boyd in his account of natural kinds. Individual manifestations of a specific disorder may vary from patient to patient, but these will tend to resemble each other to a greater or lesser extent. Murphy argues that these varying patterns of presentation of a disorder can be understood in terms of their relationship to what he terms an "exemplar" of the disorder. On his view, textbook diagnoses should be thought of as referring to idealizations or exemplars of the disorder concerned. He says: "We may think of exemplars as representing the ideal, textbook patient with a particular condition,

\_

<sup>&</sup>lt;sup>24</sup> Murphy's suggestion of symptom clusters might appear to suggest something similar to Borsboom's conception of clusters in a causally connected network of symptoms. However, although it is difficult to be clear exactly how he conceives of these clusters, Murphy seems to envisage clusters of symptoms constituting distinct diagnostic categories in a nosology of disorders which would resemble the DSM to some degree, if not actually corresponding with it. In contrast, Borsboom is quite clear that his conception is a fundamentally different approach to that of the DSM nosology.

even though such an ideal patient may never in fact enter the clinic" (2014, p105). Actual patients, therefore, resemble the exemplar to some degree, and it is this resemblance that warrants the diagnosis and allows for explanations of the patient's condition to be generated. Murphy explains further:

The bet is that real patients will be similar to the exemplar in enough respects so that the explanation of the exemplar carries over to the patient. We assume that within the individual there are phenomena and causal relations that are relevantly similar to those worked out for the exemplar (2014, p106).

The exemplar is therefore expected to encompass what is known about the causal processes underlying the stated disorder and these, or at least a significant proportion of them, are assumed to be operative in the individual case.

However, there seem to be two difficulties associated with this picture. Firstly, as Murphy indicates, this is based on assumptions about the similarity of causal relations between the exemplar and the real case. It is not entirely clear how to understand his use of the word "bet" in the quote above, but it may be intended to reinforce this assumption, while also suggesting that it may be a matter of hope, combined with some degree of confidence, that real patients will actually be similar to the exemplar in relevant respects. However, for most of the DSM diagnoses, we do not have a clear understanding of the causal processes which explain them, other than the belief that they are causally highly heterogeneous. Therefore the "bet" that real patients will be similar to the exemplar does not help us achieve any explanation of the patient's condition. This is in contrast to many physical diseases, where Murphy's conception of the role of the exemplar does seem plausible. For example, we have a good exemplar of the typical symptoms of influenza, and the similarity between this and the instantiation of it in real patients with the disease is important in explaining how the patient has become ill. This is despite the fact that the severity and course of the illness vary considerably from patient to patient, depending on such factors as the age and pre-morbid health of the patient and the strain of the virus concerned.

Secondly, it is not clear how Murphy thinks a suitable exemplar for a mental disorder can be constructed in the case of disorders where the diagnosis depends upon polythetic criteria.<sup>25</sup> A good example of this is schizophrenia, the five principal symptoms of which are listed in DSM-5 as follows (APA, 2013):

\_

<sup>&</sup>lt;sup>25</sup> Polythetic criteria for membership of a group refer to the sharing of a number of characteristics which occur commonly in members of a group, but where no single one of which is essential for membership of that group.

- 1. Hallucinations
- 2. Delusions
- 3. Disorganised speech
- 4. Grossly abnormal psychomotor behaviour, e.g. catatonia
- 5. Negative symptoms, e.g. restricted affect, avolition, asociality.

In order to receive a diagnosis of schizophrenia, the patient must display at least two out of the above symptoms, at least one of which must be from the first three listed.<sup>26</sup> The criteria for diagnosis are polythetic, in that two patients might each fulfil the criteria and therefore receive the diagnosis while displaying entirely different symptoms from the list. According to Murphy, we can construct an exemplar of schizophrenia which will contribute to our understanding of individual patients who present with the condition. However, he gives no indication of how we might do this in such a case. Would the exemplar consist of just two of the specified criteria? If so, on what basis would these two be chosen, and how would they contribute to our understanding of the presentation of the other symptoms in different patients? Alternatively, is the exemplar to be understood as encompassing all five symptom types specified by the criteria? Although Murphy does not give a clear answer to these questions, one might interpret him as favouring the latter possibility. Patients seldom present with all five sets of symptoms, but it is conceivable that an exemplar constructed in this way might constitute an advance in our ability to explain the condition in individual patients displaying a sub-set of symptoms, assuming of course (as argued above) that we do already have some information about the causal processes underlying the symptoms. However, basing the exemplar on all five symptom types in this way seems to be nothing more than a restatement of the diagnostic criteria contained in the DSM, and as such it is difficult to see how it can add anything new to our understanding of the condition, nor even clarify whether there is actually a discrete condition underlying the symptoms that meet the diagnostic criteria.

Moreover, it is not clear how any similarity between the exemplar and real individual cases is to be established without some pre-existing conception of the kind under which both are to be classified. As Quine explains in his paper 'Natural Kinds':

<sup>&</sup>lt;sup>26</sup> There are additional criteria which specify exclusions, e.g. that the symptoms must not be caused by extraneous factors such as drug or alcohol abuse. There is a further criterion that the symptoms should have been present for at least six months before the diagnosis can be made. These do not affect the argument here and consequently I will not discuss them further.

The notion of a kind and the notion of similarity or resemblance seem to be variants of a single notion. Similarity is immediately definable in terms of kind; for things are similar when they are two of a kind (1969, p7).

Consequently, any judgment of similarity must already imply that that the two items concerned are examples of the same kind. Therefore, it is not clear how the specification of an exemplar, which as argued above may be difficult to specify in any non-arbitrary manner, can explain the presentation of the patient's symptoms, because the judgment of similarity is based on the implicit assumption that the two are of the same kind. Whether or not the patient's symptoms are actually explained in this way cannot be assumed without some other kind of assessment or investigation.

I have discussed the question of whether the DSM categories can have any explanatory value without some additional information about their aetiologies, which, with a few exceptions, is not generally given by their diagnostic criteria. In conclusion, it seems that they do not. Any apparent explanatory information they may convey can be attributed to other sources, such as the results of biomedical investigations that may have been carried out during the diagnostic work-up on an individual patient. This does not exclude the possibility that some of them may constitute natural kinds. However, the conceptions of natural kinds described by Boyd and Khalidi require that they have explanatory significance by reflecting the causal structure of the world. Boyd also emphasises that natural kinds are established a posteriori. Therefore, in the absence of any a posteriori knowledge of causal explanatory information associated with a diagnosis, we cannot assume that they constitute natural kinds. This can be seen in the case of schizophrenia in particular, which I discuss further in the following section.

# 2.7 The natural kind status of schizophrenia

As noted above, the diagnostic criteria for schizophrenia are polythetic. One consequence of this is that there cannot be a single essential feature which all cases of the condition must display, but this would only disqualify it as a natural kind on an essentialist conception. However, most recent philosophers who claim it can be a natural kind (e.g. Murphy, 2014) argue that a property cluster account, such as Boyd's, is the most appropriate conception of natural kinds against which to assess it.

A key element in Boyd's account is the expectation that natural kinds will have causal functions in the theories we construct about the world. In this respect, therefore, his account

goes beyond the characterisation of natural kinds merely as homeostatic property clusters which Murphy and others tend to rely on in their arguments. Specifically, Boyd describes the "fit or accommodation between natural kind categories and induction-supporting causal powers of things" (1999, p69) as an important element in his account. Any claim that schizophrenia (or any other condition) can be a natural kind on this account would therefore have to be supported by an explanation of how it can provide for such an accommodation with the causal powers of things. In other words, it would need to show how the category illuminated relevant causal processes in the world.

There would seem to be several requirements to be fulfilled if schizophrenia is to constitute a natural kind. Firstly, if it is to function as a kind with causal properties distinct from other kinds of disorder, it ought to be a discrete disorder with relatively clear boundaries, or a zone of rarity, between it and other disorders. This is expressed, for example, in an editorial in the British Medical Journal by two leading psychiatrists, Jeffery Lieberman and Michael First, who say: "the charge that schizophrenia does not define a specific illness is clearly unwarranted" (2007, 334: 108). Similarly, Regier et al (2009) discussing the revision of the DSM prior to the publication of DSM-5 say: "Mental disorder syndromes will eventually be redefined to reflect more useful diagnostic categories... as well as dimensional discontinuities between disorders and clear thresholds between pathology and normality" (2009, p648). Despite these expectations however, there appears to be no clear boundary between schizophrenia and other psychotic disorders, either in DSM-5 or in earlier editions. Amongst other psychotic disorders in DSM-5, "bipolar disorder 1" is defined as a cyclical pattern of severe depression followed by mania.<sup>27</sup> This is broadly similar to the previous category in DSM-IV of "bipolar disorder" (which, for convenience, is the term I continue to use here). However, when experiencing episodes of mania, the patient may also display some of the symptoms of schizophrenia, i.e. hallucinations, delusions, and disorganised thoughts. Many patients in fact experience symptoms of both schizophrenia and bipolar disorder. Recognising this, the diagnosis of schizoaffective disorder was introduced in the 1930s by Joseph Kasanin (1933) for patients displaying a combination of symptoms characteristic of both conditions.

More recently, other psychiatrists (e.g. Kendell, 1991) have argued that these disorders are better understood as existing on a continuum where the majority of patients fall

\_

<sup>&</sup>lt;sup>27</sup> In DSM-5, an additional category of "bipolar disorder 2" was introduced to take account of cases where the patient has episodes of hypomania, rather than full-blown manic episodes.

somewhere close to the middle with fewer presenting as textbook cases of schizophrenia or bipolar disorder (Bentall, 2017). For example, in a large study of patients with psychotic disorders, Keshavan and colleagues rated 762 participants on a scale measuring the schizophrenia-bipolar continuum. They found no evidence for a clear zone of rarity between the two end-point diagnoses, with almost 50% of the participants falling somewhere on the continuum (Keshavan et al., 2011). Other studies have also produced similar results (Bentall, 2017). Therefore, it seems doubtful that schizophrenia, as defined in DSM-5, is a discrete disorder, and for the same reasons this also seems to be the case for bipolar disorder.

Secondly, if schizophrenia constitutes a natural kind, it would be expected to have demonstrated validity in the manner described by Robins and Guze (1970) in their discussion of how this can be established. Although their proposals for establishing validity were intended to apply to diagnostic categories generally, they were framed specifically for that of schizophrenia. Predictive validity is regarded as one of the most important components of validity (e.g. Kendler et al, 2009) and this is one that poses difficulties for the concept of schizophrenia. It seems to have limited predictive value regarding the outcomes for people with the diagnosis. Traditionally schizophrenia was regarded as a lifelong condition from which most patients could not expect ever to recover completely.<sup>28</sup> However, a very different picture emerges from a large multi-centre international study which looked at long-term prognosis over a 25-year period and found recovery rates of around 50%, but with significant variation across different cultures (Harrison et al., 2001). Recovery rates were markedly better in under-developed countries than in advanced industrialised societies. Other research has also challenged the traditional view and revealed widely varying outcomes, with studies reporting anything from 13% to 72% recovery rates, depending upon definition as well as many psycho-social factors, including geographical location (Read, 2013). Thus, the diagnosis does not predict longer term prognosis with any precision; for those diagnosed with

\_

<sup>&</sup>lt;sup>28</sup> This is a slight over-simplification, since schizophrenia has been seen for a long time as encompassing a group of conditions. According to a classic psychiatry textbook, it was regarded as "a group of mental illnesses characterised by *specific psychological symptoms* and leading, in the majority of cases, to a disorganization of the personality of the patient" (Slater & Roth, 1977, p237, italics in original). Whilst it was recognised that not all conditions covered by the term would have the same outcome, Slater and Roth state that most patients would be expected to show an overall deteriorating course over time, even if interrupted by periods of remission of symptoms. They say: "It is therefore difficult to generalize about the course and prognosis, although it remains true that every condition grouped under the term schizophrenia is associated with a general *tendency towards disintegration of the personality....* From our point of view, the tendency towards an unfavourable outcome is a general attribute of the condition, or group of conditions, which we call schizophrenia" (1977, p308, italics in original). Moreover, as this quote indicates, the "group of conditions" was nevertheless understood as being primarily a subset of one single condition called schizophrenia.

schizophrenia it seems there is no common prognosis. The course of the condition is an openended process that can be modified in many ways (Bentall, 2009; Jablensky, 2016; Read, 2013).

Moreover, the treatment that is actually given to the patient is not strongly predicted by the diagnosis itself any more than it is predicted by other diagnoses, since this is dependent on many aspects of the patient's condition. It is a feature mental of health services that only a limited range of treatment options are available. It used to be the case that antipsychotic drugs were given to patients with a schizophrenia diagnosis, but not bipolar disorder. More recently, however, they have been judged to be equally appropriate for those with bipolar disorder, as well as other conditions such as major depression and anxiety disorders (Stein, 2014). If the same drugs work for many different conditions, it is difficult to see how the specific diagnosis, rather than the facts about the individual patient's condition, can have strong implications for the specific treatment chosen. Taken together, the variable outcomes for the diagnosis and the unspecific nature of treatment prescription would seem to imply that the predictive value of the schizophrenia diagnosis is not strong.

Thirdly, as discussed in connection with the DSM categories generally, if schizophrenia is to constitute a natural kind, it would be expected to function in some manner in explaining the patient's symptoms. However, the diagnosis of schizophrenia does not do this. Despite a large amount of research over many years aimed at establishing the aetiology of the condition, no generalisable causal relationships have been uncovered. There has been a long search for distinctive changes in the brain associated with schizophrenia to support the belief that it can be explained this way. However, no consistent and replicable evidence has been found to show that any such changes are causative, rather than the effects of other factors (Murray, 2017). Whilst some changes in cortical volume and lateral ventricular volume have been observed, these have been shown to be largely associated with long-term use of antipsychotic medication (Ho et al., 2003; Zipursky et al., 2013). Moreover, whilst subtle changes have been noted at the onset of schizophrenia (Murray, 2017), there is good evidence that such changes are likely to be caused by earlier experiences. For example, abnormal brain structure and function have been detected in people with experiences of traumatic events, including childhood abuse (Read et al., 2001), and such experiences have been shown to be significant risk factors for later development of psychotic symptoms and a diagnosis of schizophrenia (Bentall, 2017; Sitko et al., 2014). Moreover, children who experience multiple kinds of trauma coupled with other adverse childhood experiences, such as being raised in poverty, are at even greater risk (Bentall, 2017). Therefore, the onset of psychotic symptoms

and the slight changes in the brain that may correlate with that are likely to be different effects of similar causal histories.

There has also been a strong belief that schizophrenia is a predominantly genetic disease (Lieberman & First, 2007). The main support for this view comes from twin, family, and adoption studies which aim to demonstrate a greater risk of schizophrenia in individuals who are genetically related to someone known to have the diagnosis. Twin studies have typically been considered to be particularly strong indicators of a genetic basis for the condition. These compare the incidence of schizophrenia between sets of monozygotic (MZ) and dizygotic (DZ) twins and have found a higher risk of the condition in the MZ twin of an affected individual than for the DZ twin of someone affected. Family and adoption studies have also shown higher concordance rates for those with closer genetic ties than for people with more distant or no genetic relationships. Nevertheless, there are disagreements about the heritability estimate that can be calculated from these studies, and it is actually the case that around 75% of the identical (MZ) twins of people with a schizophrenia diagnosis will not be diagnosed with the condition themselves (Bentall, 2017). Moreover, such risk as does exist is not confined to a diagnosis of schizophrenia. In a study of the medical histories of the entire Swedish population, Lichtenstein and co-workers (2009) found that people at higher risk of being diagnosed with schizophrenia due to having a first-degree relative with the diagnosis were also at higher risk of having a diagnosis of bipolar disorder. The reverse was also the case for people with a relative diagnosed with bipolar disorder. Therefore, the risk attached to having a relative with a diagnosis of schizophrenia is not a risk specifically for having the same condition.

In addition, genome-wide association studies have revealed no single gene or set of genes that can explain the onset of the condition. Rather, it seems more likely that a large number of small and additive genetic features contribute to the risk for developing psychotic disorders, with no specific risk either for schizophrenia or for bipolar disorder. Moreover, Kenneth Kendler (2015) notes that the genetic risk for schizophrenia is very widely spread in the population, to the extent that everyone carries some level of risk. The conclusion from recent genetic research therefore seems to be that there is no clear genetic cause for schizophrenia or bipolar disorder (Bentall, 2017).

In conclusion, the lack of a clear boundary between schizophrenia and other conditions, the limited predictive validity of the diagnosis, and the paucity of explanatory information conveyed by the diagnosis imply that it is difficult to see how it can constitute a natural kind on either Boyd's or Khalidi's account. In this respect, it clearly differs from many somatic

diseases which can plausibly regarded as natural kinds on either of these accounts, as argued in Chapter 1.

#### 2.8 Conclusions

The lack of validity of most of the DSM diagnostic categories raises the question of what they are expected to represent. It seems to be the case in DSM-5 that they are expected to represent something, given the apparent aim that they should indicate some kind of hypothesised aetiology (Cooper, 2018), even if it is not possible in most cases to say in any kind of detail what this 'something' is. Consequently, it would seem reasonable to suppose that if a diagnostic category does stand for some distinct, if currently unknown, aetiology, then it could be thought of a disease of some kind, in view of the large range of types of disease to which humans are susceptible. In this respect therefore, the current conceptualisation on which the DSM-5 is based would seem to be consistent with a view of mental disorders as being in some way analogous to somatic diseases. Nevertheless, as long as there is no clear and replicable evidence that most of the DSM categories constitute natural kinds, the plausibility of such a view will remain in doubt.

In this chapter, I have argued that DSM categories in general are implausible candidates for natural kinds. This is in contrast to my claim in Chapter 1 that many somatic disease entities can constitute natural kinds. There is reason, therefore, to think that the conditions denoted by the DSM categories are not analogous to somatic diseases. However, it has been argued by various psychiatrists that there is no clear distinction to be made between mental disorders and diseases in general medicine (e.g. Kendell, 2001). I discuss this question in the next chapter, where I review what is generally referred to as the *biomedical model* of mental disorders. I will also argue that this is still a dominant model in western psychiatry, and that (*pace* Kendell) a significant distinction can be made between mental disorders and somatic diseases.

### Chapter 3

# The Biomedical Model in Psychiatry

#### 3.1 Introduction

The aim of this thesis is to examine some of the narrative consequences of receiving a psychiatric diagnosis on those affected. A frequent assumption behind a psychiatric diagnosis is that the recipient has "a disease like any other", and recipients have often been encouraged to think this. This assumption is supported by the biomedical or disease model which is predominant in psychiatry.

In this chapter, I discuss what the biomedical model is, how it is typically understood in relation to mental disorders, and how it remains a dominant model in contemporary western psychiatry. I shall use the terms 'biomedical model' and 'disease model' interchangeably here, as both have been employed in the literature and generally with a similar meaning (in subsequent chapters, however, I shall mainly use 'biomedical model'). It should be noted that the biomedical model is also used in relation to somatic medicine as well as psychiatry,<sup>29</sup> and, therefore, in order to understand its role in psychiatry it is necessary to appreciate its role in medicine more generally.

The discussion of models can be confusing, because different terms tend to be used in the literature – e.g. disease model, biomedical model, medical model – and different writers use them with varying meanings. The term 'medical model' in particular is frequently used and tends to have a wider meaning, whereas 'biomedical model' and 'disease model' tend to be used interchangeably. This raises the question of what exactly is meant by the term 'model' in this context and consequently, at the beginning of section 3.2, I provide a brief summary of what this typically means. I also make some comments about the varying ways in which the term 'medical model' is used in the literature to demonstrate the kind of confusion that it can potentially give rise to. I then, in 3.2.1, discuss how the biomedical model is frequently conceptualised and contrast this with the more general term 'medical model'. These terms can also sometimes be confused and I attempt to clarify them as far as possible. I outline the main aspects of the biomedical model which seem to characterise it in

<sup>&</sup>lt;sup>29</sup> Somatic medicine refers to all areas of biological and physiological medicine, apart from psychiatry.

the context of psychiatry. In brief, this model assumes that mental disorders can be considered to be analogous to diseases in some sense. I also explain the reasons why the biomedical model is a dominant one in psychiatry.

Since the conception of mental disorders as being in some way analogous to diseases remains a dominant view, this raises the question of what sorts of things we think diseases are. I therefore, in section 3.3, discuss different conceptions of what constitutes a disease. In section 3.3.1, I review some accounts of the nature and ontology of diseases in the literature and conclude that a hybrid account that incorporates naturalist and normative features is more plausible than one that relies solely on normativism or naturalism for its defining features. I argue that the concept of disease requires at least a naturalist element that specifies some kind of distinct causal process as its defining feature. In 3.3.2, I describe Jeremy Simon's (2008) constructive realist account of how the causal basis of diseases can be understood and argue that it offers a plausible way of understanding how regular causal processes can operate in specific diseases when other idiosyncratic causal factors are present in individual cases. Simon uses cystic fibrosis as an example to illustrate this.

I conclude the chapter in section 3.4 by commenting briefly on the effects that a psychiatric diagnosis can have on the self-narrative of the recipient, due to the influence of the biomedical model in typical conceptions of mental disorders.

### 3.2 Models of mental disorder

Discussion of models raises the question of what exactly is meant by the word 'model'. There seems to be no precise definition of this, and most writers about this subject tend to take it for granted that the meaning is understood. An exception to this is Ahmed Samei Huda (2021). He characterises this as comprising two 'mutually influencing' senses of model, which he calls 'models of practice' and 'models of explanation' respectively. As he describes them, models of practice refer to the systems by which physicians assess their patients and prescribe treatment for them. Models of explanation refer to the nature of the disorders themselves and the causal factors that led to them. While Huda notes that these models are connected, most references to models in the context of mental health services are to models of explanation. In the latter sense, these models can be broadly understood as frameworks for conceptualising the ontology of mental disorders and the kinds of causal processes that are expected to underlie them.

The terms 'medical model', 'disease model', and 'biomedical model' are often used loosely when applied to psychiatry. In particular, the 'medical model' is a frequently used term in this context, both by its critics and those who endorse it. Nevertheless, there is considerable variability in how this term is used which can lead to some confusion in how it is to be understood. Reflecting the distinction between models of explanation and models of practice noted above, it can refer either to an epistemic paradigm for understanding the causes of human psychological difficulties and grounding inductions on this basis or, alternatively, it can refer more generally to a set of treatment and care practices to be applied within mental health services. It is the latter sense in which it is used and criticised by the psychiatrist R.D. Laing (1969), who understands it as representing a fairly linear process whereby the doctor starts by taking note of the patient's complaints, proceeding through a relatively ordered process of further investigations, making a diagnosis, and ending by prescribing a specific treatment.<sup>30</sup> More recently, Huda (2019, 2021) has described the medical model in a similar manner, although in contrast to Laing he endorses it as the appropriate basis for mental health care services. A more explicitly sociological conception of the medical model was described by Erving Goffman (1961) who regarded it as a particular way of managing people with severe mental disorders, involving incarceration in hospital and domination by medical authority. In contrast, Anthony Clare (1976) regards it as little more than the application to medicine of such typical scientific methods as observation, description, and differentiation, such as might be employed in zoology for example. On his view it represents a broad description of the process by which pathological symptoms will be observed over a period of time with gradual progress being made towards individuation of distinct conditions by aetiology and likely outcome.

Dominic Murphy also uses the term 'medical model', although in a way which is very similar to the conception of the biomedical or disease model which I describe below. Unlike the many critics of the model, he endorses it as an appropriate basis for understanding mental disorders and indeed he argues for a strong version of it. Thus, he conceives of a mental disorder as "a breakdown or suboptimal deviation in normal functioning due to a pathogenic process unfolding in some bodily system" (2013, p. 968). In our attempts to understand how the cognitive processes associated with a psychotic episode should be explained, he states:

-

<sup>&</sup>lt;sup>30</sup> R.D. Laing (1927-1989) was a well-known British psychiatrist who became known for his opposition to conventional methods of understanding and treating mental disorders in psychiatry. He was one of the founders of the 'anti-psychiatry' movement in the 1960s and 1970s.

To fit in with the logic of the medical model in its strong guise ...such processes would need to have, among their effects, a realization of a destructive disease process in the brain. The ensuing neuropathology is just what the disease amounts to (ibid, p. 970). This view, I argue, represents a clear statement of how the biomedical model of mental disorders might be typically understood in the context of psychosis. Murphy's reference to "a destructive disease process in the brain" implies a reductionist conception of the psychotic symptoms, which reflects the reductionism that is a central feature of the disease model in general (Anderson, 2017) and which I discuss further below.

What these different interpretations of the 'medical model' indicate is the inherent ambiguity in the term and therefore the potential for it to generate misunderstandings. I argue here that the term 'biomedical model' is clearer in that it focusses more specifically on the epistemic implications of what people mean when they refer to the medical model and its intended explanatory function.

#### 3.2.1 The biomedical/disease model in medicine

I've noted that the terms *biomedical model* and *disease model*, as they are used in the literature, appear to be used with a similar meaning.<sup>31</sup> I shall argue that, in conventional psychiatric practice, experiences of mental distress suffered by individuals are typically understood within such a model, according to which such experiences can be categorized within a taxonomy of mental disorders such as the DSM (e.g. Poland, 2014). My aim in this section, therefore, is to provide a broad outline of what seem to be its key features as applied in psychiatry. It should be noted in passing, however, that this model is not universally endorsed within psychiatry and indeed there are some strong critics of it, both from within psychiatry (e.g. Moncrieff & Middleton, 2015) and from other professions (e.g. British Psychological Society, 2017).

The term 'biomedical model' actually refers to the set of theories and practices that underlie Western medical science and practice in general, not just psychiatry. Sean Valles (2020) calls this set of theories and practices *biomedicine* which he describes as "the umbrella theoretical framework for most health science and health technology work done in academic and government settings" (p.1). He also describes it as an institution in the sense that it has permeated a great deal of Western social, cultural and economic life. Biomedicine is the

<sup>&</sup>lt;sup>31</sup> One writer who uses them interchangeably is Donald Kiesler in his book, 'Beyond the disease model of mental disorders' (1999). While he uses 'disease model' in the title, he reverts to the term 'biomedical model' frequently in the text.

conceptual framework underpinning the biomedical model with a distinctive set of features which he identifies. First, he notes that biomedicine encapsulates the view that the causes of diseases are to be found exclusively within biological, chemical, and physical phenomena. Second, along with many natural sciences, laboratory based research is prioritised and randomised control trials are seen as the most preferred methodology for evaluating experimental interventions and treatments. Third, it is based on a reductionist approach to explanation, in that salient phenomena are ultimately explicable in terms of more fundamental underlying processes. Reductionism in biomedicine implies that diseases are conceived of as dysfunctions in particular parts or functions of the body. Biomedicine is the framework that treats much of medicine as reducible to applied biology, with medical knowledge understood as a complicated form of biological knowledge which is reducible to the latter (Valles, 2020).

Reduction is described by Holly Anderson (2017) as a process of focussing down on smaller sizes and less complexity. It is a useful tool in medicine and a powerful method for investigating medical phenomena. In this context, Andersen cites the example of Parkinson's disease. She notes that this was once thought of as an idiopathic disease, one with an apparently spontaneous presentation and no known cause. However, the use by researchers of reductionist methodology has enabled the discovery of the genetic underpinnings in most cases of the disease, and these have been found to have several variants: one of these is monogenic, another results from the interaction of several genes, and a third is linked with environmental causes following exposure to toxins. The reductionist approach has therefore led to greater understanding of the disease and, as such, is testament to the usefulness of this approach. It has similarly yielded greater understanding of many other diseases.

The biomedical model incorporates the features of biomedicine described above. Fred Gifford (2017) notes that this model has become a dominant one in medical science due to its historical success in finding cures for many diseases and easing the suffering of patients for whom no cure is possible. These successes stemmed from the development of reductionist medicine from the 19<sup>th</sup> century onwards, which focussed on underlying biological processes that seemed to account for the illnesses of patients. The understanding gained allowed physicians to identify and treat diseases effectively, with corresponding gains in population health and life expectancy. Such results encouraged the belief that going to increasingly lower and more mechanical levels of explanation would lead to increased scientific knowledge of the causes of diseases. The biomedical model, therefore, is based on the view that diseases can be fully accounted for by deviations from normal biological functioning,

including when these may be triggered by environmental impacts on the body. Diseases on this view are seen as dysfunctions in underlying biological or chemical processes in some part of the body which can be understood on the basis of empirical data generated by the scientific methods used in biochemistry and physiology.

## 3.2.2 The biomedical/disease model in psychiatry

The application of the biomedical model to mental disorders involves similar assumptions about the fundamental causal role of biological and physiological processes to those that apply in somatic medicine. The assumption is that mental disorders can ultimately be reduced to such processes, even though we do not currently know what processes are involved in the case of the conditions described by most diagnostic categories. It is expected that defining abnormalities for each condition will be found in due course at the genetic or neurological level following sustained research programmes.

The biomedical/disease model as applied to mental disorders can be characterised in fairly general terms as incorporating the following two features, which overlap each other to some extent. First, it is implicit in the disease model that mental disorders are similar in some way to diseases in somatic medicine. It embodies the presumption that episodes of severe mental distress are "diseases" which are analogous with somatic diseases in general – i.e. they can be seen as disease entities or "diseases like any other" – "... a physical ailment no different from diabetes or cancer" (Rosenberg 2006, p412). Similarly Nomy Arpaly writes: "Many psychiatrists tell their clients that any mental disorder is 'a disease, just like diabetes" (2005, p282).<sup>32</sup> This is a dominant paradigm within psychiatric services in the western world. Thus Malla, Joober, and Garcia describe it as follows:

The almost exclusively biogenetic conceptual framework for understanding mental illness has acquired a hegemony that has influenced mental health practitioners while also influencing campaigns designed to improve public attitudes towards the mentally ill. As a result, the statement "mental illness is like any other illness" has become almost axiomatic and, therefore, by definition it embodies an accepted truth not in need of a proof (2015, p147).

This is the view that an episode of mental disorder represents the instantiation of a kind of disease entity in a given individual over a finite period of time (although this may sometimes

<sup>&</sup>lt;sup>32</sup> These statements raise the question of what sorts of things we might think diseases are. I discuss the nature of diseases in detail in the following sections.

be a whole life-time in the case of chronic diseases). Where an individual experiences some kind of mental disturbance, this is therefore conceptualised as an expression of a particular psychiatric disease kind, one which can be potentially diagnosed according to one of the categories listed in diagnostic manuals.

Second, a distinctive feature of the model is that diagnostic categories of mental illnesses can be potentially classified into taxonomies (e.g. the DSM) such that they (1) approximately represent actual or hypothesised diseases, and (2) will be refined in future diagnostic manuals so that they come to more closely correspond with actual diseases. This is particularly evident in the World Health Organisation's (WHO) multi-chapter compendium of currently identified diseases, the International Classification of Diseases (ICD). In its eleventh edition (ICD-11) (WHO, 2019/2021) the categories of mental disorder are listed in the fifth chapter. Other chapters contain taxonomies of diseases of the circulatory system, of the eye, and so on. Accordingly, it embodies the assumption that these categories can typically be discriminated as discrete disorders – what Kathryn Tabb (2015) describes as "the assumption of diagnostic discrimination". This is the assumption that individual diagnostic categories can be individuated with sufficient specificity to allow for relevant facts about them to be identified, in such a way that the categories can be further refined and discriminated in light of continuing empirical research. As such, it reflects a similar assumption of the biomedical model in somatic medicine. Thus, to repeat the example mentioned in Chapter 2, the diagnoses of pernicious anaemia, iron-deficiency anaemia, and lymphatic leukaemia, which have similar presenting symptoms, identify quite different diseases with different causal pathways which, therefore, are likely to need different treatments. The differentiation of anaemia into these individual diseases therefore allows important facts about each of them to be identified. In the context of psychiatry, the sorts of facts that psychiatric researchers might be interested in are those concerning neurological abnormalities, genetic patterns or cognitive disorders which, it is hoped, can potentially both validate the diagnosis and lead to greater specificity about causal factors and appropriate treatments for patients given the same diagnosis.

It is widely expected, despite the failure to find clear neurological abnormalities in most cases, that most of these disorders will be shown in due course to be brain diseases, and in many cases also with a genetic component in their causal history, following advances in neuroscience and molecular biology (e.g. Andreasen, 1997; Malla et al., 2015). This would seem a natural expectation to hold within a medical and psychiatric context, if it is assumed that mental disorders constitute disease entities of some kind, given the understanding of the

brain as the seat of human mental activity. This is described by Will Davies (2016) as "internalism" about mental disorders, in which aspects of neurophysiology and neurochemistry are expected to determine the nature and existence of a psychiatric illness. Moreover, one of the leading contributors in the construction of DSM-5 states this explicitly:

The implicit belief that there is an underlying, incompletely understood brain-based dysfunction for the behavioural, cognitive, emotional and physical symptom syndromes is the de facto definition of mental disorders used by most members of the DSM-5 Task Force and Work Groups (Regier, 2012, p293).

Consistent with this view, therefore, research in psychiatric science has been heavily focused in these areas over many years.<sup>33</sup> The expectation that neurobiological explanations of mental disorders will eventually be uncovered remains a prevalent view in psychiatry (e.g. Brückl et al. 2020; Charney et al, 2017). For example, Ure, Corral and Wainwright (2018) in their review of research on schizophrenia explicitly describe it, without further supporting argument, as a "brain disconnection syndrome" (p.1). More recently, Cuttle and Burn (2020) in an editorial entitled "Neuroscience: the way forward" in the journal BJPsych Advances (associated with the British Journal of Psychiatry) describe the Gatsby/Wellcome Neuroscience Project, a programme aimed at integrating modern neuroscience into psychiatric training.

Thus, the biomedical approach to instances of psychological abnormality and distress continues to be strongly endorsed in the practice of contemporary psychiatry. This was explicitly emphasised in a special article, co-authored by 37 prominent British and Canadian psychiatrists and published in 2008 in the British Journal of Psychiatry (Craddock et al., 2008). Similarly, in response to a position statement by the Division of Clinical Psychology of the British Psychological Society (2013) calling for the mental healthcare system to move away from the disease model, an editorial in The Lancet Psychiatry (2015) also reaffirmed the biomedical approach. The editorial states:

The biomedical view is an essential component. New research techniques have

year of each decade of the century (i.e. in 1905, 1915, etc.).

76

<sup>&</sup>lt;sup>33</sup> In a systematic study of research articles published in the British Journal of Psychiatry during the 20<sup>th</sup> century, Moncrieff and Crawford (2001) found a strong predominance of those focused on biological and medical aspects of treatment, with relatively few on psychological therapy or social psychiatry. The authors evaluated all research papers published in the journal during the mid-point

proliferated in the past few years, promising much information about the function of the brain and the mind; neuroscientists will press ahead with this work regardless of the philosophical bias of services (p477).

This is borne out by the continuing publication in a range of journals of studies that report results with the aim of identifying the neurological bases and other biomarkers for mental health conditions, particularly those falling under the diagnostic categories of schizophrenia, bipolar disorder, and depression.

The belief that neurological abnormalities are likely to be salient causal factors in the aetiology of mental disorders implies that, in some way, the conditions represented by the DSM diagnostic categories are analogous to diseases in physical medicine. This is the underlying assumption of the biomedical model of mental disorders and hence underpins the use of the term 'disease model' also. As noted above, this is reflected in the commonly-voiced statement that a mental illness is "a disease like any other". Psychiatric patients may be told this explicitly or, where this is not specifically stated, they may nevertheless gain the belief that this is an appropriate way of understanding their current state of mind. A frequently cited reason for conveying this message to patients is that this may relieve them of any sense of guilt or shame they may feel from the consequences of their disorder – e.g. the need for medication, hospitalisation, etc. I will return to this issue in subsequent chapters where I will argue that, apart from possibly offering some short-term relief, it may frequently fail in this regard, in particular by rendering them vulnerable to becoming victims of epistemic injustice.

Another reason often cited is that it can combat the stigma associated with mental disorder by communicate a similar message to others in the patient's circle, namely that the patient cannot be blamed for falling victim to the disease, any more than they could be blamed for contracting diabetes or Parkinson's disease. However, empirical studies of factors influencing perceived stigma indicate that the presumed analogy with somatic diseases is, if anything, counterproductive in combatting stigma. In particular, there remains a popular conception that mental disorders are a function of some kind of genetically influenced dysfunction in the brain and this attitude is more likely to lead people to react cautiously to, and even avoid contact with, those with mental disorder diagnoses (Pescolido et al., 2010; Read et al., 2006; Rüsch et al., 2005). In a systematic review of 33 population studies of public attitudes towards people with psychiatric disorders, Angermeyer et al. (2011) found that biogenetic attributions of mental disorders were not correlated with tolerant attitudes and, in the case of schizophrenia, were associated with stronger attitudes of rejection. Conversely,

a recent study in China found that people with chronic mental illnesses who attributed their difficulties to biogenetic causes were more inclined to be socially withdrawn than those who made more psychosocial attributions for their difficulties (Li & He, 2021). It is possible, therefore, that there is an interaction effect, in some cases, between the social withdrawal of people with chronic mental illnesses and attitudes of rejection towards them from the wider society.

Evidence of this nature shows that the disease model does not appear to help reduce the stigma experienced by sufferers, despite the frequent expectations that it ought to. In addition, some writers suggest that sufferers may also find themselves facing a paternalistic approach to their treatment with a corresponding diminution of their sense of agency (e.g. Moncrieff & Middleton, 2015; Tekin, 2011), along with the corresponding excusal of responsibility for their behaviour. These aspects of being diagnosed with what is assumed to be a disease-like condition may contribute to the experience of stigma. Further consequences of such paternalism and loss of agency contingent on the diagnosis may be that the sufferer is given less credit for their testimony in conversational exchanges, which is one of the characteristic features of epistemic injustice. The nature of these consequences and the manner in which they might occur will be discussed in more detail in subsequent chapters.

Despite this apparent dominance of the biomedical model, it can be objected that there is less consensus about conceptions of mental disorder among mental health professionals than the various endorsements by leading psychiatrists would seem to suggest. For example, Awais Aftab and co-workers (Aftab et al, 2020) conducted a survey of 209 health care professionals and trainees at a US academic medical centre.<sup>34</sup> They found a range of responses from strong agreement to strong disagreement to statements such as "For a condition to be a mental disorder, there must be an underlying biological abnormality" and "All mental disorders are diseases". This might suggest a more varied conceptualisation of mental disorders amongst healthcare professionals, at least within the particular academic centre where the study was conducted. The authors did not describe in detail the extent to which different professional groups agreed or disagreed with such statements, other than to note that psychology respondents were significantly less likely to agree that "all mental disorders are diseases". However, they did note that respondents generally leaned towards disease attribution for mental disorders which does suggest the continuing influence of the biomedical model.

<sup>&</sup>lt;sup>34</sup> University of California San Diego (UCSD) Health.

Another larger study conducted in Finland by Kari Tikkinen and co-workers (Tikkinen et al., 2019) looked at attitudes to what the authors called "states of being", most of which broadly correspond with categories in the DSM, although some (e.g. grief, homosexuality) did not. A questionnaire about conceptions of mental disorders was sent to a sample of psychiatrists, non-psychiatric physicians, nurses, laypeople, and all 200 members (MPs) of the Finnish parliament. A total of 3259 eligible responses were received.<sup>35</sup> The questions were framed in the form of a statement, "(This state of being) is a disease", to which respondents were asked to pick a response on a five-point Likert scale ranging from 'strongly disagree' to 'strongly agree'. Twenty states of being were chosen, which were: grief, homosexuality, absence of sexual desire, premature ejaculation, transsexualism, work exhaustion, insomnia, drug addiction, gambling addiction, alcoholism, personality disorder, ADHD, social anxiety disorder, bulimia, generalised anxiety disorder, autism, panic disorder, anorexia, depression, schizophrenia. There was a large degree of agreement amongst all classes of respondent that the first five in this list do not constitute diseases. There was also a large measure of agreement that the last five in the list do constitute diseases. For other states of being, there was more variation in responses. Overall, however, there was a clear tendency for psychiatrists and, to a slightly lesser extent, other physicians to regard many of these states of being as diseases.

In both the studies above, the conception of disease in question is not clearly elaborated or explained. Thus, Tikkinen and co-workers acknowledge that each respondent would have used their own understanding of what constitutes a disease in answering the questions and these understandings would not necessarily have coincided. Similarly, Aftab and co-workers comment that this was not explained in their study and that further discussion of what is meant by 'disease' would be desirable. Their own view is that disease is best conceptualised as a state of significant suffering and incapacity. However, other views, which I discuss in the following section, differ from this in specifying that there must be some sort of biological abnormality present, in addition to the suffering experienced by the patient, for a condition to be regarded as a disease.

This, therefore, raises the question of what sorts of things we think diseases are. If mental disorders are in some way analogous to diseases, then what sort of entity is it that mental disorders are said to be like? In the next section, therefore, I discuss this question.

<sup>&</sup>lt;sup>35</sup> Responses were deemed to be eligible for the purposes of the study if the participants demonstrated comprehension of the questionnaire and responded to all or most of the questions about states of being. Those who failed to respond to four or more such questions were excluded from the analysis.

### 3.3 What sorts of things are diseases?

I have described above a broad picture of how the biomedical or disease model of mental disorders can be characterised. If, as this model states, mental disorders are understood to resemble diseases, then the question arises of what actually constitutes a disease, as opposed to other sorts of difficulties that people might complain about. In other words, if an individual is given a psychiatric diagnosis and encouraged to believe that they have "a disease like any other", what sort of thing might they be expected to think they have?

There is a large literature in philosophy of medicine and medical science generally on what constitutes a disease. A number of writers on this topic (e.g. Ereshefsky, 2009; Sisti & Caplan, 2017; Stegenga, 2015) suggest that the various accounts proposed can be subdivided into three broad categories – naturalist, normativist, and hybrid. A full discussion of all the literature on these is beyond the scope of this chapter. My approach, therefore, is to briefly discuss the main arguments for naturalism and normativism, and then argue that hybrid accounts are more plausible than either of these. I will argue in addition that hybrid accounts are consistent with the biomedical model by virtue of the naturalistic element that such accounts contain.

## 3.3.1 Conceptions of disease

One difficulty in answering this question is that words like 'disease' and 'illness' do not have clear definitions and often tend to be used interchangeably in common parlance. In normal usage, they appear to be what Wittgenstein (1958) refers to as family resemblance concepts. Nevertheless, medical practitioners and philosophers have generally found it useful to draw distinctions between these terms and clarify the concepts they represent. Typically, therefore, a distinction is made between illness, as the discomfort and suffering experienced by the patient, and disease, as the diagnosable condition which the medical practitioner identifies as the cause of the patient's condition (e.g. Kleinman, 1988). Christopher Boorse (1975) in particular emphasises this distinction in his account of the nature of a disease. This

<sup>&</sup>lt;sup>36</sup> Wittgenstein also notes that we may reform our language to suit our purposes: "Such a reform for particular practical purposes, an improvement in our terminology designed to prevent misunderstandings in practice, is perfectly possible" (1958, §132).

<sup>&</sup>lt;sup>37</sup> This is a slight oversimplification. In most cases, the doctor's diagnosis will be expected to identify the cause of the patient's symptoms, but this is not always the case. Sometimes a diagnosis will just serve to indicate a pattern which is believed to be similar to the set of symptoms observed in other patients. This may be the case particularly in primary care services when more precise diagnoses are not required to guide treatment. I discuss this issue further in Chapter 4.

distinction is likely to be important in the context of debates about the distinction, or lack of it, between physical and mental illnesses. For example, Kendell (2001) argues that there is no coherent distinction to be made in the latter case, although he does not articulate a clear account of disease which might distinguish it from illness.

## i) Naturalism

The account of disease that Boorse describes is a *naturalistic* one. According to this, a disease is the kind of internal state which is responsible for impairing the health of the patient in any particular case. On this account, a disease is construed solely as an abnormality in typical biological functioning. This is defined as a deviation from statistically normal functioning with respect to a reference class of an appropriate sex and age group. In contrast to disease, health is normal functioning ability in a member of the relevant reference class. The specification of the reference class is important to account for the fact that some of our body parts and systems function differently, depending upon our sex and age. A normal function is a part of a process within an individual which makes a statistically typical contribution, however small, to individual survival or reproduction for the reference class. Disease, therefore, can be described as "a failure of parts of the body to perform biological functions which it is statistically normal for them to perform" (Boorse, 1977, p561). Boorse's account is a naturalistic one in the sense that what constitutes a disease can be determined solely by an examination of the manner in which a given body part is failing to perform the function to the level of efficiency, relative to the appropriate reference class, that is statistically normal. In this sense, the determination of a disease state is a matter of empirical discovery.

#### ii) Normativism

Naturalist accounts, such as Boorse's, are open to the criticism that mere deviation from a statistical norm cannot itself indicate that a given abnormal bodily state is a disease. For example, having red hair is a statistical abnormality, but does not thereby qualify as a disease, even though it is caused by a genetic abnormality. It does not diminish the individual's capacity for survival or reproduction. Similarly, a man of unusually short stature does not have a disease just because of his biological abnormality, even though his short stature might create difficulties for him in finding a mate with whom he could reproduce. Objections of this kind reflect a *normativist* conception of disease, according to which a disease is something which is disvalued or a source of harm to the sufferer. The main features of this conception

are described by William Stempsey (2000) as, firstly, the idea that disease must imply something bad, and, secondly, that statistics alone cannot determine that an abnormal process is a disease. An example of this view is the conception of diseases described by Caroline Whitbeck (1978). The key elements of her view are as follows:

A disease is any type of psycho-physiological process such that:

- 1. People wish *to be able* to prevent or terminate that process because it interferes with the bearer's psycho-physiological capacity to do those things that people commonly wish and expect to be able to do;
- 2. Either the process is statistically abnormal in those at risk or people have some other basis for a reasonable hope of finding means to prevent or effectively treat the process (1978, p211, emphasis in original).

Clause 1 here implies that a value judgment is inherent in the judgment of a given condition as a disease. What people wish and expect to be able to do are partially determined by specific socio-cultural contexts. Clause 2 notes that the condition must either be abnormal in some way, or where it is not abnormal, as the second part of the disjunct indicates, it must be judged as something one would hope to change or ameliorate. Under conditions which might be considered statistically *normal*, but potentially amenable to treatment of some kind, Whitbeck includes the sort of conditions which are associated with normal aging, such as arteriosclerosis. Whilst such conditions might in the past have been regarded as inevitable consequences of aging which would just have to be accepted by the sufferer, more recent advances in medical science have opened up the possibility that their specific causes might be identified and potentially, therefore, become treatable. Whitbeck's explication of the valuejudgments inherent in the attribution of a disease state to a bodily condition goes beyond a judgment of mere badness. There is also a "capability" sense to her conception, according to which a condition warrants being called a disease if people have an interest in being able to influence it in some way. That is, it warrants being understood as such if people have reason to find ways to prevent it or treat it effectively (Stempsey, 2000).

Nevertheless, besides specifying that a disease is some sort of psycho-physiological process which is abnormal in some way, there is little else in Whitbeck's account to make clear what else might distinguish a disease from other bodily states that we might want to prevent or treat. It is not clear that obesity, for example, ought in itself to constitute a disease, although it can be a cause of various diseases and it appears to meet Whitbeck's criteria. Therefore, her criteria seem to be over-inclusive of the conditions that we normally describe as diseases. In particular, she does not include any criterion that there must be some more

specific biological processes or abnormalities that would normally be considered to be necessary for distinguishing one disease from another.

Another normative account of disease is described by Rachel Cooper (2002). In her account of what constitutes a disease, Cooper identifies three main criteria for diseases: they are things that are bad to have; to suffer from them is in some way a matter of being unlucky; and they are potentially (if not always in practice) treatable by medical services. The first of these, that diseases are in some way a bad thing, is intended to distinguish diseases from other biological abnormalities that a person might have – e.g. ginger hair. The second, that diseases represent a matter of bad luck, distinguishes diseases from other biological conditions which are normal and expectable in life – e.g. baldness in men, infirmities due to senescence. The sufferer is to some degree unlucky, whether temporarily or permanently, compared with most other people of the same sex and age. The third, that diseases should be potentially medically treatable, distinguishes diseases from other sources of ill fortune that are unconnected with any kind of bodily functioning. As Cooper describes it: "This condition is required to distinguish diseases from other types of misfortune—economic problems, social problems and so on" (2002, p277). Cooper is careful to note that such conditions should be "potentially" medically treatable. Therefore, the range of conditions that constitute a disease is not exhausted by those that are currently treatable. However, it is not clear on this account how we are to assess whether or not a condition is potentially medically treatable in the absence of any actual available treatment. We could make this judgment if there is already a known or hypothesised cause of the condition, but in that case the account risks being a circular one, since we will have already identified what we believe to be a disease and, on that basis, concluded that it could be treatable. Moreover, what counts as being "medically treatable" is a vague criterion. Not only would this include specific surgical and pharmacological interventions, but also general advice given by doctors regarding dietary and exercise recommendations. When faced with a patient complaining of symptoms that may not allow a confident diagnosis to be made, the doctor may nevertheless prescribe some sort of treatment, if only of a palliative nature to relieve symptoms. Therefore, on this basis, any kind of condition which the patient reports to the doctor could be judged to be a disease, including, for example, everyday aches and pains for which people sometimes consult their doctor.

On this account, one could also wonder whether the common cold would count as disease, despite the fact of it being known to be caused by a virus, since doctors typically say that there is no treatment for it. The response to this worry might be that the doctor would

likely recommend analgesics to relieve the symptoms, and therefore be offering treatment in some sense, but such treatments are palliatives and do not cure the condition itself. Medical practitioners naturally understand the distinction between aiming to cure the condition by means of the treatment they prescribe (even if the treatment fails to achieve a cure) and offering merely palliative remedies which they do not expect to be curative. But it is difficult to see how they could have any such clear distinction if their conception of diseases was a purely normative one. On a normative account, it seems a disease just *is* a condition that is disvalued by the patient such that it might lead her to consult a doctor about it.

#### iii) Hybrid accounts

Despite the objections above to a purely normative account, what such considerations illustrate is that a purely naturalistic account fails to take account of the fact that the things we call diseases are not just deviations from normal biological functioning, but are also unpleasant or harmful in some way for the afflicted individual. Therefore, it would seem that a satisfactory account of diseases would need to incorporate both naturalist and normativist elements within it. The alternative to both, therefore, is a hybrid account. On such an account, each of the above elements is a necessary condition for disease attribution, but individually neither is sufficient. One such account is provided by Jerome Wakefield's (1992, 2007) harmful dysfunctional account. Wakefield frames his account with particular reference to mental disorders and for this purpose he generally refers to 'disorders' rather than diseases. Nevertheless, his account can be understood to apply equally to somatic illnesses. On his account, for a condition to be regarded as a disorder it must meet two criteria. First, a condition is a disorder if it involves an abnormality in some internal mechanism that fails to perform its naturally selected function (i.e. as determined by evolution). Second, the dysfunction causes harm to the individual's wellbeing as defined by social values. The first of these is a naturalistic criterion and the second a normative one.

Wakefield frames his account specifically in terms of functions that are presumed to have been selected by evolution. However, it is not clear why an account of diseases needs to be framed in these terms, even if it's the case that evolution is responsible for the shape of most of our current biological and psychological mechanisms. While many diseases will be characterised by biological abnormalities in internal mechanisms that are a product of evolution, it does not follow that they must necessarily be so in all cases, unless one makes the prior assumption that it is so in all cases, in which case the appeal to evolution would become circular. For example, the appendix is not generally considered to have any adaptive

function, but we still regard acute appendicitis as a disease. Even if turns out from future research that the appendix does have an adaptive function, we can still call appendicitis a disease without knowing that.

Another hybrid account, which avoids including any appeal to evolutionary processes, is the one proposed by Jacob Stegenga (2015). His account incorporates the insights of both naturalism and normativism. As Stegenga describes it, a hybridist account requires that a disease be identified as a condition stemming from an abnormal biological function, in the sense of statistically abnormal as described by Boorse in his naturalist account, and one that is considered to be harmful. Both criteria are necessary and jointly sufficient for such an attribution. Stegenga justifies his account by tying it to an account of what makes medical interventions effective. He argues that the purpose of medical interventions – by which he means therapeutic interventions<sup>38</sup> – is to improve health by targeting diseases. On this account, the naturalistic criterion for disease attribution – that a physiological mechanism is functioning in some abnormal manner – constitutes what he calls the "causal basis of disease". Similarly, he describes the normative criterion for a disease as the "normative basis of disease". For a medical intervention to be fully effective, it needs to target a disease and hence both these criteria must be satisfied for this to be the case. It is sufficient, however, for an intervention on a disease to be effective to some degree if it has an impact either on the causal process underlying the disease or on its harmful effects. Effectiveness does not necessarily imply that the disease would be cured completely by the intervention. For some diseases the intervention targets the causal basis of the disease and will lead to a cure - e.g. pneumonia, which can typically be treated successfully with antibiotics. For others, the target will be the normative basis of the disease when, despite knowledge of the causal processes, no effective intervention on the causal basis is currently possible. Thus, patients with Type-1 diabetes are prescribed insulin to replace the insulin which their bodies are failing to produce in adequate quantities to control blood glucose levels. This treatment does not target the causal basis of the disease, which is understood to be an auto-immune disorder. It does, however, target its normative basis inasmuch as it nullifies the dangerous and potentially fatal

<sup>&</sup>lt;sup>38</sup> This seems a slightly narrow understanding of medical interventions. For example, he excludes vaccinations, although there is a plausible argument that these also target diseases, albeit by doing so *before* the disease has actually become instantiated in the patient. Similarly, the prescription of antibiotics to a patient after surgery is a medical intervention, even if no discrete disease is actually targeted at that point. However, his emphasis on therapeutic interventions is intended to focus on the nature of the disease itself and the processes that define it.

symptoms. The causal basis could, of course, become a target for medical interventions in the future, following new advances in medical science.

One consequence of adopting a hybrid account like Stegenga's, which includes a naturalistic element as a necessary condition, is that it makes the process by which we individuate diseases comprehensible in a way that it is less clear on a normative account. As is demonstrated by the examples of pernicious anaemia, iron-deficiency anaemia, and lymphatic leukaemia, what might be considered to be same disease, and in the past was viewed as the same disease, are actually quite distinct diseases, despite have similar presenting symptoms. On a purely normative account it would be difficult to explain why they should be considered distinct diseases. However, they have clearly different causal bases, which is what individuates them as separate diseases. Moreover, different medical interventions are effective for each disease, since the causal bases of the diseases are different. This is what a naturalistic or hybrid account would predict, and it offers a clear account of how disease entities can be understood in medical science and practice. A hybrid account is also consistent with the biomedical understanding of diseases, due to the naturalistic element which it incorporates.

# 3.3.2 The causal basis of diseases

The naturalistic criterion in Stegenga's account of diseases concerns the biochemical or physiological abnormality which constitutes its causal basis. The dysfunctional processes that can lead to abnormalities are as many and varied as the number of diseases listed in medical taxonomies, and typically these have been studied and understood by means of reductionist methods, as the biomedical model would prescribe. I argue that a plausible framework for understanding how the causal processes underlying diseases can be conceptualised has been proposed by Jeremy Simon (2008) in what he calls his *constructive realist* account. While his aim is to provide a realist account of diseases as abstract entities,<sup>39</sup> he also offers an account of how in general the causal bases of diseases can be understood in view of the multiple causal factors that can lead to their presence in individual cases.

-

<sup>&</sup>lt;sup>39</sup> Simon is interested in whether a realist account of diseases is possible. This leads into metaphysical debates about disease realism and anti-realism, which are not relevant to my thesis. I therefore don't explore this issue further here.

Simon bases his account of models of diseases on the conception of abstract models described by Ronald Giere (1999, 2004),  $^{40}$  in what the latter refers to as constructive realism. Giere discusses how abstract models are used by scientists to represent aspects of the world in which they are interested. Models are used in this way quite generally across the sciences. Giere gives several examples from classical physics. Thus, for example, he argues that the equation for a simple harmonic oscillator (F = -kx) is a general model which can be applied to various physical systems (e.g. a mass hanging from a spring). The model is a simplified description, which when tested empirically against real objects gives an approximately correct value for the variables in the equation – approximately, because there are always extraneous variables in reality which are not encompassed by the equation. Nevertheless, the model has value in that its similarity with salient aspects of the causal structure of the world enables scientists to generate hypotheses and make reliable predictions about future empirical findings. As such, Giere argues, models allow scientists to represent real-world systems.

Simon argues that diseases can be represented in the same way. He conceptualises diseases as abstract entities of some kind in which causal processes are a defining element in the individuation of specific diseases. He elaborates in some detail what he means by 'abstract entity' in terms of models:

Ontologically, a disease is an abstract entity that specifies the structure of part of an otherwise unspecified human organism. The behavior of this model can then be predicted based on what we know of human physiology. To the extent that a given model, by embedding the relevant causal structures, allows us to predict and affect the clinical course of a group of patients, that model will represent a (constructively) real disease (2008, pp362-3).

However, creating a model of a typical disease presents additional complexities, beyond the simpler systems in classical physics described by Giere, in that larger numbers of variables need to be assessed for inclusion in the model – patients with the same disease may present varying symptoms or different degrees of severity for reasons relating to their previous level of health and other idiosyncratic factors. He sees his conception of diseases as abstract entities as being able to account for the differing presentations of the same disease in different patients, where the abstract entity is the disease type which is manifested as disease

87

-

<sup>&</sup>lt;sup>40</sup> Note that Simon's use of the term 'models' in relation to diseases is not the same as the broader usage of the term 'model' (biomedical model, medical model, etc.) which is the theme of this chapter. He conceives of disease models as models of causal relationships which are specific to individual diseases. In this section, my reference to disease models is confined to Simon's usage of the term.

tokens in individual patients. He illustrates this with the example of cystic fibrosis which is a chronic respiratory disease characterised by nasal congestion and breathing difficulties. These symptoms are caused by increased thickness of the mucus in the respiratory tract which the sufferer has difficulty clearing in the way a healthy individual normally can do. Other physiological systems affected can include the gastrointestinal tract and, in some cases, the reproductive system. These symptoms are caused by a disorder in ion transport across cell membranes which alters the biochemistry of the cell environment. This in turn is caused by a genetic mutation, although not all sufferers have the same mutation. The disease is therefore defined by the particular causal structure – i.e. the defect in the ion transport system – that generates its presenting symptoms. Similarly, Simon argues, other diseases can also be defined by their distinctive causal processes, despite the existence of extraneous variables in individual cases.

For many diseases, the causal history is inevitably very complex and not necessarily explicable solely in terms of a single aetiological agent (although in some cases it may be). Thus, Caroline Whitbeck (1977) points out that most diseases have a range of causes which interact with each other and not every causal factor is equally salient in each manifestation of a particular disease entity. What causal factor counts as salient for any particular disease is likely to depend on the relevance it is expected to have for prevention, treatment or cure. Typically, Whitbeck suggests, the aetiological factor considered most salient will be a proximate cause rather than a more remote one. Thus, in the case of infectious diseases the identification of a proximate cause, the infective agent, opens up the possibility of preventive measures and treatment. In such cases, the infective agent may be thought of as "the cause" of the disease. Nevertheless, the disease entity itself is characterised by a multiplicity of aetiological factors, several of which will account for the occurrence of the disease in some instances (see Figure 2 in Chapter 1 for a diagrammatical representation of the aetiological factors leading to duodenal ulcers). This will particularly be the case in instances of complex diseases, such as autoimmune diseases where many causal factors can operate – e.g. Graves' disease (Dragulinescu, 2010). Whilst defining such diseases might be more difficult to do by means of simple causes, these can still be encompassed by the kind of abstract model that Simon proposes, in that abstract models of the causal processes can still be constructed to show the common causal processes that operate. Thus, Whitbeck's cautions about the complexity of the aetiologies of some diseases does not undermine Simon's account.

In addition, Simon's account has the virtue of distinguishing between diseases with distinct causal processes and syndromes of medically unexplained physical symptoms

(MUPS), about which there is no medical consensus that they constitute diseases. Such conditions include chronic fatigue syndrome and irritable bowel syndrome, which are characterised by loosely correlated clusters of symptoms with no known common causal history. These conditions present medical practitioners with dilemmas about how to prescribe effective treatment for their patients in the absence of a clear aetiology which can form the target for medical intervention (Cournoyea & Kennedy, 2014). Such syndromes do of course seem like real conditions to those people afflicted by them. However, because of the current state of medical knowledge, this does not of itself mean that they can be regarded as discrete diseases, in the same way that, for example, influenza, cystic fibrosis, and Graves' disease can be so regarded.

It might be objected that Simon's account is too restrictive in terms of what disease models can look like. It might, for example, be thought that the same disease could have quite different sets of causal relationships as their basis in different patients. However, this would then raise the question of why they should be regarded as the same disease. If the same disease could be understood as having quite different causal processes in different patients, this would seem to lead to a purely normative conception of diseases, which I argued against above. On such a conception, the only way to define a disease would be on the basis of a particular cluster of symptoms and the associated negatively valued suffering experienced by the patient. But to return to the example above, pernicious anaemia, iron-deficiency anaemia, and lymphatic leukaemia are regarded as separate diseases, precisely because there are distinct causal bases underlying each of them. Consequently, Simon's view that causal processes define specific disease entities seems to be a plausible one. It is also consistent with the biomedical model in its use of reductionist explanations of diseases.

#### 3.4 Conclusion

That the biomedical model is the dominant one for somatic diseases in the western world is beyond doubt, due largely to its successes in predicting the course and appropriate treatment for such illnesses. Given this dominance and the fact that psychiatry is a medical specialty, it would be expected that the biomedical model would be a prominent one in psychiatry also. The frequent statements by psychiatric scientists and neuroscientists that neurological abnormalities will in due course, and following further research, be found to be causally relevant in the genesis of mental disorders supports the view of the biomedical model as an important one in modern psychiatry.

The implication of this for someone given a DSM-based diagnosis, such as schizophrenia or bipolar disorder, is likely to be that their condition can be understood as a function of some sort of neurological abnormality, even if this cannot be described at present. It may be stated that the condition has been triggered by some environmental events, such as a bereavement or loss of a job, but the condition itself will be understood as being caused by the presumed abnormality in the patient's brain. Moreover, understood in this way, the diagnosis is liable to be reified in the manner described by Hyman (2010) and discussed in Chapter 2, such that the diagnosis may be understood as representing a distinct disease process. When people become ill and are given a diagnosis, they generally form some kind of idiosyncratic narrative about how their illness may have come about (Kleinman, 1988). This may include a story about, for example, how they came to be infected with a virus or how they might have developed an uncommon disease such as motor neurone disease, for which the aetiology is largely unknown. Being given the diagnosis enables the patient's selfnarrative to incorporate an important biomedical element. The diagnosis will thus have some effect on the patient's self-narrative to a greater or lesser degree, depending upon the nature of the diagnosis and its implications for the patient's future health status. In many cases of psychiatric diagnoses, there will be an implicit biomedical element in the narrative as well, even though this cannot be clearly described, in view of the prominence of the biomedical model in psychiatry and the medicalised nature of most mental health services. Consequently, the receipt of a psychiatric diagnosis is liable to lead to a narrative for the patient with a strong biomedical element to it. I discuss the consequences of such biomedical narratives on the patient's self-narrative in chapters 5 and 6.

In both somatic medicine and psychiatry, receiving a diagnosis can have various consequences for the patient, beyond those of being given a name for the condition and any relevant treatment that the diagnosis may entail. In the institution of biomedicine, diagnoses are recorded in case files and may be difficult to change even when there are concerns about their validity. In some cases, where a diagnosis of a notifiable disease is made, there may be additional consequences falling on the patient, such as the mandatory need to isolate for a period of time. In the next chapter, I discuss the function of diagnosis in the systems of biomedicine, the speech act of delivering a diagnosis, and the institutional fact that is created by the diagnosis.

### Chapter 4

## **Diagnosis in medicine and psychiatry**

#### 4.1 Introduction

In my discussion of the biomedical model in Chapter 3, I argued that a hybrid account of disease, which includes both naturalist and normative conceptions, is a plausible account of how medical scientists and practitioners typically think of diseases. The naturalistic element in this account conceives of diseases as being individuated by their specific causal histories. The naming of these diseases in individual patients is a matter of diagnosis, which is an important part of medical practice. In this chapter, I examine the role of diagnosis in medical practice and the sorts of social consequences it can lead to.

I start in section 4.2 by arguing that diagnoses in somatic medicine are expected to convey some kind of explanatory information regarding the patient's condition. I make some preliminary observations in section 4.2.1 about the different ways in which the word 'diagnosis' can sometimes be used in healthcare settings, and I note that my main focus in this thesis is how diagnosis is used in specialist health services where there typically needs to be an emphasis on obtaining precise diagnoses. I then in section 4.2.2 discuss in detail the arguments as to why in general we expect a medical diagnosis to be explanatory. In sections 4.3.1 and 4.3.2, I consider a pair of related objections to this view, each arguing in a different way that a diagnosis need not have explanatory value. I argue that both objections can be answered.

In section 4.4, I discuss some of the social consequences of a medical diagnosis which have been identified by sociologists. These consequences arise out of the epistemic authority which society invests in the medical profession. This raises the question of how the assignment of a diagnosis can have the kind of social consequences that follow from it. As part of the answer to this, I claim in section 4.5 that the statement of a diagnosis can constitute a perlocutionary speech act, in the terms of Austin's (1962) speech act theory, in that it can change reality for the patient in various ways as a consequence of its announcement.

The personal consequences for the patient of the diagnosis can, of course, be very varied depending on the diagnosis itself and other personal factors. These consequences also arise from the way in which diagnosis carries a formal status in healthcare systems and

bureaucracy. In section 4.6, I argue that the social fact of the diagnosis can be understood as an institutional fact, as this is described by John Searle (1995, 2010) in his account of how social reality is constructed. I argue that it is the particular medical context in which it is made that leads to it becoming an institutional fact.

In section, 4.7, I go on to argue that the social consequences of diagnosis that apply in medicine generally also apply to the field of psychiatry, despite psychiatric diagnoses not having the same explanatory function as those in somatic medicine. I also note some additional implications which apply specifically to psychiatry and have a distinctive impact on the recipient. In section 4.8, I conclude with a summary of these arguments and note the importance of diagnosis in medicine and psychiatry.

## 4.2 The explanatory function of diagnosis in somatic medicine

### 4.2.1 Different meanings of 'diagnosis'

The first thing to note in any discussion of diagnosis is that the word 'diagnosis' can have different meanings. For example, this word can be used by medical practitioners to describe both the process of assessment they go through with their patients and the statement they might make about their conclusions. In this respect, at least as it is used in medical practice, it might appear to express another family-likeness concept, in which no single set of necessary and sufficient conditions defines it. However, I argue below that the apparent looseness in its usage is partly a function of the specific context in which the word is used, and further that, despite this looseness, there is generally an expectation amongst patients that a diagnosis will be explanatory to some degree. This also seems to correspond to what people expect of a diagnosis in other (non-medical) contexts.

As noted, in medical contexts, a diagnosis can refer to the *statement* with which the clinician informs the patient of what name is being given to the condition from which she is suffering and any other associated information as appropriate. Alternatively, it can also refer to the *process* of assessing and deciding on what that condition is – i.e. the process of diagnosis. It can also, though less commonly, refer to the diagnostic categories in a diagnostic manual, although generally the term 'diagnostic category' seems to be preferred. The context usually makes clear what meaning of the word is intended. In what follows, I will generally take 'diagnosis' to refer to the statement made by the clinician of the patient's condition, unless the context clearly indicates otherwise. Most of my discussion will be about the

statement of diagnosis, although some comments about the diagnostic process will also be necessary.

A further point to note about the process of diagnosis in medicine is that it can take place in different settings. This may matter because, as noted, the word 'diagnosis' can be used rather loosely at times. Depending on the severity of their condition, the patient may receive a diagnosis from their GP in a primary healthcare setting or from a specialist in a secondary or tertiary level service, which will usually be in a hospital (although secondary and tertiary mental health services are often located in community units). Appointments with a GP normally last about ten minutes which limits the time available for assessing the patient's complaint. If the patient does not need to be referred for more specialist assessment and treatment, the GP may often make a rapid diagnosis without requesting any further tests, if this is sufficient to guide treatment. The diagnosis made may therefore be quite imprecise, but adequate for the purpose at hand. However, when the patient is referred to a specialist service due to the nature of their condition, the issue of obtaining a more precise diagnosis becomes important. This is also the case for those referred to mental health services.

Consequently, in the remainder of this thesis, my discussion will refer to diagnoses being made in the context of specialist services, unless otherwise indicated.

### 4.2.2 Disease identification and explanation

Except perhaps in some cases where a relatively imprecise diagnosis is deemed adequate for determining the patient's treatment, it is generally expected that, as Stegenga describes it, "The main point of diagnosis is... to establish what disease a patient has" (2018, p176). This statement does not clearly distinguish between the process and the resulting statement and could be read as referring to either. Nevertheless, the implication is that diagnosis is strongly associated with disease identification. As such, the identification of the relevant disease is understood to have some explanatory value on the basis of a hybrid account of disease, as discussed in Chapter 3 and endorsed by Stegenga (2015, 2018), in which a naturalist conception plays a role. However, this statement of the purpose of making a diagnosis requires some qualification in cases where no distinct disease entity can be identified, but there appears to be a group of symptoms with a recognizable pattern which is named by the diagnosis. Typically, these patterns are understood as *syndromes* for which no medical

<sup>&</sup>lt;sup>41</sup> I am describing here the normal arrangements in the NHS in the UK. However, many other developed countries have broadly similar distinctions between primary healthcare providers and more specialised, higher level services.

explanation is available. I discuss this in more detail below, where I consider an objection to the claim that diagnoses are generally expected to provide some explanation for the patient's symptoms.

Nevertheless, whatever qualifications one makes along these lines, the process of diagnosis plays a central role in the identification of diseases in medical practice. Indeed, the concepts of disease entity and diagnosis seem to be mutually dependent. Charles Rosenberg, a historian of medicine who has written extensively about diagnosis, argues that this dependence is linked with the notion of disease specificity. He describes this notion and its relationship with diagnosis in the following terms:

... diseases can and should be thought of as entities existing outside their unique manifestations in particular men and women. During the past century especially, diagnosis, prognosis, and treatment have been linked ever more tightly to specific, agreed-upon disease categories (2002, p237).

This follows from the enormous empirical and theoretical advances in biomedical sciences from the 19<sup>th</sup> century onwards, which have led to diagnosis gaining the importance it now has in medical practice. Disease specificity, therefore, reflects the conception of disease entity, according to which disease entities are individuated in terms of their aetiologies and natural course. This approach to their individuation allows them to be organised into a systematic nosology, as for example in the WHO's International Classification of Diseases (ICD-11). The role of diagnosis therefore is to identify which disease entity, out of a range of possible ones, is instantiated within the patient concerned and which, therefore, can explain the patient's illness. This frequently calls for a process of differential diagnosis where the patient is showing symptoms that might, on initial presentation, be explained by different disease entities. It is the notion of disease specificity which makes differential diagnosis into a purposeful activity.

In other cases, the physician may settle for a provisional diagnosis, which may be described as such, where the precise aetiology of the patient's symptoms is not established, but sufficient information is known to guide effective treatment. It can be argued that diagnoses are always provisional (e.g. Huda, 2019, p.27), in that new information can always come to light which might cause the diagnosis to be revised. However, in reality in the case of well understood and commonly occurring diseases, this happens quite infrequently. Such an argument would appear to conflate diagnosis as the act of naming the disease from which the patient suffers, which in most cases is unlikely to be revised, and the further additions to the clinical information about the specific patient which may entail modifications to the

treatment plan. Inasmuch as a diagnosis may also be regarded as a fully comprehensive statement of explanation about the patient's condition, the diagnosis may sometimes be open to revision in the latter sense.

Hane Maung (2017) discusses the manner in which diagnoses may explain the patient's symptoms and endorses the view that in general a diagnosis does this by identifying the actual causes. He writes:

In clinical practice, the diagnostic process is normally aimed at discovering the pathology that is causing a particular patient's symptoms and signs. The diagnosis, which is the outcome of this process, often denotes this cause.<sup>42</sup> For example, the diagnosis of acute appendicitis points to inflammation of the appendix as the cause of a patient's abdominal pain (2017, p51).

One consequence of the advances in medical science is that patients and clinicians now expect there to be causal explanations of episodes of disease when these occur, and the diagnosis given in a particular case can be thought of as providing this. This is consistent with the naturalistic and hybrid conceptions of disease discussed in Chapter 3. It seems a natural expectation that we should want the diagnosis to give us at least some indication of the causes of the disease which is present. Simply being informed of the medical term for the collection of symptoms we might be experiencing seldom tells us anything about the actual disease we have. Consequently, the identification of a specific disease entity affecting a given patient can be expected to indicate that that disease has its own discrete aetiology, even if this is not fully understood by medical scientists. Thus, for example, a diagnosis of pneumonia can be understood as a statement about the cause of the patient's respiratory difficulties.

Elsewhere, Maung (2019) also notes that diagnostic categories can be abandoned and replaced with more precise categories if, following new research findings, these turn out to have more explanatory value than the ones being replaced. Thus, to use the example he gives, the obsolete diagnosis of dropsy has been replaced by the diagnostic categories of congestive heart failure, cirrhosis of the liver, and nephrotic syndrome. These new diagnoses represent diseases with different causal processes underlying superficially similar clusters of symptoms, which used to be labelled as dropsy.

This view of diagnosis as an explanatory act is based on the notion of an explanation being expected to indicate the cause of a *particular* event – e.g. abdominal pain in a particular

<sup>&</sup>lt;sup>42</sup> Maung states here that the diagnosis "often" denotes the cause. This implies that in some cases, it does not. This is the basis of a key objection to the account of diagnosis I provide here. I address this objection in the section 4.3.

patient at one point in time. The diagnosis therefore carries causal information, and the initial task of the clinician when confronted with a patient is to make a diagnosis that provides an explanation of the patient's symptoms (Thagard, 1999). This is consistent with David Lewis's (1986) account of explanation, according to which explanations answer questions about particular events and are expected to give information about causes of those events. In his paper 'Causal Explanation', he states as his main thesis: "to explain an event is to provide some information about its causal history" (1986, p217). Any such information only counts as an explanation if it includes *causal* information. Lewis argues that without such information the account given could not be regarded as an explanation of the phenomenon.

An explanation is typically an answer to a why-question. As Lewis observes, an explanation can take many forms and be as brief or as complex as the questioner requires to answer the question. The explanation may include many causes acting independently of each other, or else a single salient cause that assumes prime importance with other causal factors appearing relatively unimportant. A diagnosis need not provide a full causal history of the patient's condition and indeed it cannot do this if the full causal history is thought of as a complex causal chain. However, a diagnosis is intended to encapsulate a sufficient amount of explanatory information to allow the patient's condition to be recognised as relevantly similar to a pattern of symptoms suffered by other people and caused by the same disease entity; consequently, it can serve as a guide to treatment decisions.

There is one qualification to this picture of diagnosis as explanation that needs to be added. Maung argues that a fully adequate causal explanation should also include additional information regarding the mechanisms and processes through which the disease entity impacts on the individual organism. This is illustrated by Thagard's (1999) example of the causal mechanisms that explain duodenal ulcers, as shown in Figure 2 in Chapter 1. Thus, a full diagnostic explanation of the patient's condition will need to include reference to the relevant theoretical background in biochemistry, histology, physiology, and other sciences as appropriate. Therefore, a diagnosis is, strictly speaking, a short-hand indication of an explanation, rather than a fully articulated explanation in itself. As Annemarie Jutel describes it: "A diagnosis is a kind of short-hand for encapsulating as much information as possible into a word or phrase" (2016, p165). Nevertheless, in normal clinical contexts a diagnosis can be regarded as providing the explanation needed to guide the patient's treatment, taking account of relevant other aspects of the clinical picture as may be required.

### 4.3 Objections to the view of diagnosis as explanation

So far, I have argued that diagnosis represents a causal explanation of the patient's symptoms, or at any rate that it is short-hand for such an explanation. However, this statement about diagnosis is open to a pair of related objections, each of which contends that not all diagnoses need be explanatory. Since the arguments for each are slightly different, I address them separately here.

## 4.3.1 Non-explanatory diagnoses

It is often stated by medical practitioners that diagnoses are not necessarily explanatory in any sense – in other words, that some diagnoses do not actually indicate an explanation, nor are they intended to. William Stempsey (2000), for example, points out that some diagnoses do not name diseases, but just name groups of symptoms. An example of such a diagnosis which is frequently cited is that of hypertension. This is understood as purely a description, not an explanation, of a patient's condition in which their blood pressure is raised above a certain threshold level. This level is defined by medical consensus as representing a significant health risk to those affected. Nevertheless, it does not follow that a diagnosis of hypertension never carries any causal information. For example, in the case of a patient who suffers from severe headaches, chronic fatigue, and chest pains, this diagnosis may well provide important explanatory information concerning the symptoms. These symptoms can be caused by excessively high blood pressure.

Another example often mentioned is that of obesity. This is a potentially confusing example, in that it was classified as a disease by the American Medical Association in 2013 and is said to be "diagnosed" when the individual's body mass index exceeds 30 (e.g. Kyle, Dhurandhar & Allison, 2016). However, it is not clear that obesity can really be considered a disease in terms of the account of disease discussed in Chapter 3, and it would seem odd for a condition to be classed as a disease purely on the basis of a committee decision to stipulate it as such.<sup>43</sup> If part of the criterion of a disease is that it should have a causal basis by which it can be individuated from other diseases (Stegenga, 2015), then obesity would not meet this criterion, as it lacks any such distinct causal basis. Only in a few cases is there a known

-

<sup>&</sup>lt;sup>43</sup> Despite this stipulation by the American Medical Association, there remains controversy amongst medical scientists over whether obesity is a disease, rather than just a risk factor for various different diseases. For example, Stoner and Cornwall (2014) point out that some people who are obese seem to be fairly fit and healthy, whatever risks their weight may indicate for their future health.

biological cause, such as the rare genetic condition of Prader-Willi syndrome, in which constant hunger is one of the symptoms. Moreover, the process of diagnosing obesity simply involves measuring body mass and height. Doing the relevant calculation then determines whether the individual meets the stipulated criteria for the diagnosis, which has no explanatory value. Describing the act of identifying obesity as a *diagnosis* seems therefore to be an atypical use of that word in somatic medicine.

Typically, the number of such examples of non-explanatory diagnoses cited tends to be small, and it is significant that hypertension and obesity are very frequently chosen as examples in this context. This suggests that such diagnoses are not typical of the function that diagnosis is expected to fulfil in medicine. What these examples illustrate is just that 'diagnosis' is a word that can be used loosely, and hence what may be stated to be a diagnosis in everyday practice does not always name a defined disease entity with a known aetiology. This may be particularly the case in primary healthcare settings, where a more precise diagnosis may be unnecessary. Other examples where a diagnosis does not name a defined disease entity include conditions such as chronic fatigue syndrome and irritable bowel syndrome, for which there is no medical consensus that they constitute actual diseases. Both of these, as their names indicate, are syndromes rather than distinct diseases, and in these cases the diagnosis has no explanatory function.<sup>44</sup>

This is also the case for diagnoses that name syndromes characterised by medically unexplained physical symptoms (MUPS). Cournoyea and Kennedy (2014) argue that, despite their lack of explanatory value, diagnoses of MUPS can still have pragmatic value, in that they can "build patient communities, fuel advocacy and encourage research" (p931). Nevertheless, they state clearly that explanatory diagnoses constitute the most desirable end state of research into illness conditions: "It is uncontroversial in the medical literature that the ideal diagnosis is a biomedical explanation" (p928). Non-explanatory diagnoses can at best only be "optimistic placeholders for future causal explanations" (p930). They also cite evidence to show that an explanatory diagnosis is valuable for the patient as well as the clinician, and that understanding the nature of their disease can have therapeutic benefit for patients in terms of speeding up their recovery. This is a particularly pertinent issue for

-

<sup>&</sup>lt;sup>44</sup> It should be noted that the use of the word 'syndrome' can in some cases be confusing, in that some conditions described as syndromes have known aetiologies. For example, Prader-Willi syndrome and Williams syndrome are rare genetic disorders with an identifiable causal basis. In such cases, the conditions are often named after the scientists who first identified them and the word 'syndrome' has been retained in the name after the genetic aetiology was discovered.

psychiatry. As Craddock and Mynors-Wallis (2014) note in an editorial in the British Journal of Psychiatry:

A fundamental issue in psychiatry is that current classification schemes are of clinical syndromes. Diagnosis is based on descriptive data elicited from clinical observation rather than measurements that relate directly to brain function and pathology (p.93).

As noted in Chapter 2, most psychiatric diagnoses are not based on aetiology and provide no information about the causal history of the individual's symptoms. Hence, they are analogous in this respect to MUPS. What these examples seem to demonstrate, therefore, is that in actual medical practice, which is oriented towards the pragmatic goal of treating patients as effectively as possible, there can be some degree of looseness in how the word 'diagnosis' can be used. Although medical science is constantly revealing new information about the causes of patients' illnesses, medical practitioners have to make diagnostic judgments based on the best available information. Where possible, however, a diagnosis that explains (at least to some extent) the patient's illness is felt to be desirable by patient and doctor alike.

Moreover, this expectation seems intuitively to conform to how we think of the meaning of 'diagnosis' in everyday life away from medicine. When we encounter mechanical or electrical failures in pieces of equipment or machinery, the relevant expert will seek a diagnosis of the problem, which will entail finding out which part of the apparatus is faulty. If one's car fails to start in the morning, a statement that simply places the problem into a pattern of other old cars of the same type with the same problem does not qualify as a diagnosis. Only a statement of the type 'the battery has short-circuited' constitutes a diagnosis, which is what is needed if the problem is to be rectified. Similar expectations of a diagnosis apply when faults occur in computers, washing machines, and so on. We can even talk about diagnosing systemic problems in organisations, such as businesses, in order to identify the cause of the dysfunction when the organisation is judged to be failing to achieve its goals. The point here is not so much about what diagnosis really means, but about our expectations of the function a diagnosis ought to perform. I argue therefore that our expectations are not different in any substantial way when we think of medical diagnoses. Whilst medical practitioners may have varying conceptions of what any particular diagnosis might mean when they make one in practice, the understanding of patients when they receive a diagnosis is likely to be that it is in some way explanatory of their condition.

## 4.3.2 Diagnosis as one element in clinical reasoning

The reference to the pragmatic value of non-explanatory diagnoses leads into the second objection that not all medical diagnoses must be explanatory. This is described by Caroline Whitbeck (1981) who argues that diagnosis is just one aspect of clinical reasoning in general and, consequently, that the view that diagnoses should be explanatory can be over emphasised. She does not deny that diagnoses in many cases do name disease entities, but she argues that they often need not do so. Amongst the latter she includes syndromes which she regards as complete diagnoses, and which would appear to correspond closely to the MUPS syndromes discussed by Cournoyea and Kennedy. 45 These diagnostic judgments may often be assumed to refer to existing but unidentified disease entities and may contain the words 'idiopathic' or 'nonspecific'. As Whitbeck describes it, these terms reflect the inherent uncertainty of such diagnoses, where 'idiopathic' indicates that the aetiology is largely unknown and 'nonspecific' reflects the fact that the infecting agent is unknown in cases where an infection is presumed to have occurred. One might, for example, re-describe the medical diagnosis of the common cold as a 'nonspecific rhinovirus'. For many cases, Whitbeck argues, a diagnosis of this nature is quite sufficient for the clinical purposes of prescribing treatment for the patient. Therefore, in many clinical situations the doctor will make a diagnosis which is idiopathic, nonspecific, or provisional in some other way, and this will often be sufficient for treatment to proceed without demanding that the diagnosis needs to explain the patient's symptoms.

Consequently, Whitbeck offers a definition of diagnosis which avoids any mention of identifying the nature of the disease:

Diagnosis is the process of inquiry aimed at discovering *the causes* and *mechanisms* of a patient's disease insofar as this information is needed to inform treatment and management decisions to achieve the best medical outcome for the patient, and to prevent the disease in others (1981, p324, italics in original).

Stated in such a way, this definition is surely a reasonably accurate and concise description of the way in which clinicians go about assessing the complaints reported by their patients in everyday medical practice. However, as she states explicitly, the meaning of diagnosis she has in mind here is that of the *process* of reaching a diagnostic statement, not the statement

illness assumes greater importance.

<sup>&</sup>lt;sup>45</sup> It has been estimated that between 15 to 30% of consultations in primary care are for medically unexplained physical symptoms (Kirmayer et al. 2004). However, as I stated earlier, my concern here is with secondary and tertiary level services where the search for a precise explanation of the patient's

itself. Therefore, her definition of diagnosis does not contradict the idea that the statement of a diagnosis is one that indicates the cause of the patient's condition. Moreover, diagnosis on her conception would apply to all levels of healthcare service, including primary care, where the diagnosis need not be particularly precise to guide treatment.

Although Whitbeck emphasises that the purpose of diagnosis is essentially a pragmatic one, in terms of its role in enabling effective treatment of the patient's condition, it seems clear from her definition that it must aim at identifying the causes of the illness to some extent if it is to perform this role. Some kind of explanation of the patient's symptoms is necessary, if "the causes and mechanisms" of the patient's illness are to be identified to enable a course of treatment to be instigated. On the other hand, where she says "insofar as this information is needed to inform treatment and management decisions", Whitbeck seems to be suggesting that in some cases it is not necessary to identify any causes or mechanisms for the patient's condition. If that is how her definition is to be understood, then we could respond by saying that treatment might proceed in such cases without a diagnosis at all, rather than saying a non-explanatory diagnosis has been made. The doctor would simply recognise the pattern of symptoms and prescribe the relevant treatment based on previous knowledge. In other words, no diagnosis would be felt necessary for treatment to be offered. Therefore, if a diagnosis is to be made at all, then in order to qualify as a diagnosis on Whitbeck's terms, some degree of explanatory function in the diagnostic assessment would generally seem to be unavoidable. As she says in her definition, it is "aimed at discovering the causes and mechanisms of a patient's disease". Consequently, it is still plausible to argue that, whilst the clinician may well make do with an incomplete explanation or provisional diagnosis as long as it suits the purposes at hand, the general expectation of a diagnosis is that it will explain the presenting symptoms.<sup>46</sup>

It should be added that diagnosis does serve other functions in medical practice as well, such as enabling access to appropriate treatment.<sup>47</sup> It is also a central element in patient records held in hospitals and public health agencies. Correspondingly, diagnostic categories are the basis for calculations by epidemiologists of estimated incidence and prevalence. Similarly, aggregated diagnostic data from patient records frequently form the basis for public health planning and health service development (Jutel, 2017; Rose, 2013).

<sup>&</sup>lt;sup>46</sup> Or to be more precise, the diagnosis will be short-hand for an explanation in the manner described by Maung (2017).

<sup>&</sup>lt;sup>47</sup> In insurance-based health systems, such as in the USA, a diagnosis can also confer eligibility for treatment costs to be covered by the insurer.

Nevertheless, these functions of diagnosis depend upon the assumption that a diagnosis typically identifies a disease of some kind and are therefore secondary to that function.

To summarise, a diagnosis in medicine is typically expected to refer to a disease entity and to have explanatory implications for the patient's symptoms. In everyday clinical practice, however, the doctor's statement of a diagnosis may not necessarily have this implication, since the primary goal of the doctor is to decide on the appropriate treatment for the patient without wasting too much time aiming for unnecessary precision in the diagnosis. In such circumstances the diagnosis given may not go beyond a descriptive label, such as may correspond with a MUPS syndrome. Nevertheless, it remains the case that diagnosis has a fundamental role in medicine as the means by which disease entities can be named and their causes identified. As such, the statement of a diagnosis, particularly in specialist services where precision is important, is important in providing the patient with explanatory information about their condition. Similarly the statement of diagnosis is important in informing other healthcare staff who may be involved in the treatment of the patient. Certain consequences for the patient, in relation both to treatment and more broadly, may therefore follow when the diagnosis is announced.

### 4.4 The social consequences of diagnosis

The discussion above treats diagnosis primarily as a statement of information from one person to another about the patient's disease, briefly noting also some of the direct consequences of that. However, several sociologists have observed that diagnosis is also a social act with social consequences, in that it serves a number of social functions which are not always evident when it is viewed solely as an act of disease identification (e.g. Jutel, 2017; McGann, 2011; Rosenberg, 2002). Thus, for example, Talcott Parsons (1951) has observed how a diagnosis provides what he calls a 'claim for exemption'. The sufferer is enabled to adopt the 'sick role' upon receipt of a diagnosis, which may thereby exempt them from the need to work and to participate in the normal requirements of adult life for a certain period of time. It also, however, carries an expectation that the sufferer will do whatever may be necessary to recover, including complying with treatment regimes, as well as avoiding any behaviour that might transmit the disease to others.

Diagnoses also have the effect of providing indicators of the kinds of conditions society regards as abnormal and potentially in need of treatment (Jutel, 2009). This is particularly evident in the field of psychiatry, in which there have been either regular

additions of new diagnostic categories or modifications to diagnostic criteria whenever the DSM has been revised. One example of the latter concerns changes made to the diagnosis of major depressive disorder, for which the experience of grief following bereavement is no longer excluded from the criteria for diagnosing the condition (Horwitz, 2015).

Consequently, what would previously have been viewed as a natural human reaction to a tragic life event can now be seen as potentially pathological. In addition, the revisions of the DSM have in some cases led to the abandonment of old diagnoses. A well-known example is that of homosexuality, which used to be a diagnostic category in DSM-II, but was dropped from DSM-III following sustained campaigning by the gay rights movement and is no longer treated as a kind of psychopathology. Similarly in the general medical field, new diagnoses can sometimes be created following medical consensus conferences. As discussed above, the decision in 2013 by the American Medical Association stipulating that obesity should regarded as a disease is one such example. One consequence of such changes is to inform affected individuals that certain aspects of their physical or mental states are now to be regarded as pathological, when previously they might not have been.

Diagnoses create a narrative for the institution and healthcare services generally, requiring such things as treatment protocols to be drawn up, illness prevention and vaccination programmes to be implemented, and statistical data regarding disease incidence to be compiled and reported. Such features may have implications for how the patient's experience is altered by the act of diagnosis. Rosenberg describes the set of bureaucratic consequences as follows:

In the act of diagnosis, the patient is necessarily objectified and recreated into a structure of linked pathological concepts and institutionalised social power. Once diagnosed, that bureaucratic and technically alienated disease-defined self now exists in bureaucratic space, a *simulacrum* thriving in a nurturing environment of aggregated data, software, bureaucratic procedures, and seemingly objective treatment plans (2002, p257, italics in original).

In other words, Rosenberg describes patients as being "objectified" within the bureaucratic structures of the healthcare system, in the sense that following diagnosis they become disease-defined representations within that system. The diagnostic category that is recorded is inevitably what gets counted as salient for treatment planning and resource allocation, rather than the patient's subjective experience which cannot be summarised easily – this becomes lost from the system's bureaucratic structures. This is not inevitably the case in the medical consulting room where the patient's experience is often very far from that. However, in a

certain respect, some degree of "objectification" of the patient in the disease-defined sense serves a function, in that the clinician needs to treat the patient's disease as effectively and speedily as possible, without this necessarily meaning that the patient herself actually *is* objectified in the clinical encounter. Nevertheless, Rosenberg's reference to the patient being objectified seems to reflect what Kidd and Carel (2017) observe about ill people being "typically regarded as the objects of the epistemic practices of medicine rather than as participants in them" (p.181) in their account of how patients can become victims of epistemic injustice in healthcare contexts. I argue in Chapter 6 that diagnosis is a specific factor when epistemic injustice occurs in mental health service contexts and that this kind of objectification associated with it is a contributor to such a consequence.

The assignment of a diagnosis can structure the patient's experience in various ways and expose her to a range of practical and social consequences that follow from the diagnosis. The patient becomes exposed to the various organisational practices and structures of healthcare systems as a result of the diagnosis. These concern not only the determination of the appropriate treatment, as would be expected, but also the institutional or bureaucratic consequences that might accompany this, such as requirements for regular health monitoring or checks, expectations to conform to prescribed assessment and treatment regimes, and in some cases curtailment of aspects of patients' everyday lives in response to risk factors that may be associated with a diagnosis (e.g. restrictions on driving cars for people with a diagnosis of severe aortic stenosis).

In addition, if a patient in the UK is diagnosed with a notifiable disease, this must be reported by the clinician to the appropriate officer (as designated in local areas) in the individual's local council or health protection team. Currently there are 33 notifiable diseases in the UK. Similar systems also operate in other countries. This can mean, depending on the diagnosis, that patients may be required to be in quarantine and their rights to patient confidentiality waived in respect of the disease in question. An obvious example recently is that someone who received a diagnosis of Covid-19 was required to self-isolate for a specified period (Covid-19 is no longer a notifiable disease). Thus, a diagnosis can have significant repercussions on the individual's life, beyond those that are a direct consequence of the pathology itself.

#### 4.5 Diagnosis as a speech act

It might seem that the statement to a patient of a diagnosis is nothing more than a simple objective statement about the medical reality affecting the sufferer. However, the social consequences that follow from the diagnosis suggest that there is something more to a diagnosis than a mere statement of the patient's condition. I argue, therefore, that the statement of a diagnosis can be seen as a *performative* in the terms of J.L. Austin's speech act theory (1962) – it is a statement that does something by virtue of its utterance. A similar argument is offered by Kazem Sadegh-Zadeh (2011) based on Austin's speech act theory. Using this framework, Sadegh-Zadeh argues that the statement of a diagnosis in a medical encounter is a *verdictive illocutionary* act.

In Austin's account of speech acts, an illocutionary act represents the force or intent of the act in saying something, such as a promise or a verdict by a judge in a court of law. This is contrasted with *locutionary* acts which are defined purely by the content of the sentence. An illocutionary speech act goes beyond the locutionary act: it is uttered with the intent of achieving the effect on the listener of recognising the particular force of the statement, in addition to that of simply understanding what has been said. Austin describes this as involving "the securing of uptake" (1962, p.117). This can happen in certain ways. One example he gives is that of giving an order which is intended to have the effect on the listener of realising that obedience to the order is expected. Another example is that of promising which is intended to lead the listener to expect that the promise will be fulfilled. Austin also identifies several relatively distinct types of illocutionary act, one of which he labels as verdictive to represent those acts which give a verdict, such as a verdict in a court, which might affect the personal narrative or social status of the individual concerned. Sadegh-Zadeh argues that the pronouncement by a physician of a diagnosis – e.g. "I diagnose you as having diabetes mellitus" (2011, p46) – is not just the communication of a piece of factual information, but also an action in which a certain kind of verdict is announced by a person with the kind of social consequences described in the previous section. As he describes it, the assignment of a diagnosis can be seen as a medical verdict that changes the way the recipients view themselves and the way others view them.

On this view, a diagnostic verdict in particular places expectations on the patient. It takes place within a broader social system and has the effect of altering the position of the patient within that system, even if in many cases this effect has little lasting significance for

the individual concerned.<sup>48</sup> As noted previously, the assertion of a diagnosis alters the social status of patients by assigning the sick role to them, if only temporarily, along with whatever expectations for behaviour may be associated with that diagnosis, such as treatment compliance, attendance for follow-up appointments, and declaration of the diagnosis on applications for a driving license or insurance policy where appropriate.

At the same time, the act of diagnosis influences a wider range of actors, particularly those within the healthcare institutions and systems, such that the appropriate responses to the patient's diagnosis are brought into play, including for example registering the diagnosis on case files, organising treatment, and ensuring relevant disease control measures are in place. Thus, a diagnosis is not just a factual statement about the patient's condition, but, as Sadegh-Zadeh explains:

"it *generates* facts in that it triggers individual, group and organizational behaviour...

The doctor's utterance 'you have diabetes' or 'you have myocardial infarction' makes it appear so in the real-world context' (2011, p416, italics in original).

These facts are the various responses that follow contingently on the diagnosis, and this is also the case even if the doctor has made a misdiagnosis. Inasmuch as the doctor's pronouncement of the diagnosis is intended to make such things happen, the statement is a speech act with distinctive illocutionary features. While the claims about the social functions of diagnosis may appear intuitively plausible on their own, Sadegh-Zadeh's argument in terms of speech act theory supports them by showing how the pronouncement of a diagnosis may have such functions.

Sadegh-Zadeh focuses on the illocutionary act of diagnosis and he sees this as accounting for its social functions. He also notes that diagnosis constitutes a *perlocutionary* act, though he places less emphasis on this. However, I argue that the individual and group reactions he notes as consequences of diagnoses can be better understood in terms of perlocutionary speech acts. As Austin describes the latter, these are identified by the consequential effects brought about by an illocutionary act, such as the defendant in a trial assuming the mantle of guilt entailed by the judge's verdict. However, Jennifer Hornsby (1994) argues that the distinction between illocutionary and perlocutionary acts needs to be made more clearly than Austin appears to do. Hornsby notes that Austin offers a number of

-

<sup>&</sup>lt;sup>48</sup> For example, a diagnosis of a transient infectious disease may lead to the doctor to prescribe an antibiotic, which is successful in curing the disease. Nevertheless, the diagnosis and the prescription will be recorded in the patient's case notes and this information may be relevant in some future healthcare episode.

ways of explicating the illocutionary-perlocutionary distinction, with the consequence that his understanding of this can be open to varying interpretations. One that seems prominent in his work is that he regards an illocutionary act as being essentially conventional, whereas a perlocutionary act is not.<sup>49</sup> On this interpretation, the illocutionary act is conventional in the sense that it achieves its purpose directly as a consequence of social conventions relating to certain forms of expression, such as those associated with words such as 'promise' and 'warn'. Hornsby rejects this interpretation on the grounds that in many cases of such acts no obvious conventions seem to be clearly operative. Instead, she proposes that the success of an illocutionary act can be better understood as dependent on *reciprocity* between the speaker and listener. What she understands by reciprocity in this context is the situation pertaining, such that when a speaker utters a statement with a specific intent (e.g. to convey some information or achieve some effect) the listener will recognise what the speaker is doing. This is sufficient to establish reciprocity, and the listener need not agree with the speaker or change their behaviour in any way. The listener's recognition of the speaker's communicative intent is what makes the illocutionary act successful.

Having delineated illocution in this way, Hornsby then argues that perlocutionary acts can be identified by the additional consequences going beyond anything that reciprocity on its own can effect. Such consequences include the listener being persuaded by the speaker's message, other cognitive and emotional reactions that that may engender, and other behaviours of the listener that may result directly from the speech act concerned. In the case of diagnosis therefore and following Hornsby's argument, the patient's acceptance of the diagnosis and their conforming to the particular role implied by this are what identify the perlocutionary act effected in the statement of the diagnosis. The perlocutionary act is also evident in relation to the corresponding consequences on other agents in the healthcare system who may be involved with the patient in any way. Whilst Sadegh-Zadeh agrees that the consequent behaviour of the patient and of those in the patient's social environment are part of what are generated by the perlocutionary act of the diagnostic statement, he differs from Hornsby in not viewing the patient's acceptance of the diagnosis this way. Instead he says: "The diagnosis proper, e.g. the supposed 'fact' and the induced social belief and role that the patient has diabetes, is the illocutionary act of the physician" (2011, p416, italics in original). In contrast, Hornsby views any cognitive reaction, such as a belief induced by the

<sup>&</sup>lt;sup>49</sup> Austin states: "Illocutionary acts are conventional acts: perlocutionary acts are *not* conventional" (1962, p121, italics in original).

diagnostic statement, as being a consequence generated by the *perlocutionary* act. Understood this way, all direct consequences of the statement are what identify the perlocutionary act.

Austin's account of speech acts tends to deal with these as relatively discrete acts generally delivered by individual actors. However, the communication of diagnoses may often be more protracted than this, particularly in complex cases, and may involve a number of agents in the healthcare system. Nevertheless, there is generally a point at which the physician, having settled upon the appropriate diagnosis, communicates it to the patient and other healthcare workers who may be involved. These acts of communication need not occur simultaneously, but each one can still function as illocutionary and perlocutionary acts in the manner described by Hornsby. The process of diagnosis takes place in a social environment and the resultant statement of diagnosis, therefore, is also given its status within a social environment. The perlocutionary act in particular is what can lead to various social consequences which can have important implications for the patient, beyond simply affecting the treatment prescribed. In the next section, I claim that the diagnosis constitutes an institutional fact within the healthcare system, and in some cases, such as those concerning notifiable diseases, in the wider social system as well.

### 4.6 Diagnosis as an institutional fact

The consequences of a diagnosis described above indicate that a statement of diagnosis is more than just a statement about empirical reality or a locutionary act in Austin's terms. Rather, as argued in the previous section, it constitutes a perlocutionary act. Inasmuch as it does so, it is a statement with implications which go beyond a mere statement about the patient's illness. As Sadegh-Zadeh, in the quote above, says: "it *generates* facts in that it triggers individual, group and organizational behaviour". In this section, I argue that a diagnosis constitutes a particular kind of socially recognised fact, in that the pronouncement of the diagnosis is expected to generate acknowledgement and agreement within a defined social group (the patient and the healthcare staff involved), as well as collective agreement that something appropriate needs to be done in response to it, due to the authoritative status it carries.

A diagnosis can only be given by the right kind of person – one with the socially recognised qualifications to do it. As Annemarie Jutel notes: "Medicine has an officially approved monopoly over the right to define health and to treat illness" (2009, p284). If an

unqualified person attempts to make a diagnosis, it cannot be recognised as a real diagnosis, even if subsequent medical examination happens to confirm it as correct. Similarly, one cannot diagnose oneself.<sup>50</sup> There may be some mundane exceptions to this, such as when we can diagnose our current illness as the common cold without needing any professional assistance to do so. In general, however, any non-qualified person attempting to make a diagnosis of any illness which is more than such a simple and frequent occurrence cannot be said to have made a medical diagnosis, even if they claim to have done so. What makes the statement into a diagnosis is the fact of it being given by a person with the appropriate epistemic authority in the appropriate context.

The pronouncement of a diagnosis in the appropriate context can be seen as creating what John Searle (1995, 2010) describes as an *institutional fact*. That the statement only assumes the status of a diagnosis when given by a person with the generally recognised qualifications to do so and following the generally recognised processes is what makes it an institutional fact, in Searle's terms. According to Searle, institutional facts are created by the formula "X counts as Y in C", such that a socially recognised fact becomes invested with a certain kind of institutional status conferred by a socially recognised authority. The X term can refer quite generally to any physical object or person, the Y term refers to the institutional fact created, and the C term refers to the context in which the X term can assume the status implied by the Y term. The institutional fact is then recognised as carrying the appropriate authority within the community concerned.

One of the examples that Searle uses to clarify this is that of money. Thus, a \$20 bill is in one sense just a piece of green paper of a certain size and with certain markings on it. Searle observes that any kind of object can in theory become a form of money. What makes the \$20 bill into an item of currency is the validation given to it by the US treasury, which is conveyed by the writing on it, and the collective intentionality of the community who treat it as a medium of exchange and value. This, in brief, is the context within which the piece of green paper can become the institutional fact of being a designated item of currency. Another example Searle uses is that of the person who becomes the president of the United States. What makes one individual, say John Smith, count as the US president is the particular

-

<sup>&</sup>lt;sup>50</sup> One might wonder whether doctors can diagnose themselves. They might, of course, be very good at hypothesising what disease they are suffering from, but without consulting another doctor for assessment and being referred for the relevant biomedical tests, any such hypothesis by the afflicted doctor would not count as a diagnosis in the terms I describe it here. Moreover, the reason why doctors who are unwell are advised to consult their own doctor in the same way as anyone else is because the illness can impair the objective approach to diagnosis which any patient requires.

context and socially accepted process by which this title can be conferred on him. When he becomes the president, his doing so creates an institutional fact.

Searle states that the notion of "object" represented by the X term in his formula is sufficiently general to include speech acts. There are many such speech acts that can become institutional facts in the appropriate context. He cites as an example the speech act in which a priest or a marriage registrar announces, "I hereby pronounce you husband and wife", which is not just a statement of empirical fact. This statement has no institutional force when spoken by somebody without the relevant designation, nor even when spoken by an authorised person in a context other than a wedding in a designated location. Without the relevant context it would be meaningless. It only counts as an institutional fact in the appropriate context, and the institutional fact thereby created is one in which two previously unmarried people become legally married to each other.

The pronouncement of a medical diagnosis is clearly a speech act and one with social implications, such as those discussed above. Certain conditions must apply before the statement can become one of stating a diagnosis – as described above, these include the requirement that the right person with the right kind of epistemic authority makes it, following the right kind of processes. The statement therefore is the X term in Searle's formula, the necessary medical system and processes forms the context which is represented by the C term, and the formal diagnosis is the Y term. The diagnosis will be recorded on the patient's case file and will be instrumental in determining what treatment may be prescribed. The fact of the diagnosis may also be used for other, more general purposes, such as the collation of epidemiological data and health service planning. The significance of the institutional fact created will depend on the particular diagnosis given. For diagnoses of nonserious or transient diseases, the significance will be quite limited. When the patient recovers from the disease, the diagnosis is no longer applicable and the institutional fact is no longer salient. For diagnoses of chronic or life-limiting diseases, the institutional fact of the diagnosis remains in existence for the remainder of the patient's life. In any case, the fact of the diagnosis shapes, not just the patient's way of life, but also the responses of the social environments (i.e. family, healthcare system, and other social groups) to it, for as long as the diagnosis is applicable.

Searle (2010) also lists the kinds of institutions which can generate institutional facts. This includes professional activities which, although not institutions themselves, contain or encompass institutions that underpin their functioning. He includes in these such activities as

law, medicine, academia, theatre, carpentry, and retail trade.<sup>51</sup> The institutions encompassed by medicine include hospitals, community health services, medical schools, professional bodies, and ethical and regulatory councils. Most of these institutions impact on the process and conclusion of diagnosis in one way or another: for example, diagnoses are entered in records in hospitals and healthcare systems, and medical schools and professional bodies award the qualifications which underpin the epistemic authority of doctors and hence the diagnoses they make. Thus, it can be plausibly argued that diagnoses do create institutional facts.

Moreover, diagnoses meet an important test specified by Searle for whether a phenomenon or fact can be understood as institutional. The test for the fact is: "Does its existence imply deontic powers, powers such as those of rights, duties, obligations, requirements, and authorizations?" (2010, p91). Medicine, or rather the set of institutions that medicine encompasses, is a field that confers these sorts of deontic powers and obligations which apply to the praxis of diagnosis, as well to other activities such as treatment prescriptions, service planning, etc. The process and statement of diagnosis issue from individuals with socially recognised and authorised powers to undertake these activities. Corresponding with these powers are the obligations to prescribe the most appropriate treatment for the patient's condition and to conduct all dealings with the patient in accordance with the ethical codes laid down by the professional regulatory bodies.

The social and institutional consequences of the fact represented by a formal diagnosis can be illustrated by considering the case of misdiagnosis. When this happens, the clinician makes a diagnosis which incorrectly identifies the patient's disease. Nevertheless, it remains the patient's diagnosis, which is recorded in the patient's case file and which generates further courses of action, including treatment plans which are believed to be appropriate. If and when the mistake is recognised and the patient given a new diagnosis, this is then recorded in the case file and new treatment plans will follow. However, the original diagnosis remains on the file as the diagnosis which the patient previously had. It will always be true to say that the patient *had* that diagnosis, even if it has now been replaced by a new one. In this respect, a diagnosis has more important social implications than other statements by experts in a particular field – e.g. a chemist identifying a gas, and perhaps doing so incorrectly, in a chemical experiment in which no other social consequences follow from it. There may of

<sup>&</sup>lt;sup>51</sup> Whereas it seems clear how law, medicine, and academia can contain institutions, it is not clear in what way carpentry, for example, is an activity that does so. Searle does not say anything more about these examples.

course be instances where social consequences do follow from general scientific findings. For example, a social fact is created by the scientific fact that smoking cigarettes is a major risk factor for developing lung cancer. However, this social fact is just a commonly accepted fact which influences the behaviour of many people, but not everyone. It also influences the behaviour of doctors working in healthcare institutions, in terms of the questions they frequently ask patients, but this is not sufficient to make it an institutional fact in Searle' terms.

When the conditions described above regarding the process of diagnosis obtain, there is then a corresponding expectation that the patient will accept the diagnosis. <sup>52</sup> Acceptance of the diagnosis by the patient also implies acceptance of the authority behind it and of the associated narratives. These narratives cover the expected behaviour and experiences of patients with the given diagnosis. Thus, diagnoses such as epilepsy or diabetes have implications for the patient in terms of how their lives will be shaped subsequently. In some cases also, there may be specific personal consequences in addition to treatment prescriptions arising from the diagnosis, such as increased premiums for car, travel, and health insurance. In such cases, the individual will be expected to declare any relevant diagnoses on insurance applications.

## 4.7 Psychiatric diagnoses

There are some specific issues arising with psychiatric diagnoses which do not generally arise in somatic medicine. Psychiatry is a medical speciality and psychiatrists are medical practitioners. Hence, the usual characteristics of diagnosis in medicine described above might be expected to apply equally to diagnosis in psychiatry. Thus, Maung (2019) comments about psychiatry, "...its diagnoses are sometimes presented as if they serve the same sorts of function as diagnoses in bodily medicine" (p.508). However, as I have argued in section 4.2, diagnoses in somatic medicine are usually explanatory to some degree. This is not the case for psychiatric diagnoses. As Craddock and Mynors-Wallis (2014) have noted, psychiatric diagnostic categories refer to clinical syndromes, not to identifiable disease entities with discrete aetiologies. In this respect therefore, they resemble the MUPS conditions discussed

\_

There may be strong emotional reasons why the patient may find this hard to do when the diagnosis is of a serious or life-limiting disease. Nevertheless, even in these cases, the patient is likely to recognise the pressure on them to accept the authority of the physician and to accept that there is some kind of serious disease that they need to come to terms with.

above. As discussed in Chapter 2, these diagnoses do not offer an explanation of the patient's condition in the way that somatic diagnoses generally do, nor do they predict the course and outcome of the condition with any degree of reliability. Inasmuch as they fail to do this, they lack validity.

There is ongoing debate within psychiatry and clinical psychology about the issue of validity of diagnoses and their lack of explanatory value, but many psychiatrists claim that this is not how the use of psychiatric diagnoses ought to be evaluated. This is expressed, for example, in a paper by two leading psychiatrists, Robert Kendell and Assen Jablensky, who concede that these diagnoses generally lack validity, but argue that they are primarily justified by their utility:

Thoughtful clinicians have long been aware that diagnostic categories are simply concepts, justified only by whether they provide a useful framework for organizing and explaining the complexity of clinical experience in order to derive inferences about outcome and to guide decisions about treatment (2003, p5).

The central argument that Kendell and Jablensky propose is that utility should be considered separately from validity. Therefore, as long as diagnostic categories actually enable such inferences, they can be considered to be useful, without the implication that they represent distinct disease entities or have any kind of validity. Nevertheless, as Schaffner (2012) argues, these considerations cannot be distinguished so easily. A diagnostic concept which succeeds in generating reliable predictions about outcome, as a consequence of the treatment decisions which are judged to follow from it, can be said to have predictive validity. It is difficult to see how diagnostic categories could have much utility if they lack predictive validity. Moreover, as this quote states, Kendell and Jablensky regard diagnostic categories as useful for "organizing and explaining the complexity of clinical experience". What they mean by "organizing" here is unclear, but this quote clearly implies that these categories are expected to have some explanatory function. As well as seeming to imply that they likely therefore do have some degree of validity, contrary to what the authors state elsewhere, it would also seem that this expectation would be communicated in some way to the patients who receive them. It is difficult to see how patients could avoid having some expectation that a diagnosis represents an explanation when the clinicians themselves, explicitly or implicitly, suggest the same thing.

As noted in Chapter 3, patients with a psychiatric diagnosis are frequently encouraged to believe they have "a disease like any other" (e.g. Malla et al., 2015; Rosenberg, 2006). This is liable to suggest that the diagnosis does have some explanatory function, even though

patients do not have a clear understanding how it does. The belief that these diagnoses correspond to specific disease entities is likely to be reinforced by the widespread prescribing of psychoactive drugs for many diagnosed conditions, in the same way that drugs are frequently prescribed to target identified somatic diseases. As Rosenberg observes: "...depression is legitimated ontologically by the drugs that treat it" (2006, p418). However, the assignment of a psychiatric diagnosis, though it may encourage the narrative of "a disease like any other", does not generally correspond with a distinct disease entity. This is not a case of misdiagnosis, in which the physician mistakes one disease for another, but one where it has not been established by psychiatric science that any distinct disease entity is present at all. Nevertheless, the assignment of such a diagnosis can still be seen as constituting an institutional fact in the same way that a somatic diagnosis does, because the same kind of context (the 'C' term in Searle's formula) – i.e. a formalised diagnostic process – has been used. This is despite the diagnosis lacking any explanatory implications for the patient's condition.

The broader consequences of a psychiatric diagnosis, such as the conferral of the sick role and other bureaucratic and social consequences, are likely to impact on patients in similar ways as for diagnoses in general. However, there are additional potential consequences arising for the patient associated with psychiatric diagnoses which do not generally arise with diagnoses for somatic illnesses. First, the receipt of a psychiatric diagnosis is likely to lead to the patient experiencing some degree of stigma as a consequence (Ben-Zeev et al., 2010; Rüsch et al., 2005). The stigma associated with a diagnosis seems paradoxically to be increased by the popular belief that the sufferer has some kind of inherent personal or biological defect which underlies the psychological disturbance (Haslam, 2014), despite the supposition that attributing the disorder to an illness of some kind would seem to absolve the sufferer from any responsibility for it. In this respect, a psychiatric diagnosis differs markedly from somatic diagnoses which do not, except in a few cases (e.g. a diagnosis of AIDS), lead to any stigma.

Second, individuals with serious psychological disorders are at risk of being compulsorily detained in an institution under the terms of mental health legislation. Although it is not obligatory in the UK that individuals who are detained under the Mental Health Act should receive a diagnosis before this can happen, it will very often be the case that a psychiatric diagnosis will form part of the justification for doing this, along with other relevant information about them. Moreover, their continued detention in an institution can also be justified in part by the diagnosis assigned to them, and any appeal they may make to a

Mental Health Act Tribunal for their compulsory detention to be lifted may be countered by the psychiatrist's report of their diagnosis. Therefore, assigning a psychiatric diagnosis to a patient can have very far reaching implications for the patient including the deprivation of their liberty for unspecified periods of time. In this context, the power of the psychiatrist is quite overt and extends beyond the usual epistemic medical power involved in making a diagnosis and prescribing relevant treatment.

Finally, a psychiatric diagnosis can have significant implications for the self-narratives which patients may generate for themselves which go beyond the narrative implications that may follow from medical diagnoses in general. These implications include the vulnerability that patients may have for experiences of epistemic injustice. I discuss these implications in more detail in the following chapters.

### 4.8 Conclusion

To receive a diagnosis is to be told something important about oneself. It is to be told, in effect, that there is something wrong with one's body or one's mind, which to some extent is understood, however loosely, to explain the pain or distress one is currently experiencing. The notion of diagnosis typically conveys the idea that an explanation (or at least a partial one) has been given. Diagnosis in medicine, therefore, is a significant event for the patient and carries a number of social consequences which can easily be overlooked due to our tendency to take them for granted in most of our medical encounters. This is particularly the case for psychiatric diagnoses, where the implication that one has some kind of malfunction in one's mind or brain, understood as a disease of some kind, is likely to have a far-reaching impact on the patient's self-narrative. The statement of a diagnosis is a speech act, one with illocutionary and perlocutionary force. This statement therefore has an effect on the recipient that changes them in some way. It turns what initially appears to be a straightforward factual statement into one with institutional significance that cannot be readily refuted or rejected by the listener. This is achieved by virtue of the recognised authority of the person making it, the process through which it was arrived at, and the medicalised context in which the whole process takes place. This is what makes it into an institutional fact in Searle's terms. The patient may in some cases want to reject the diagnosis, but the fact of the diagnosis being made and recorded in the patient's case file cannot normally be annulled, except when it is replaced by another diagnosis which is considered to be more appropriate. It is very difficult for the patient to question the doctor's judgment about a diagnosis.

The impact of a psychiatric diagnosis on the patient's self-narrative is discussed in depth in two papers by Şerife Tekin (2011, 2014) in which she explains the importance of narratives in shaping people's lives. She argues that such a diagnosis can be damaging to the patient's self-insight, as well as to their self-confidence and expectations of recovery. I therefore discuss these issues and Tekin's argument further in Chapter 5, to explain how people find meaning in their lives through the narratives they have about themselves. I will then argue in Chapter 6 that it is the impact these diagnoses have on people's self-narratives that can lead to the recipients of the diagnoses becoming victims of hermeneutical injustice.

## **Chapter 5**

### The effect of a psychiatric diagnosis on the recipient's self-narrative

### 5.1 Introduction

Medical diagnoses are obviously important for those who receive them due to the meaning a diagnosis conveys about the state of one's body or mind. This is particularly the case for diagnoses of chronic or life-threatening diseases, as well as for people with mental health difficulties. In this chapter, I argue that psychiatric diagnoses are likely to have significant consequences for the self-narratives of the individuals concerned and for the psychological sequelae which result. More precisely, I argue that self-narratives are a fundamental feature of human psychology and that the receipt of a psychiatric diagnosis can alter a person's self-narrative in a profound manner by implying that the patient's mental health difficulties are a function of some kind of biomedical process.

In section 5.2, I start by discussing the general shape of narrative theories and the common themes they embody, in particular arguing that self-narratives are a universal or near-universal means of understanding ourselves. There are many published accounts of narrativity which overlap with each other, but also differ in various ways. Due to space limitations I cannot cover them all here. Consequently, the accounts I discuss are those that are frequently cited in the literature as relevant to the narrative effects of receiving a diagnosis. However, these accounts differ in the extent to which they regard narratives as somehow *constituting* the self, compared with those treating narratives primarily as sources of meaning and intelligibility in the individual's life. For ease of exposition, I first, in section 5.2.1, discuss accounts placing emphasis on the meaning and intelligibility deriving from narratives, and then in 5.2.2 those that treat narratives as constituting the self. I claim that the former are plausible accounts of the manner in which people make sense of the chronology of events in their lives, rather than arguing for the view that narratives actually constitute the self which is a more contentious view. I summarise these arguments for narrativity in section 5.2.3.

In section 5.3, I discuss objections made by Galen Strawson (2004) to the effect that, firstly, self-narratives are not the typical feature of human psychology that narrative theorists claim they are, and secondly, to the extent that they may be a typical feature, this is a trivial

thesis. I treat each aspect of these objections in turn in sections 5.3.1 and 5.3.2. I argue that these objections can be met as long as claims for narrativity avoid making strong claims about narratives and the self.

In section 5.4, I consider Tekin's (2011, 2014) idea that a diagnosis of mental disorder based on the DSM categories constitutes a source of narrative that has consequences for the individual's subsequent development and their chances of recovering from the condition that led to the diagnosis. I describe, in section 5.4.1, how the diagnosis may affect the patient's self-respect, feelings of agency, and self-insight. I argue that these effects can additionally be understood in terms of the concept of self-efficacy (Bandura, 1982, 1997) in that the patient's self-narrative will reflect in some way his beliefs about his ability to influence outcomes that matter for him. The narrative that the diagnosis conveys can imply that some unspecified biomedical condition is acting in such a way as to diminish his ability to do this. The diagnosis may also lead the patient to question whether his experiences are really his own or are a part of the illness, a state described as 'self-illness ambiguity' by Sadler (2004, 2007). This raises the question of how a diagnosis can have such effects. I therefore discuss in section 5.4.2 how the particular features of the DSM diagnoses and the manner in which they are determined are liable to have the effect of marginalising the self-narrative of the patient.

I conclude in section 5.5 by briefly discussing an alternative formulation-based approach to assessing human psychological distress which can strengthen the patient's self-narrative and self-understanding and thus lead to appropriate psychological therapy without the need for a DSM diagnosis.

### 5.2 Accounts of narrativity

The central idea in theories of narrativity is that we live and order our lives according to implicit or explicit narratives of some kind. We experience our lives in time – narratives have a temporal or linear structure – and the stories we construct about own lives in some way shape who we are by describing how we have developed over time. This is reflected in very many aspects of our culture in which narratives and stories are omnipresent, such as in myths, literature, drama, films, television soap-operas, biography, and so on. The autobiographies we construct for ourselves may not and do not have the ordered aesthetic structure of narratives in literature or drama, but they are nevertheless vehicles that carry meaning for us. Our lives are more than just a seemingly endless sequence of days with one following another without

any meaningful connection. What strings our days and years together into more than just a mechanical sequence in time is the self-narrative we have that connects them for us.

Many philosophers and some psychologists have offered accounts about the importance of narratives in our lives. These accounts vary in many ways, although they can be roughly divided into those that emphasise the meaning or intelligibility that narratives provide for their protagonists and those that claim that self-narratives constitute the self in some way. In the following two sections, I discuss these two families of accounts in turn. Then, in section 5.2.3, I argue in more detail that theories of narrativity that focus primarily on the intelligibility and meaning that are generated by narratives offer a plausible account of how we make sense of our lives over the course of time.

## 5.2.1 Intelligibility and meaning in narratives

What is meant by 'meaning' in this context is the sense in which self-narratives represent to us who we have become, where we think we ought to be going in our lives, and more generally how we make sense of our lives. Charles Taylor (1989) argues that having a narrative that reflects one's life story to some extent serves both these functions. Our narratives also contribute to our sense of identity as individuals and the self-concepts we form. As Taylor describes it, one's identity is given particular shape by the narrative that one has of one's life – where one has come from, what one has done and experienced, and where one envisages going in future. As he describes it:

My life always has this degree of narrative understanding, that I understand my present action in the form of an 'and then': there was A (what I am), and then I do B (what I project to become). But narrative must play a bigger role than merely structuring my present. What I am has to be understood as what I have become (1989, p.47).

On this view, the self-concept that one acquires is closely bound up with the narrative one has. Taylor suggests that how one understands one's self emerges from the narrative. This is not to claim that the self is *wholly* created by one's narrative, nor that there is any kind of metaphysical self that can be represented by the narrative. Instead, this is a claim about the psychological significance of narratives for a person's self-concept – i.e. about how one understands *oneself*, rather than any claim about the nature of one's *self*. Taylor also argues that conceiving our lives in the shape of a narrative is a basic condition of making sense of ourselves. For him, only a coherent narrative can answer fundamental questions about who we are, even though aspects of our selves may not be fully represented by it. Nevertheless, on his account, an individual's narrative is an essential element for her or his self-understanding.

The importance of narratives in generating intelligibility and meaning for their possessors is also highlighted in the account of Alasdair MacIntyre (2007), who emphasises the importance of the concept of intelligibility for understanding human actions, both of ourselves and of others. He says: "... the concept of an intelligible action is a more fundamental concept than that of an action as such" (2007, p.209). Narratives represent the vehicle by which this is attained. For MacIntyre, this is an important thesis: "...man is in his actions and practice, as well as in his fictions, essentially a story-telling animal" (ibid, p.216). Actions only become intelligible by virtue of their place in a narrative – otherwise they are nothing more than disconnected elements with the potentiality to form part of a narrative. Unless they are incorporated within a narrative, however brief and simple, they are actions without meaning for the agent. Moreover, our own experiences of seeing our lives in narrative form and recognising the actions of other people as intelligible are interconnected: "It is because we all live out narratives in our lives and because we understand our own lives in terms of the narratives that we live out that the form of narrative is appropriate for understanding the actions of others" (ibid, pp211-2). In the same way, our self-narratives make our lives intelligible to ourselves. As Anthony Rudd notes, understanding other people may be optional, but "having a more or less coherent narrative to tell myself (or others) is a part of what is involved in living my life as a self-conscious agent" (2007, p.62). Thus, we don't normally perform our actions without knowing why we're performing them, which we understand in terms of a narrative of some kind, albeit one which may often be very shortterm and of no lasting significance. When we occasionally do things without knowing why we're doing them, we tend to describe such behaviours in everyday life as being performed absent-mindedly – in other words as actions that seem to lack a place in any kind of narrative.

We do not begin life by creating our narratives. Rather we, as children, enter into a society and a set of stories that have been provided for us. As MacIntyre notes: "We enter human society... with one or more imputed stories – roles into which we have been drafted – and we have to learn what they are in order to be able to understand how others respond to us and how our responses to them are apt to be construed" (ibid, p216). As we grow older, we develop our own narratives to a gradually increasing extent. However, we are always constrained in the narratives we construct for ourselves, by the settings we live in, and by the narratives others have of us – we are "co-authors" of our narratives. Similarly, other people play inescapable roles in our own narratives, and we cannot avoid being aware to some extent, at least when we are not in a state of psychotic delusion, of what their narratives of us are likely to contain. This means among other things, as I argue below, that the speech act of

a psychiatrist in pronouncing a diagnosis on a patient is an important event in that person's life which may then lead to changes in her self-narrative.

One psychologist who has emphasised the importance of narratives is Jerome Bruner (1990). Where MacIntyre talks about the role of narratives in making their subjects' actions intelligible, Bruner explains how narratives create meanings for their authors. He describes this as follows:

... a narrative is composed of a unique sequence of events, mental states, happenings involving human beings as characters or actors. These are its constituents. But these constituents do not, as it were, have a life or meaning of their own. Their meaning is given by their place in the overall configuration of the sequence as a whole – its plot or *fabula* (1990, p43, italics in original).

This reflects MacIntyre's claim that actions only come to be intelligible when located within a narrative which links them together. Bruner likens self-narratives to autobiography which is an account given by a narrator who is also the protagonist. As such, it has the form of a story terminating in the present when the narrator and protagonist come together. The narrative does more than just recite a chronology of actions and events - it expresses a larger purpose and a moral framework for the individual. In the narrative of a person's life, or part of a life:

... the larger story reveals a strong rhetorical strand, as if justifying why it was necessary (*not* causally, but morally, socially, psychologically) that the life had gone a particular way. The Self as narrator not only recounts but justifies (ibid, p121, italics in original).

The story, however, is not fixed. It is always open to re-interpretation at a subsequent telling, whether this be to other people or to oneself. New and unexpected events will change the narrative and may also change the sorts of psychological justifications that the narrator invokes.

The Self as Bruner understands it is a complex notion and can be criticized for not being very clearly defined.<sup>53</sup> As the quote above indicates, he sees the Self as the person, or rather the psyche, that does the narrating. But the consequence of the narrating is to construct the Self, or more precisely to re-construct it. As such it is a dynamic Self, one that can be subject to modification with each re-telling of the protagonist's narrative, or segment of narrative. In any case, it seems clear that the Self is not conceived by him solely as some

\_

<sup>&</sup>lt;sup>53</sup> Bruner uses a capital letter in his discussion of the self. I therefore follow his practice when discussing his conception of it.

mental entity that emerges as the product of narrative, as if there is no Self existing prior to and independently of the narrative.

The process of narrative formation starts from a young age. Like MacIntyre, Bruner observes that children initially gain their narratives from their parents in the process of their acquisition of language. Narratives are expressed in language, and children learn their native language in part from the manner in which their parents use simple story-telling to comment on their activities and experiences. Children's narratives are further shaped by other social contacts, including in school, such that they come to adopt, in some form, the prevailing narratives in their culture. Bruner describes this as the "cultural shaping of meaning-making" (ibid, p.xii). Through the process of receiving culturally appropriate narratives and adopting them as their own, children gain a sense of their own identity. This gives continuing meaning to their lives and their actions, and it forms the basis of their own narratives into adulthood.

## 5.2.2 Narratives constituting the self

Some philosophers have emphasised the importance of narratives as in some way *constituting* the self. For example, Daniel Dennett (1992a) has described how one's self-narrative forms an ongoing autobiography with the self at its centre. The self, as Dennett conceives it, is not a metaphysical entity, but a useful abstraction, analogous to the centre of gravity of an object which fulfils a distinct explanatory purpose in physics. On this analogy, the self is seen as the centre of narrative gravity. Moreover, as he describes it, the process of creating stories is a "fundamental tactic of self-protection, self-control, and self-definition" which is how "we tell others – and ourselves – about who we are" (1992a, p418). The self thus defined is a fiction, but a useful fiction nevertheless. Elsewhere, Dennett elaborates on the idea of our stories as autobiographies, saying that we seem to be like "virtuoso novelists" (1992b, p114) in the way in which each of us tries to make the elements of our life come together into a single coherent narrative around the fictional protagonist: "The chief fictional character at the center of that autobiography is the self" (ibid, p114). However, Dennett notes, to try to know what the self really is to make a category mistake.

This view of the narrative self as fictional is criticised by Owen Flanagan (1994) for being misleading. He notes that authors of real fictional narratives have rather more freedom over what they create than we have in our self-narratives. We are always constrained in our self-narratives by the reality of our lives and we will be caught out if our self-narratives stray very far from this. Similarly we can deceive ourselves about who and what we really are. This is what gives meaning to the notion of self-deception. Hence, given the constraints on

self-narratives imposed by reality, the self at the centre of the narrative cannot be entirely fictional in most cases, unless the individual concerned is suffering from a serious delusion of some kind.

A stronger version of the narrativity thesis is offered by Marya Schechtman (1996) who sees the self that is created by narratives as more than a fictional self. Rather, she claims that the content of a self-narrative is what constitutes that person's identity.

... a person's identity is created by a self-conception that is narrative in form. Most broadly put, this means that constituting an identity requires that an individual conceive of his life as having the form and the logic of a story – more specifically, the story of a person's life – where 'story' is understood as a conventional, linear narrative (1996, p96).

It is not entirely clear what she means by *identity* here. At times she seems to mean that of psychological identity – i.e. who one thinks one is as a person. At other times, however, she seems to think of identity as more a question of persistence of the person over time – what Olson and Witt (2019) refer to as the persistence question.<sup>54</sup> She describes her view as the *narrative self-constitution* view, in that the self-narrative is what constitutes the *self* of the person concerned. Reflecting MacIntyre and Bruner, Schechtman also notes that narratives are what make episodes in a life intelligible:

To say that a person's life is narrative in character, then, is at least in part to claim that no time-slice... is fully intelligible – or even definable – outside the context of the life in which it occurs. To say that a person's *self-conception* is narrative is to say that she understands her own life in this way (ibid, p97, italics in original).

A person's self-conception, therefore, derives from her self-narrative, in which her experiences and activities are not viewed as isolated events, but as a connected series of episodes that form a meaningful story about her life. The sequence of episodes has to generally follow in a coherent and logical manner, in that her actions are understood in terms of her motivations and values. This does not, however, rule out occasional discontinuities or conflicts when the individual may be unsure of who she really is at certain points in time, such as in adolescent identity crises. Moreover, Schechtman acknowledges that people

\_

<sup>&</sup>lt;sup>54</sup> Schechtman appears to think that having a narrative is a condition for persistence of personal identity. However, I am not arguing that having a self-narrative is what underpins identity in a metaphysical sense – i.e. in the sense of specifying what it takes for the persistence of a person over time. This is not necessary for my thesis on the effects of a psychiatric diagnosis on one's self-narrative.

seldom narrate much of their lives in an explicit and self-conscious way. Nevertheless, in normal circumstances they have the ability to articulate their self-narrative, or relevant parts of it, when it may be desirable to do so. In less normal circumstances, such as when they are experiencing conflicts about their identity or mental health problems, this may become more difficult for them to do. At such times, the narratives may appear less coherent or logical than usual.

The claims Schechtman makes about the role of self-narratives in constituting personal identity are bold and have been criticized by Galen Strawson. I discuss his criticisms more fully in section 5.3. Briefly, however, he argues that not everyone lives their lives according to a narrative in the way that Schechtman claims, other than in a trivial sense. In response to this line of criticism, Schechtman modifies her claims to some extent in a subsequent paper (2007) by drawing a distinction between *persons* and *selves*. It is not self-evident, she says, that persons and selves are identical. There is therefore a corresponding distinction to be made between a narrative account of persons and a narrative account of selves. To constitute oneself as a *person*, on this account, one must have a narrative that recognizes oneself as a continuing individual with one's past and present experiences having implications for one's future behaviour. However, she says: "One need not deeply identify with past or future actions and experiences, care about them, or take an interest in them, but one does need to recognize them as relevant to one's options in certain fundamental ways" (2007, p170).

In contrast, Schechtman says, a narrative account of the *self* entails that one must have a much stronger sense of identification with one's past experiences and actions. Nevertheless, she does not explain very clearly what this stronger sense of identification actually means in practice. She suggests that this sense of identification occurs when one has what she terms "empathic access" to the experiences that make up one's narrative, but again it is not clear how this is to be determined. A sense of identification with or empathic access to one's past experiences would not seem to be a feeling with an all-or-nothing character about it. When I think back to experiences I had in childhood, how can I tell that I am having empathic access to them? Does it depend on how strongly or for how long I immerse myself in them? It is not clear how one could answer these questions. Rather, one might expect such feelings to be quite variable, such that one has a strong sense of identification or empathy with them on one

occasion and a much weaker sense on another.<sup>55</sup> Therefore, the distinction she makes between person- and self-narratives is not entirely clear.

# 5.2.3 Summary of the argument for narrativity

In summary, there is a considerable amount of overlap between the different accounts of narrativity outlined above in respect of the centrality that narratives play in shaping our understanding of who we are. However, there are different emphases in these approaches to narrativity. Moreover, as Rudd (2007) notes, there is not just one single approach to narrativity, but several which make quite varying claims. I have drawn a distinction here between those that emphasise the way in which narratives provide for meaning or intelligibility in the person's actions and those that focus more on the role of narratives in constituting or characterizing the self in some form. Where they agree is in the view that the formation of self-narratives in some form is a typical feature of human psychology for people with normal language abilities. If it is accepted that people have a need to make sense of their lives and of the sequence of events in those lives as they occur through time, then a narrative form would seem to be indispensable to this task. Narratives have the characteristic of being temporally organised (except in more sophisticated literary works, the typical narrative follows a chronological order) and are therefore suited to reflecting the unfolding of a life over time. Moreover, as Bruner explains, the manner in which a child learns his native language is closely bound up with the stories given to him about his activities by his parents and other care-givers. By understanding the language he hears around him and hearing the various stories in his social environment, he comes to understand how stories can be used to make sense of his own position in that environment. I claim, therefore, that narratives are an important means by which we make sense of our lives – they make the actions and events in our lives intelligible and meaningful. Without the events in our lives being linked in a narrative of suitable length and complexity, they would just be isolated events with little clear significance for us. Unimportant events happen in our lives all the time, but it's precisely because they are unimportant that we don't incorporate them into any sustained narrative. Hence we tend to forget about them.

-

<sup>&</sup>lt;sup>55</sup> There are debates in philosophy and psychology about what precisely constitutes empathy, but Schechtman does not make clear what her position is on this. However, I do not think it is necessary here to settle on any precise definition of it. It is enough for my purposes to stick with the common sense understanding of empathy to mean close emotional connection with another person. Schechtman's usage in this context is a little unusual, in that she seems to mean some kind of emotional connection with oneself.

A similar argument can be made that the psychological states we experience are made intelligible by their place in a narrative. The strength of the narrative approach is that it adds an extra dimension to the conception of human psychology described by folk psychology (e.g. Botterill & Carruthers, 1999; Fodor, 1987). According to folk psychology, inner mental states of beliefs and desires are organised in such a way as to have causal efficacy in the subject's actions. The beliefs and desires we have are naturally linked with the ongoing events in our lives which reinforce or change our beliefs and satisfy or frustrate our desires, and these events also give rise to new beliefs and desires. What makes an individual's beliefs and desires have any lasting significance beyond the moment in which they occur, in the sense of the significance they may have for the individual's life choices in the shorter or longer term, is the place they occupy in his self-narrative. Otherwise they can be little more than brief passing desires and beliefs.

Moreover, it is not necessary on this account to claim, contra Dennett and Schechtman, that a narrative constitutes the self. This point is emphasised by Tim Thornton (2010) who highlights what appear to be three sources of ambiguity in any such claim. First, he asks whether selves are the narratives themselves or rather the authors of the narratives and concludes they cannot be both. Second, he asks how narratives, if they literally are selves, can have any meaning. There has to be a self or embodied agent, he argues, to actually interpret the narrative and derive meaning from it. Without an interpreter, how can there be any meaning? Third, he points out that a narrative can hardly avoid including elements representing psychological states of various kinds. Given that it does so, it must therefore presuppose a concept of self that can experience such states. Without a self to embody them, it's hard to see how there can be any psychological states to comprise a narrative. Instead, Thornton argues, an account of narrativity that makes no such claims about the self (that Dennett and Schechtman make) is sufficient for an understanding of the psychological states and motivations of patients being assessed in mental health services. He describes his view of narrativity as follows: "I will take it to be the kind of understanding that connects together beliefs, desires, intentions and so forth in rational patterns" (p.258). In effect, therefore, this means that narrativity provides an additional dimension to folk psychology in that it makes intelligible the various individual beliefs and desires a person may have.

To see how an individual's beliefs and desires become intelligible when seen within an encompassing narrative, one can consider the following hypothetical example. One can imagine a medical student with aspirations to become a successful surgeon who desires to pass his degree with distinction. However, he believes he has just failed a critical exam and

knows that failure will entail a re-sit and possible failure again. He begins to wonder whether his desire to be a surgeon is a realistic goal for himself. When the results are given out, he finds out that he has passed after all and his life plan is consequently still on track. The various beliefs and desires he has during this episode are made intelligible by the self-narrative he has. This may, for example, include the expectations placed on him by his parents, one or both of whom may be senior medical professionals. This helps to make his desire to be a successful surgeon fully intelligible in the context of his life. Similarly, his belief about his expected exam failure has a particular meaning for him given by his self-narrative about his aspirations and the expectations placed on him. The various emotions he experiences at each point in this sequence of events are also intelligible in light of his self-narrative. Without the narrative that linked the relevant beliefs, desires, events and associated emotions, it would not be clear what the significance of each of this student's mental states and emotions would be.

However, this view is open to the objections that self-narratives are not as typical as this view implies, or at least that any such claim for narratives is merely trivial, and that narratives are not essentially constitutive of selfhood or personal identity. I therefore discuss these objections in the next section. Before doing so, however, it is important to note that narrativity does not demand that individuals should have a whole-life narrative, although some narrative theorists (e.g. Schechtman, 1996) seem to imply this. It would of course be impossible for any individual to have such a narrative, because of the enormous amount of time needed to narrate one's whole life either to somebody else or even to oneself. Instead, as David Lumsden (2013) argues, there may be a multiplicity of shorter narratives available to the individual at either a conscious or unconscious level – what he calls narrative threads. The narrative threads may connect up to represent the person in some way, he argues, but there is no need to conclude that they form some kind of master narrative, nor that they constitute a metaphysically realist self. It is plausible, however, to argue that invoking narrative threads is what we seem to do when we want to explain some aspect of our lives. The narrative thread may be as long or as short as the specific circumstances require and display an appropriate degree of complexity. In the case of an individual receiving psychotherapy, for example, she is likely to be prompted by the therapist to produce quite a lengthy narrative to help her locate the source of her psychological distress in the relevant part(s) of that. This approach to narrativity views it as a psychological thesis, not a metaphysical one about the constitution of the self. Since my concern in this chapter is the effects of a biomedical narrative on an individual's self-narrative in cases where the person is

suffering from psychological distress or disorder, it is not necessary for me to argue for any claim beyond the psychological one about the function that narratives have in people's lives.

# 5.3 Strawson's objections to narrativity theory

Galen Strawson (2004) has raised objections to narrativity theory which can be understood as being directed at both sets of accounts of narrativity described above. He identifies two distinct narrativity theses and argues that both are false. First, he criticizes what he calls the psychological Narrativity thesis (PNT)<sup>56</sup>, which expresses the view that human beings typically experience their lives in the form of a narrative or a set of narratives. The manner in which he characterises this is not completely precise and can be understood to encompass both families of accounts I have described. Second, he criticises what he identifies as the ethical Narrativity thesis (ENT), which claims that experiencing one's life in a narrative form is in some way a good thing and contributes to a well-lived life. I will restrict myself to examining his criticisms of the PNT, since this is my focus in this chapter. I am not making any attempt to defend the ENT, which is not necessary for my thesis. For the remainder of this section where I refer to the *narrativity thesis*, I intend this to mean the PNT as Strawson characterises it, except where a more precise description is necessary. Strawson's objections to the PNT essentially comprise two elements. First, he argues that the PNT is not true for everybody – it is not a universal aspect of human nature. Second, he argues that, to the extent that everybody can create some form of narrative about themselves, this is a trivial thesis. I will address each of these aspects in turn in sections 5.3.1 and 5.3.2.

In addition to Strawson's objections, there have been other critics of some of the claims of narrative theorists, particularly in respect of claims that narratives constitute the self or personal identity in some way. Thus, John Christman (2004) rejects the notion that personhood depends upon narrative unity. As such, he seems to be criticizing a strong version of narrative theory in which the unity of the narrative is what is needed to constitute the self. Karsten Witt (2020) criticizes the view that our identities are constituted by narratives, which he attributes to narrative theorists.<sup>57</sup> He states that this is a "basic tenet" of narrativity.

\_

<sup>&</sup>lt;sup>56</sup> Strawson says he uses a capital letter for Narrativity "to denote a specifically psychological property or outlook" (p.428). However, it is unclear what he means by this, and I will therefore use lower case letters for these words from now on.

<sup>&</sup>lt;sup>57</sup> Witt actually refers to "characterization identities", although his usage seems to broadly coincide with the usual psychological concept of *identity* – i.e. who we think we are – although this is not a precisely defined word.

However, neither MacIntyre nor Bruner can be accurately described as holding such a view. In any case, I am not defending the strong views of narrativity which Christman and Witt are criticizing, as these are not necessary for my thesis about the effect of a psychiatric diagnosis on an individual's self-narrative. Consequently, I will not discuss them any further here.

# 5.3.1 The psychological narrativity thesis (PNT) is not true of everybody

In support of his claim that not everybody conceives of their life in narrative form, Strawson states that he regards himself as an exception to this supposed rule. He says:

I have a past, like any human being, and I know perfectly well that I have a past. I have a respectable amount of factual knowledge about it, and I also remember some of my past experiences 'from the inside', as philosophers say. And yet I have absolutely no sense of my life as a narrative with form, or indeed as a narrative without form (2004, p.433).

Thus, being fully aware of his past history does not entail that he must have any kind of narrative formulation of his life. In order to illustrate why he thinks the PNT is false, he quotes the novelist, Henry James, who said about one of his earlier works: "I think of... the masterpiece in question... as the work of quite another person than myself... a rich... relation, say, who... suffers me still to claim a shy fourth cousinship" (1864-1915/1999, p562-30). Therefore, Strawson argues, James did not seem to experience his own life in the kind of narrative form implied by the PNT.

Strawson goes on to say: "it seems clear to me, when I am experiencing or apprehending myself as a self, that the remoter past or future in question is not my past or future, although it is certainly the past or future of GS the human being" (ibid, p.433). The implication seems to be that there are different ways of experiencing oneself and that doing so "as a self" is just one of those. In order to try to explain this, Strawson introduces a starred pronoun, 'I\*', which is intended to represent "that which I now experience myself to be when I'm apprehending myself specifically as an inner mental presence or self" (ibid, p433). Similar meanings are attached to 'me\*', 'my\*', 'you\*', 'oneself\*', and so on. In this respect, he agrees that events in his remoter past did happen to him, but those events are not experienced as 'his\*' past. However, he does not make clear what is involved in apprehending or experiencing oneself as an inner self, as opposed to any other mode of experiencing, nor does he give much indication of how one can tell that one is experiencing oneself in this particular way. How do I know, when I think of parts of my life in the past, that I am apprehending myself as 'I\*', rather than merely as 'I'? Of course, I can think of

myself as some kind of inner self if I so choose, but it is not clear why events in my remoter past should not have happened to 'me\*' as well to 'me'. Moreover, it is not necessary for the narrativity thesis, except perhaps to Schectman's rather strong version of it, that events in Strawson's past should have happened to 'him\*' as well as to 'him'. It is not implied by the accounts of narrativity described by MacIntyre and Bruner. In other words, neither of these writers makes any distinction about the manner in which one recalls past events, in the way Strawson does. It is enough to simply know about one's past for this knowledge to be part of one's self-narrative.

Despite criticizing the PNT, Strawson does not deny that some people very likely do experience their lives as in some way an unfolding narrative. His claim is that not everybody does so. In order to substantiate this, he suggests that people can be understood as dividing into two kinds in this respect which he labels diachronic and episodic – he calls these "styles" of temporal being" (ibid, p430). To be diachronic is to conceptualise oneself as a self that was there in the (further) past and will be there in the (further) future. Diachronics are likely to understand their lives as having the form of a narrative. By contrast, to be episodic is not to conceptualise oneself in this way. Instead one has little sense that the self that was there in the past will also be there in the future, although one is fully aware of one's continued existence as a human being through time. Episodics are not likely to experience their lives in a narrative form. He describes these outlooks as "radically opposed, but they are not absolute or exceptionless" (ibid, p430). Therefore, episodic individuals may occasionally feel a very strong emotional connection with events in their past, as though those events were very recent in terms of their impact. Similarly, diachronics may at times fail to experience any strong emotional linkage with events in their past when they recall them.<sup>58</sup> Strawson notes, therefore, that the episodic/diachronic distinction does not correspond exactly with the narrative/non-narrative distinction, although he says the two are correlated. Nevertheless, he regards himself as an episodic and as being non-narrative in his attitude.

The manner in which Strawson describes the diachronic and episodic styles of temporal being suggests that they can be understood as personality traits. He says: "I take it that the

-

<sup>&</sup>lt;sup>58</sup> Indeed, one can add that even most diachronics, as understood in Strawson's terms, must surely fail to experience significant emotional connections to events in their past when they recall them much of the time. It would seem to be difficult to think and talk objectively about one's past if one experienced strong emotional reactions every time one does so. Strawson, of course, refers to these modes as "styles of temporal being" which does not necessarily imply that the distinction between them is grounded in emotional reactions. However, it is difficult to understand what this distinction does consist in, if not some kind of emotional response to aspects of one's past (and future) life.

fundamentals of temporal temperament are genetically determined, and that we have here to do with a deep 'individual difference variable', to put it in the language of experimental psychology" (ibid, p431). Consequently, this would seem to be the kind of question that would be amenable to empirical investigation by experimental psychologists working in the field of individual differences. However, Strawson does not cite any empirical evidence to support the idea that styles of temporal being can be understood as psychological characteristics which occur in different proportions in different people, despite his claim that these seem to be fundamentals of human temperament. It is difficult, therefore, to know whether we should accept that these styles really are distinctive personality traits as he seems to think. Moreover, if they do represent actual personality traits, then we would expect them to be characterised as variables which fall on a continuum of values. Thus, some people would be more diachronic in their styles and others more episodic, rather than there being two distinct categories where every individual is unequivocally characterised as one or the other. This does seem to be Strawson's view, since he also states: "one's exact position in Episodic/Diachronic/Narrative/non-Narrative state-space may vary significantly over time according to what one is doing, thinking about, one's state of health..." (ibid, p413). In other words, the episodic and diachronic thinking styles do not seem to be as fixed as he initially seems to suggest. It is difficult to see, therefore, how this distinction supports his argument that people can be characterised in terms of whether they are 'Narratives' or 'non-Narratives'.

Strawson's critique of the PNT seems to be based on the idea that there are distinctive differences between individuals in terms of their propensity to apprehend their lives in a narrative form, and in addition in terms of their styles of temporal being. Nevertheless, as the last quotation in the previous paragraph indicates, Strawson does not seem to think individuals are fixed either in their styles of temporal being or in their tendency to view their lives in narrative terms. When he criticizes the PNT, he appears to think that those endorsing the thesis claim that having a narrative view of one's life implies that one must also feel some kind of close connection with one's more distant past when one is thinking about one's self-narrative. He says:

...for a life to be a narrative in the required sense it must be lived Narratively. The person whose life it is must see or feel it as a narrative, construe it as a narrative, live it as a narrative (ibid, p440).

Such a view seems to form part of Schectman's conception of narrativity, but this is not an essential feature of the narrativity thesis. It does not feature as an essential element in Bruner's and MacIntyre's account, for example. There is nothing in their accounts that

demands that the individual must "feel it as a narrative", although some people may well do so. Therefore, even if we accept his claim that he does not apprehend his life in this way – in the way of feeling it as a narrative – and consequently that it is not true of everybody, he has not given us a good reason to reject the PNT in his objections so far.

### 5.3.2 If the psychological narrativity thesis is true at all, it is a trivial thesis

While Strawson says that the PNT is false in any non-trivial sense, he does concede that we can all have narratives in a more trivial sense. He agrees that we can all create narratives about what has happened to us, in the sense that we know what has happened to us in the past and what we hope for in the future. He regards this kind of recalling of events as trivial, in the same sense that a narrative one might make about the steps involved in making a cup of coffee is trivial. However, this is clearly a misleading example of a possible narrative, because the event in question is insignificant in the way that a story of an individual's life is not. To cite such an example of a narrative does not seem to be sufficient to render the narrativity thesis trivial in its entirety. Moreover, it could be argued that the narrative of 'making a cup of coffee' is what, in MacIntyre's terms, makes the apparently disconnected behaviours of filling a kettle, putting coffee granules in the cup, and so on into the *intelligible* actions of coffee-making.

Rudd (2007) points out that Strawson tries to isolate such apparently trivial short-term narratives from longer-term narratives in which they may be embedded and in so doing he misses the significance of the longer-term narratives. There may be all sorts of such narratives within which the coffee-making takes place. It may be a larger narrative in which my making coffee at this time maintains the regular structure of my day. Alternatively, I may be making it to waste time and avoid doing a difficult task, or I may be trying to impress a friend with my ability to make a particularly good cup of coffee. Clearly many more extended narratives are possible. By separating out such everyday narratives in an arbitrary manner from any larger scale narrative, the narrativity thesis can be made to seem trivial. However, Rudd argues, placing them in a larger narrative context is what can reduce the apparent triviality of the narrative of coffee-making.

The common theme among the various accounts of narrativity described above is that the narratives of our lives are closely bound up with who we consider ourselves to be. This is not to claim that our narratives fully constitute our *selves*, as Schechtman argues, but rather to

say that they play an important role in shaping our self-concepts.<sup>59</sup> It is difficult to see how Strawson could deny that in his own case. As argued in the previous section, to the extent that he has a full awareness of his own past and his future hopes and expectations, he has a self-narrative which to some extent is likely to exercise an influence on his beliefs, actions, and values. Whilst this conception of narrativity need not be as strong as to imply that the narrative constitutes the self, it is difficult to see why it should be regarded as trivial. Insofar as one's self-narrative shapes one's self-concept, this would seem to be a psychologically significant thesis. Moreover, it is not clear whether, when Strawson claims it is a trivial thesis, this is merely because he finds it uninteresting in any form other than the strong form in which Schechtman expresses it. If so, it does not follow that it is trivial in the way that Strawson believes it to be, simply because he might find it uninteresting. What I argue in my thesis is that theories of narrativity are not trivial in the context in which a biomedical narrative may be given to the mental health service user. Rather, I argue that narrativity can help us to understand how the receipt of a psychiatric diagnosis may affect the recipient.

# 5.4 The effects of a psychiatric diagnosis

There can be a range of psychological effects produced by a psychiatric diagnosis on the recipient. In the first part of this section, I review some of the effects that are reported and argue that these are a consequence of changes that the diagnosis can effect on the recipient's self-narrative. Such changes follow from the fact that we are not the sole authors of our narratives. Rather the narratives we have for ourselves are strongly influenced by a range of external factors, including the events that befall us and the opinions that other people hold of us. In the second part, I argue that the diagnostic process by which DSM diagnoses are reached and the symptom-based categories on which they are based make such effects likely to occur.

# 5.4.1 What kinds of effects may be produced in the recipient?

The experience of a severe or chronic illness is likely to have a significant impact on the self-narrative of the sufferer. The psychiatrist and anthropologist Arthur Kleinman describes in his book 'Illness Narratives' (1988) many cases from his clinical experience of people with

-

<sup>&</sup>lt;sup>59</sup> I am using the notion of self-concept here as defined by Roy Baumeister (1999) to mean "The individual's beliefs about himself or herself".

chronic diseases and the narratives of their illnesses they created. An important influence on the patient's narrative is the diagnosis given to them. Kleinman describes diagnosis as a semiotic activity, in which the patient's complaints are translated into a diagnosis by means of the signs or biomedical indicators of the disease named by the diagnosis. The diagnosis alters the meaning of the illness for the patient. One way in which this can happen is through the stigma that some diagnoses carry. An example that Kleinman cites is that of leprosy. The visible marks of the disease provide the basis for the stigma perceived by the social world of the sufferer. In addition, the knowledge that he has the disease creates an internalised stigma that the sufferer can also experience. The stigma, therefore, need not just be the societal reaction generated but also the sufferer's own acceptance of the stigmatised identity. A more recent example of a diagnosis inducing this kind of effect is that of HIV/AIDS where many of those affected have reported feelings of internalised stigma.

I argue that this can equally be the case for mental disorder diagnoses. As the accounts of narrativity I have discussed above indicate, we are very far from being the sole authors of our self-narratives. These are heavily influenced by the social world in which we move and which in many ways defines who we are. Unless I am seriously deluded, I cannot incorporate being a concert pianist or the president of the USA into my self-narrative – the social environment in which I live could not support any such elements in my self-narrative. Even at the level of assessing what kind of person I am, I cannot disregard the opinions of others close to me, even if I might disagree with them in specific respects. Hence, the manner in which my self-narrative helps me make sense of my life is partly a function of events happening in my social environment which I may, or may not, have some control over. In the case of a patient in the mental healthcare system, the stated opinion by the clinician of the patient's diagnosis will be one of those events, and indeed a particularly powerful one in view of the epistemic authority of the clinician and the institutional fact created by the diagnosis. 60 The biomedical narrative associated with the diagnosis is therefore likely to be influential in its effects on the patient's self-narrative. An example of this is expressed by one individual talking about his diagnosis of bipolar disorder in a video recording of service users' reactions to their diagnosis: he says "people suddenly realised I wasn't doing things for attention... It was because I have a brain disorder" (BBC, 2018). The diagnosis therefore allows him to adopt a biomedical narrative to explain his unpredictable and extreme mood swings, and in this way it alters the meaning of this condition for him.

-

<sup>&</sup>lt;sup>60</sup> As discussed in Chapter 4.

Note that, as the above example indicates, I am not claiming that the biomedical narrative associated with the diagnosis is one that the patient *invariably* finds unhelpful or stigmatising. Rather, the effect of receiving such a diagnosis can be very varied, as a range of surveys has shown (e.g. Perkins et al., 2018). There is some evidence that the nature of the diagnosis influences whether or not the recipient views it in negative terms. Thus, a survey by Thomas et al (2013) of people who had received a diagnosis of schizophrenia revealed a substantial amount of negative reactions. They report that many people found the diagnosis stigmatising. In some cases, respondents reported that the diagnosis felt like "a life sentence from which there was no recovery" (p.137). In contrast, very few people reported finding that diagnosis helpful.

However, describing such effects on the sufferer as internalised stigma is liable to gloss over the quite varied negative reactions that individuals will experience. The manner in which a psychiatric diagnosis can impact on an individual's self-narrative is discussed in more detail by Şerife Tekin (2011). She reiterates the features of narrativity discussed by other philosophers, and in addition emphasises the distinction between the narrative authored by the individual and that received from their social environment. This distinction can assume particular salience when an individual receives a diagnosis, whether medical or psychiatric.

Tekin argues that a DSM diagnosis can function as a source of narrative for the person concerned. She suggests that, in some cases, the patient may be comfortable in understanding their experience as some kind of illness in terms of an established medical diagnostic category. In other cases, however, patients may find that the diagnosis imposed upon them prevents them framing their experience in any other way than as a kind of neurochemical imbalance, which becomes the dominant narrative. Tekin argues that this kind of narrative may impede the individual from developing sufficient cognitive and affective resources of her own that could help her recovery in future. To the extent that she is encouraged to think of her psychological states as merely a function of unbalanced brain chemistry distinct from the environmental and social contexts that may have precipitated her condition, she may be led to question the reality of her own experiences, with a consequential loss of self-respect and feelings of agency. The belief that her psychological states are outside her control may undermine her sense of autonomy and responsibility. This may particularly be the case when patients are led to believe that their condition is a chronic one. The biomedical determinism which such a narrative implies can limit the hopes for recovery among these patients, and the disempowering self-narrative generated can become self-reinforcing (Yanos et al., 2010). If patients come to believe that their condition is entirely, or even partially, caused by an

underlying medical condition which they cannot understand, they are liable to feel there is little they can do to counteract its effects.

The feelings of agency in the individual, which Tekin argues may be impaired by the receipt of a psychiatric diagnosis, can be understood as similar to the concept of *self-efficacy*, described by the psychologist Albert Bandura in many publications (e.g. 1982, 1995, 1997). The concept was developed in a large number of empirical studies in which he and his coworkers examined the importance of self-efficacy in many different situations that people typically face in their lives. This can be understood as the appraisal by the individual of her or his ability to carry out given tasks or overcome specific obstacles to achieving intended goals. Bandura defines self-efficacy as follows: "Perceived self-efficacy refers to beliefs in one's capabilities to organise and execute the courses of action required to produce given attainments" (1997, p.3). It is situation-specific, in that the individual may feel a high level of self-efficacy in tackling one activity but a much lower level in another. Examples of the situations in which the effects of self-efficacy have been studied include the academic performance of students, the parenting behaviour of parents, the manner in which people recover from heart attacks, and many others (Bandura, 1982, 1995). It also affects the manner in which people cope with the consequences of psychological disorders (Bandura, 1997).

Importantly, the individual's perceived self-efficacy does not necessarily correspond with his or her actual ability to perform the given activity. Moreover, self-efficacy can become over-generalised by some individuals to the extent that they either feel themselves incapable of doing many activities successfully or, at the other extreme, become over-confident about their ability to do too many things. Bandura (1982) notes that self-efficacy mediates between the relevant knowledge gained by the individual and the behavioural output that may occur as a result of that knowledge. It is a mental state of belief that may be conscious or unconscious. For frequent and effortlessly performed activities it is usually unconscious, whereas for more challenging or unfamiliar tasks the individual is likely to consciously assess her ability to perform it.

Patients suffering from psychological disorders and distress are likely to experience diminished self-efficacy, purely by virtue of their disordered emotions and thought states. The claim by Tekin that a psychiatric diagnosis can lead to a loss of feelings of agency can be understood as meaning that the diagnosis can lead to a further diminution in the patient's sense of self-efficacy. What this means is not that the patient can suffer a loss of agency *per se*, but rather a reduction in her *feelings* of agency. Bandura, explaining that perceived self-efficacy is central to human agency, states: "If people believe they have no power to produce

results, they will not attempt to make things happen" (1997, p.3). As noted, self-efficacy is a belief state which can therefore be expected to feature in the patient's self-narrative, either in relation to specific activities or more generally across a broad range of activities or life-tasks. The biomedical narrative associated with the diagnosis conveys the implication that the patient's condition is a consequence of some unspecified biological process, which is something that is seen by the patient as not obviously under her control. The self-narrative therefore includes an element that can run something like this: "I am the way I now am because something I neither understand nor am able to control has happened to me". In addition, the biomedical narrative may often convey the implication that the condition is a long-term one, thus reinforcing the loss of self-efficacy experienced.

In another paper, Tekin (2014) develops this further by arguing that a DSM diagnosis may contribute to a diminution of the patient's self-insight. She conceives of self-insight as the degree of understanding that individuals have regarding their sense of their own identity, their mental states (particularly those that are confusing or distressing), and their interpersonal relationships. To have a well-developed degree of self-insight is likely to contribute to a high quality of life and personal flourishing, as well as the ability to meet the various demands and challenges that one encounters in life. Inasmuch as the statement of the diagnosis focusses the recipient's thinking on the notion that his distressing experiences are caused in some way by the condition that the diagnosis denotes, he is liable to lose some degree of self-insight into what might have led him to be in the state he find himself in. As argued in Chapter 2, a DSM diagnosis does not typically convey any information about the explanation of the patient's condition, although it may be believed to imply some sort of brain disorder. Nevertheless, it may direct the patient's attention towards a supposed biomedical conception of his condition and away from the actual set of experiences, with a likely history of long duration, which have actually contributed to his current psychological state.

Another way of conceptualising such effects is described as *self-illness ambiguity* by John Sadler (2004, 2007). This can arise, according to Sadler, when the receipt of a psychiatric diagnosis leads the recipient to question whether what she is experiencing is her personal 'self' or her illness. Sadler notes that the word 'self' has many meanings in Western culture. In this context, he says he intends it to refer to the common sense notion that people typically have of it, in which it comprises feelings of agency, identity, a unique personal history, an expected life-trajectory into the future, and a distinct personal perspective on the world. These aspects of the self, as Sadler describes them, are loosely specified, but taken

together they resemble what psychologists refer to as an individual's self-concept – i.e. how the individual sees him or herself. The ambiguity between self and illness that Sadler describes refers to cases where patients ask what has "come over" them. He explains: "In these cases, one's personal self may be fully intermingled with one's illness; the boundaries of self and illness are not clear" (2007, p.115). The patient therefore has difficulty knowing how to understand his mental states. Are they part of who he believes himself to be or are they part of the illness? Sadler notes that "Patients often ask their doctors, 'Is this me or is this my disorder' " (ibid, p.118). This seems to suggest some degree of confusion in their self-narratives and an impairment in their self-insight, at least in some cases, as Tekin argues in her paper. The mental distress experienced by the patient would partially account for this, but the added effect of the diagnosis with its biomedical implications would further reinforce the idea that his mental states are the consequence of a presumed illness entity.

In summary, the biomedical narrative associated with a DSM diagnosis can have a significant impact on the patients' self-narratives with corresponding effects on their psychological states. In particular, it can lead to a diminution in their feelings of agency and hopes for recovery. A further consequence is that it can leave patients confused about the source and meaning of their distressing experiences and conflicting emotions. In the next section, I discuss the features of DSM-based diagnoses that can lead to such effects.

## 5.4.2 How can psychiatric diagnoses generate such effects?

It is plausible that psychiatric diagnoses are much more likely to have such an effect than diagnoses of somatic conditions. For chronic or life-threatening diseases the individual's self-narrative is bound to be altered to some degree, possibly to the extent of requiring the sufferer to re-conceive the meaning of her life in the most severe cases. This does not necessarily result in any loss of self-insight or experience of self-illness ambiguity, other than by the direct consequences of the pain or suffering caused by the disease. In the case of psychiatric diagnoses, however, the diagnosis can have additional significance. In their systematic review of published surveys of the experiences of mental health service users, Perkins and coworkers (2018) note that the disclosure of the diagnosis frequently emerged as "a pivotal moment" for the patient (p.9). Moreover, as Maung (2019) notes, psychiatric diagnoses are often communicated as if they refer to a hidden disease process that explain the symptoms, despite the failure so far to show that they are explanatory (as discussed in Chapter 2) and even if the clinician does not consider them to be explanatory. Patients understand that their diagnosis signifies something important. Although they are unlikely to think of it specifically

as constituting an institutional fact, in the terms discussed in Chapter 4, they will nevertheless recognise it as something with a formal or official status recorded in their case file.

Therefore, they are likely to understand the diagnosis as saying something important, albeit inexplicable, about themselves. To the extent that a diagnosis does this, it is likely to change the degree of insight that the recipient has into her mental states and her emotions, such that her self-narrative is changed by the biomedical narrative the diagnosis implies.

The manner in which clinicians interview their patients is likely to contribute to this effect. Reliability of the diagnostic categories is one of the main virtues claimed for the DSM nosology. In order to maximise the reliability of these diagnoses, best practice requires that diagnostic interviews be conducted according to a standardised system. The typical purpose of these interviews is to establish whether the patient's symptoms conform to one diagnostic category or another. As a consequence, the personal meanings that the symptoms may have for the patient are liable to be neglected, particularly when, as is often the case, the clinician is working under considerable time pressures. Giovanni Stanghellini (2004) observes that the use of standardised psychiatric interviews, which consist of a set of prescribed questions, are likely to contribute to this tendency. He notes that the meaning of a symptom is interpreted by the clinician in terms of the properties corresponding to a given category, leaving little space for meanings and narratives that are more salient to the patient. He also emphasises how narratives play a central role in creating coherent meaning for people in their lives. If the manner in which the psychiatric interview is conducted ignores this at a time when the patient concerned is experiencing severe distress, and instead gives her a biomedical narrative, her confidence in her own psychological resources for making sense of her experiences is likely to be diminished.

Diagnostic interviews do not typically take the form of a normal conversation, in which the participants may elaborate on their statements to whatever extent they wish to attain a degree of mutual understanding. Rather, such interviews follow what resembles a stimulus-response process in which the clinician's question constitutes the stimulus and the patient makes some kind of response which is expected to answer the question in some way. As Stanghellini argues, the nature of the interview requires that answers be given in either a "yes-no" form or in a relatively circumscribed manner pertaining to a narrowly specified question. Thus the larger narratives that the patient might want to convey will not be purposely elicited. The purpose of the interview is to arrive at a diagnosis, rather than to allow the patient to articulate a richer and more complex narrative.

This process can constitute a kind of epistemic silencing of the patient. The conversation between the clinician and the patient is not an equally balanced one, in that the patient's contribution to it is restricted by its structure. Any contribution the patient might want to make will only be incorporated into the diagnostic decision making if it contributes in some way to that end. It is not the specific intention of the clinician to restrict the participation of the patient in this way, but a consequence of the manner in which DSM-based diagnoses are arrived at, compounded by the time pressures faced by clinicians in most mental health service systems.

A similar analysis is offered by Thomas Fuchs (2010) who identifies three approaches to the assessment of mental disorders. These he describes as, respectively, the positivistic objectifying approach, the phenomenological subject-oriented approach, and the hermeneutical intersubjective approach. The first of these, the positivistic approach, is essentially that which is characterised by the DSM diagnostic system. His description and critique of this approach reflects Stanghellini's, in that the need to focus on a defined list of potential symptoms of a diagnostic category in order to make a diagnosis precludes a thorough exploration of the patient's subjective experience and the complexity of his narrative. A consequence of this is that a large amount of subjective experience, including personal meanings, self-concept, and subtle changes in emotional states, can be marginalised from the assessment process. While Fuchs does not completely clarify the difference between the second and third approaches, as he describes them they are both concerned with placing the focus of the assessment on the patient's subjective experience. The difference is that the hermeneutical intersubjective approach, as he sees it, is more concerned with the patient's manner of interacting with others, which would be assessed within a psychodynamic form of psychotherapy. In order to fully capture the complex characteristics of the patient's narrative, he argues, assessment needs to conducted in accordance with all three approaches. Otherwise, the positivistic, DSM-based approach on its own will fail to capture important features of the patient's self-narrative.

The manner in which assessment conducted in accordance with the DSM diagnostic categories marginalises the patient's narratives is described as *hyponarrativity* by Sadler (2004). The prefix 'hypo' indicates a state that is under an expected level or lower than normal. In this context, it is intended to signify the manner in which DSM diagnoses underrepresent the patient's narratives. Thus Sadler observes that a DSM diagnosis fails to characterise or do justice to the reality of patients' narratives and the meanings these have for the individual. Rather, the DSM system directs attention to common symptoms, as specified

in the diagnostic criteria for each diagnosis, and the patients' stories that do not immediately fit with these symptom descriptions are down-played. As Sadler notes: "The DSM's descriptions have no plot lines, no particularized conflicts with unique others, no climax, no denouement" (p.177). In other words, the DSM descriptions overlook precisely those features that make up an individual's self-narrative and that form the basis for the meanings the individual finds in her life. They offer no kind of rich or extended narrative beyond the implication that the patient's condition is the result of some kind of unspecified biomedical process.

#### 5.5 Conclusion

I have argued that the creation of narratives is a fundamental feature of human psychology and these narratives function as a source of meaning in the lives of their authors. The extent to which individuals will understand their lives in narrative form will vary from person to person, and it is more plausible to regard such narratives as threads of greater or smaller lengths, rather than as complete whole-life narratives. While some theorists regard these narrative constructions as constituting the self or identity of the person, such a view is not held by all narrative theorists. In particular, it is plausible that individuals find meaning in their lives, at least in part, by means of their self-narratives, without claiming that these narratives constitute the self.

When people present themselves to mental health services, they are suffering from significant, and sometimes severe, levels of mental distress. The narratives, or shorter narrative threads, by which they try to make sense of their experiences will tend to be confused and disordered. The receipt of a psychiatric diagnosis, which in many cases will be communicated to the patient as representing some kind of medicalised entity which purports to explain his distress, offers a biomedical narrative about his condition which is superimposed on his previous narrative or may replace it altogether. This has the potential to reduce his insight into his own condition and his feelings of autonomy and responsibility, despite the fact that DSM diagnoses are not explanatory.

Alternative approaches that can mitigate or avoid such a consequence make greater use of formulation, without the imposition of diagnoses that convey a biomedical narrative (e.g. Johnstone & Dallos, 2014). Broadly speaking, a formulation is a detailed psychological hypothesis about the causes and maintaining factors that are responsible for the patient's condition and which is arrived at in a collaborative conversation with the patient. Reinforcing

the individual's psychological resources to help her develop her own self-narratives, such that she feels more empowered to overcome her difficulties, is a key aim of formulation in psychological therapy. Such an approach involves helping the patient to articulate her feelings and experiences in more depth and to gain a greater understanding of how she might have come to be in this condition – that is, to have a story about it that she can understand and assent to. Treatment and therapy is then determined on the basis of such a psychological formulation. Cognitive behaviour therapy (CBT) is now a frequently used set of methods for people with psychotic symptoms (e.g. Hagen et al, 2013) and its success depends upon an adequate formulation having been agreed between patient and therapist.

The process of formulation seems to reflect the approach of narrative medicine, as described by Rita Charon (2006) for example. In her description of this, medical assessments and treatment need to be individualised to the patient and based around his particular narratives. She sees this approach as applicable to medical practice in general, not just to psychiatry, and contrasts this with what she calls the "logico-scientific" approach of most contemporary medical practice. This is open to the objection by Miriam Solomon (2015) that the two approaches are not mutually exclusive, and more specifically that contemporary scientific medicine is well equipped to take account of individual differences between patients. However, this objection seems less applicable to psychiatry where the diagnostic categories have little scientific validity, and where treatment needs to be particularly tailored to the individual's own narrative. Solomon notes, for example, that psychodynamic psychiatrists are among those specialists who have developed the narrative medicine approach, since such an approach is central to psychodynamic psychotherapy. Similarly, narrative therapy is another approach to therapy which aims for a specifically narratively framed formulation on which therapy is explicitly based (e.g. Harper & Spellman, 2014). These kinds of methods can be employed without any formal diagnosis being included.

In the following chapter, I will cite some excerpts of patients' reports and argue that, as a consequence of the effects of the diagnosis on her self-narrative, the patient may become a victim of hermeneutical injustice.

## Chapter 6

# Hermeneutical Injustice as a Consequence of a Psychiatric Diagnosis

### 6.1 Introduction

I have argued in Chapter 2 that psychiatric diagnoses based upon the DSM do not represent natural kinds, except in a small number of cases where the disorder is explicable by biological processes – e.g. Alzheimer's Disease. This is in contrast to somatic diseases which, as argued in Chapter 1, can in most cases be considered as natural kinds in medical science, in that they are characterised by clusters of properties which reflect the causal structure of the world and have explanatory value (Boyd, 1991, 1999b; Khalidi, 2013, 2018). Despite this, as I argued in Chapter 3, psychiatric diagnoses are frequently conceptualised within a biomedical or disease model. In Chapter 4, I argued that the statement of a diagnosis in medicine constitutes a perlocutionary speech act and that one consequence of this speech act is to create an institutional fact about the patient's diagnosis. This applies as much to psychiatric diagnoses as to other medical diagnoses. Following from this, I argued in Chapter 5 that the receipt of a psychiatric diagnosis, by virtue of its institutional status and its associated biomedical narrative, can have a significant impact on the patient's self-narrative in such a way that patients can come to conceive of their psychological difficulties as a function of some unexplained biological process in their brains. A consequence of this is that their feelings of agency and their hopes for recovery can become diminished. In this chapter, I further elaborate on this idea and argue that, because of such consequences, the receipt of a psychiatric diagnosis can lead to the recipient becoming a victim of epistemic injustice, and specifically of hermeneutical injustice. I argue that, where this happens, what constitutes the injustice is the effect that the diagnosis has on the recipient in her or his particular case.

I start in section 6.2 by giving a brief overview of the concept of *epistemic injustice*, as introduced by Miranda Fricker (2007). I then, in 6.2.1, discuss her concept of *hermeneutical injustice* in more detail and argue that this can be understood more broadly than Fricker seems to imply in her book, "Epistemic Injustice: Power and the Ethics of Knowledge". In particular, it can be understood as applicable, not just when suitable hermeneutical resources for understanding a given predicament are unavailable to the individual, but also in cases where he or she may previously have had some such resources available, but these have been

obscured or diminished by new circumstances. This, I shall argue, can be the case when people receive a psychiatric diagnosis.

In section 6.3.1, I discuss how epistemic injustice can occur in healthcare contexts and what characteristics of these contexts facilitate this. Such characteristics are also operative in mental health services, in particular because doctors are invested with the same degree of epistemic authority in all medical contexts. There are, however, additional features specific to mental health services which follow from the statement of a psychiatric diagnosis which can also contribute to experiences of epistemic injustice. I discuss these features in 6.3.2, with particular reference to the manner in which the presence of psychotic symptoms in the patient can generate a prejudicial credibility deficit in the listener leading to testimonial injustice. I also discuss how the process of diagnostic assessment can overlook the patient's personal narrative and effectively marginalise her from this process. Instead, it can encourage her to adopt a biomedical narrative about her condition.

In section 6.4, I discuss the importance of intellectual self-trust in supporting one's sense of agency and self-efficacy, which I base on Karen Jones' (2012) explication of this concept. Jones describes how being a victim of epistemic injustice can damage one's intellectual self-trust. I note the high level of trust that we normally give to medical practitioners by virtue of their epistemic authority. I argue that, precisely because of the trust we give to the medical profession and the epistemic authority given to doctors, one consequence of the diagnosis can be to diminish the recipient's intellectual self-trust, since the patient will likely have no good reason to mistrust the opinion of the clinician. Instead, the patient will be likely to downgrade any trust she has in her own views about her condition, where these appear to conflict with the judgment of the clinician.

The sorts of reactions that service-users experience after receiving a psychiatric diagnosis demonstrate in many cases feelings that their diagnosis has in some way undermined their sense of themselves and led them to question their previous self-narratives. In section 6.5, I argue that this can constitute hermeneutical injustice in that the effect of the diagnosis is to diminish the patient's confidence in their own hermeneutical resources and their ability to make sense of the difficulties in their lives. I discuss the results of several surveys of mental health service-users regarding their reactions to their diagnoses to illustrate this. I also cite some examples of testimonies by former service-users who have publicised their own experiences which support this view. On the basis of such surveys and reports, I claim that the receipt of a psychiatric diagnosis can render the recipient vulnerable to becoming a victim of hermeneutical injustice. While I mostly discuss psychiatric diagnoses in

general for ease of exposition, it should be noted that the DSM diagnostic categories are a very heterogeneous group. I am not claiming that all diagnoses are equally likely to lead to hermeneutical injustice for their recipients, nor that this is invariably a consequence for any diagnosis in particular. Rather, some diagnoses seem to be more likely to have this effect than others.

In section, 6.6, I discuss an objection that giving a psychiatric diagnosis to a patient might actually increase their hermeneutical resources. On similar lines, it might also be argued that denying a patient a diagnosis might constitute hermeneutical injustice. I argue that such objections can be met. Patients who want a diagnosis can still be given one, along with a full formulation of their difficulties agreed collaboratively with them. I also argue that much depends on the manner in which the diagnosis is communicated to the patient. Where the patient is told that the diagnosis is explanatory or that her condition is a chronic one, then she may be a victim of hermeneutical injustice. This can be avoided when the limits on what the diagnosis actually means are carefully explained to the patient, and this can often be done when a formulation of her difficulties is constructed collaboratively with her.

I conclude in section 6.7 by briefly reiterating the manner in which the receipt of a psychiatric diagnosis might lead to the recipient becoming a victim of hermeneutical injustice, particularly in the circumstances of over-stretched and under-resourced mental health services.

#### **6.2** Epistemic injustice

Miranda Fricker (2007) introduced the concept of epistemic injustice to describe one aspect of the ethical dimension of the epistemic activities in which we are habitually involved – i.e. the activities of reasoning, believing and knowing, giving testimony, and interpreting our experience. Inasmuch as we undertake such activities, we are epistemic agents. Epistemic injustice, therefore, is an injustice done to someone in their capacity as an epistemic agent. It is generated by some kind of negative identity prejudice towards the victim, and in some cases the victim may have internalised the negative identity, whether consciously or unconsciously. Fricker emphasises that it occurs in contexts of a power imbalance, in which the victim is situated at a power disadvantage in some manner.

Fricker identifies two forms of epistemic injustice: *testimonial* and *hermeneutical*.

Loosely speaking, testimonial injustice occurs when the credibility given to an individual is deflated due to negative identity prejudice and their credibility in giving testimony is thereby

undermined. Again speaking loosely, hermeneutical injustice refers to the marginalization of an individual's social experience due to structural identity prejudice of some kind (I discuss this in more detail in the following sub-section). It arises where the individual's psychological resources for understanding or interpreting his experiences are impaired or missing in some respect and his ability to have these experiences recognized in a wider social milieu is unjustly obstructed, due to the influence of dominant social groups. The individual is harmed when his experiences are marginalized in this way.

Anastasia Scrutton (2017) and Rena Kurs and Alexander Grinshpoon (2018) have claimed that having a mental illness can leave the person concerned vulnerable to hermeneutical injustice. In this chapter, I expand on these claims and argue in addition that the receipt of diagnosis makes a distinctive contribution to the hermeneutical injustice experienced by the individual, due to the biomedical narrative implications which the diagnosis conveys. However, it is not immediately evident from Fricker's account of hermeneutical injustice that this can be applied to the situations in which individuals are given a diagnosis that purports to explain their experiences where they may previously have had difficulty doing so. In answer to this, it can be argued that she has not fully encompassed the various kinds of situation that can lead to hermeneutical injustice as she seems to understand it. I therefore discuss some alternative ways in which this can be manifested in the following sub-section.

#### 6.2.1 Hermeneutical injustice

Fricker introduces the notion of hermeneutical injustice by considering what she calls "The Central Case". This is based particularly on two examples highlighted in the women's liberation movement in the USA in the 1960s and 1970s. The first concerned Wendy Sandford who suffered severe depression following the birth of her son. Feeling blamed for her inability to cope, both from her husband and from herself, she had no way of conceptualising her experiences. Only when she shared her feelings in a women's workshop did the concept of postpartum depression become known to her. This was a relatively new concept amongst mental health professionals at the time. She then found that she had a new way to make sense of her experiences and to communicate these to other people.

The second example concerned Carmita Wood who was employed in an administrative role at an American university. However, she was forced to leave her job because of the continual sexual harassment from her boss. She was unable to clearly make sense of her experiences, nor to take any remedial action through her employer, because the concept of

sexual harassment had not at that time been formulated and entered into a common conceptual vocabulary. The concept only emerged subsequently out of activities of the women's liberation movement in the USA.

In both these cases, Fricker argues, there was a lacuna in the collective hermeneutical resources available at that time which would enable people to comprehend the issues at stake in a coherent manner. Because of this lacuna, as Fricker describes it, Wendy Sandford was hermeneutically disadvantaged in not being able to name her difficulties as postpartum depression. In a similar manner, Carmita Wood was hermeneutically disadvantaged, as were the many other victims of sexual harassment before this concept emerged in general discourse. As a consequence both were victims of hermeneutical injustice. The injustice in each case was generated by a structural inequality of power between men and women and by what Fricker calls "a background inequality of hermeneutical opportunity" (2017, p.53). Without such a structural inequality of power, the situation might simply have been a case of epistemic bad luck. An important feature of the situation in which women like Carmita Wood found themselves in was that of being a disadvantaged group in a context of unequal power relationships, as is generally the case in employer-employee relationships and particularly salient in cases of female employees with male managers. Both Sandford and Wood were hermeneutically marginalized due to the background inequality of hermeneutical opportunity that prevailed in their cases and the unequal power relationships that prevented them from rectifying this marginalization. There was a consequent lacuna in their hermeneutical resources, such that they were unable to communicate their experiences to the wider community. They lacked the conceptual lexicon that would have allowed them to do that.

Such cases are examples of what Fricker calls *systematic* hermeneutical injustice. What distinguishes them from mere epistemic bad luck is the systematic way in which they can occur. More specifically, she argues that hermeneutical injustice arises from structural prejudice in the hermeneutical resources available to the community and that people may become victims to it by virtue of some aspect of their social identity. She offers a definition of this as follows: "the injustice of having some significant area of one's social experience obscured from collective understanding owing to a structural identity prejudice in the collective hermeneutical resource" (2007, p155). This, for her, defines hermeneutical injustice in those cases where the subject suffers from prejudice in relation to their membership of a relatively powerless group. The hermeneutical marginalization that the subjects suffer is fundamental to their being victims of hermeneutical injustice. Such

examples are characterised by a conceptual absence within the community concerned, in that a relevant concept is absent from the collective hermeneutical resources of that community.

However, as Rae Langton (2010) has pointed out, hermeneutical resources can contain harmful presences, as well as damaging absences, and these can also lead to hermeneutical injustice. Langton does not give any examples of such presences, but one can suggest how it might arise in a situation such as might occur with an unemployed person claiming welfare benefits. A common label for such people in parts of the media and some political circles is that of 'benefit scrounger' which is intended to suggest that they are unwilling to work and happy to live off benefits, despite evidence that this is seldom the case (Mulheirn, 2013). Most people who claim welfare benefits will be aware of this slur, as well as the stigma that is frequently associated with needing benefits. This might affect individuals in various ways. They may internalise the stigma, such that they believe themselves to be deficient in some way for having to rely on benefits. They may also face hostility from staff at the benefits offices and other negative consequences, such as sanctions when they arrive late for an appointment because their scheduled bus failed to arrive on time. In these sorts of ways, the label 'benefit scrounger' is a harmful conceptual presence in the collective hermeneutical resources that they unavoidably encounter. It can diminish their beliefs in their own hermeneutical resources and hinder their attempts to communicate their narratives to the dominant community in which they find themselves. I shall argue below that a similar sort of case can arise when individuals receive a psychiatric diagnosis.

In addition, José Medina (2013, 2017) argues that we should not be restricted to seeing people's hermeneutical capacities limited by the kinds of concepts and terms dominant in the broader culture. His argument here is similar to that of Rebecca Mason (2011) who distinguishes between *dominant* and *non-dominant* hermeneutical resources. According to her distinction, non-dominant groups may still have sufficient hermeneutical resources to understand their experiences, but these will be marginalised by the dominant hermeneutical resources possessed by more powerful groups. She notes that certain "relations of power allow some to neglect or claim interpretive authority over the experiences of others" (p.295). As an example, she cites Charles Mills (2007) who explains that black people, during the period of slavery, were severely oppressed, but still had ways of expressing their suffering and in some cases were able to speak out about it as well, despite the hermeneutical marginalization from which they suffered. Medina observes that there are a range of communicative processes in which people might try to make sense of their experiences, both to themselves and to others, and that these are more heterogenous than Fricker's account

seems to suggest. It is not the case that only those concepts that have a place in the broader hermeneutical resources can allow people to make sense of their individual experiences. Thus he says: "it is dangerous to establish too close a link between intelligibility and linguistic labels" (2013, p. 98). In particular, he notes that oppressed people often have their own means to render their experiences intelligible to themselves, if not to others in the dominant culture, as Mills' example of slavery demonstrates.

Medina also observes that there are cases where hermeneutical injustice arises as a result of institutional dynamics in contexts "when there are structural conditions or institutional designs that prevent the use of certain hermeneutical resources or expressive styles, or simply when those conditions or designs favor certain hermeneutical communities and practices and disadvantage others" (2017, p46). These can plausibly include medical communities in general, and mental health service communities in particular, which are characterised by very well established conceptual lexicons and expressive practices. In these contexts, such hermeneutical practices can become strongly dominant and marginalise other less well established ones.

Similarly, Komarine Romdenh-Romluc (2017) notes that the picture of hermeneutical injustice painted by Fricker is one that involves a lack of concepts. She argues, however, that the issues that matter here are "competing views of the world". What this means is that it is not enough for a potential victim of hermeneutical injustice to have the relevant concept in her conceptual lexicon, but also that the wider culture should be willing to use it to describe the experience at issue. Thus, in the case of Wendy Sandford cited by Fricker, it is not enough for her to have the concept of postpartum depression. It is also necessary for the wider culture to accept that this is a valid description of her experiences. If, for some reason, the culture is persuaded by a group of doctors that postpartum depression is not a real phenomenon, then the harm she suffers will be the same as if she did not have access to that concept. She will be blamed for her perceived inadequacies in the same way as before the concept entered the conceptual lexicon. Romdenh-Romluc describes this as "the problem of authority", in which the meanings imposed on people are those determined by the dominant cultural group. In the context of mental health services, the dominant group is the psychiatric and associated professional community, by virtue of their epistemic authority and their power to prescribe treatment for the patient.

Insofar as Fricker's account of hermeneutical injustice appears to focus mainly on the lack of concepts with which hermeneutically marginalised individuals might name their experiences, this seems to miss out the cases where individuals are marginalised because their

own hermeneutical resources are non-dominant in relation to the larger community of which they are a part. Nevertheless, the common theme in both Fricker's examples and those emphasised by Mason and Medina is one of hermeneutical marginalization. The effect of this is that the subject's social experience is obscured or omitted from collective understanding, consistent with Fricker's definition of hermeneutical injustice. The fact that subjects are hermeneutically marginalised, even when they are able to understand their own experiences, means that their understanding of their experiences are prevented from entering collective understanding where this is governed by dominant groups in society. Moreover, this does not necessarily prevent those who are marginalised from losing confidence in their own hermeneutical resources and becoming victims of hermeneutical injustice in the manner that Fricker describes.

# 6.3 Epistemic injustice in medicine and psychiatry

# 6.3.1 Epistemic injustice in healthcare contexts

Havi Carel and Ian James Kidd (2014; Kidd & Carel, 2017, 2018) have described how patients in general medical settings can become victims of epistemic injustice. They argue that both testimonial injustice and hermeneutical injustice can be experienced. Such instances are common and result from the structures and practices which typify modern healthcare systems. They label such injustices as pathocentric epistemic injustices. It is central to Fricker's account that epistemic injustice is liable to occur in situations where there is an imbalance of power. Ill people are vulnerable to this, since patients are inevitably placed at a disadvantage by virtue of their illness and their need for access to specialist knowledge and resources possessed by the clinicians. This puts clinicians in a position of "epistemic privilege" compared with their patients, such that the clinician can claim epistemic authority over the patient. Moreover, the experience of illness, particularly severe illness, is likely to be a source of anxiety and distress to the patient, which can impair her ability to describe her complaints in a clear manner. This can give the impression to the clinician that the patient is unable to report their symptoms reliably and hence predispose the patient to suffering testimonial injustice when their reports are not given due credibility. The diagnosis itself can also be a trigger for significant emotional distress when it indicates a serious condition for the patient.

Kidd and Carel (2018) suggest that a fundamental source from which pathocentric epistemic injustice can be generated is the naturalistic conception on which modern medicine

is based. This is broadly the view (discussed in Chapter 3) that diseases can be understood as disorders of biological functioning in relation to a relevant reference class. Many philosophers of medicine favour hybrid accounts of disease which incorporate normative elements in addition to naturalist conceptions, in preference to purely naturalist accounts. Nevertheless, the naturalism underpinning modern medical science remains a dominant mode of thinking. This conception assumes a broadly reductionist account according to which diseases are a product of abnormalities in complex biochemical or physiological processes and which therefore have to be addressed by highly trained practitioners with expertise in biomedical science. It is from this background of training that doctors derive their epistemic privilege and authority, and which can lead them to marginalise or exclude the perspectives of their patients.

The same set of circumstances also leave patients liable to become victims of hermeneutical injustice (Kidd & Carel, 2017; 2018). This can arise because people with serious illnesses may have difficulty in articulating or adequately communicating certain aspects of their experience. This can occur because the hermeneutical resources they have available for understanding their condition are not such as are likely to be recognised as an important means of understanding by the dominant professional healthcare community. Kidd and Carel refer to Mason's (2011) distinction between dominant and non-dominant hermeneutical resources. On Mason's analysis, hermeneutical injustice can arise when the non-dominant hermeneutical resources of the less powerful group are not recognised or acknowledged by the dominant epistemic authorities. In the healthcare context, the medical profession is a powerful group that possesses dominant hermeneutical resources. Kidd and Carel, therefore, argue that patients in healthcare settings can suffer from hermeneutical injustice in the way Mason describes. Most ill people can describe their experiences clearly, although they typically do so in non-expert terms. However, these may often be felt to be inappropriate for public discussion and play little role in clinical decision-making.<sup>61</sup> In such circumstances, their stories will be relatively marginalised and they may become victims of hermeneutical injustice. It can also be the case that some aspects of the ill person's experience may be incapable of expression in propositional form – e.g. in cases of extreme or

\_

<sup>&</sup>lt;sup>61</sup> One might question the assertion that patients' experiences play little role in clinical decision-making by pointing out that when patients complain of pain, for example, they will typically be given analgesics. However, the point here is that the major clinical decisions – e.g. whether to perform surgery or prescribe radiotherapy for cancer – are based on the findings of a range of biomedical tests or imaging results which the patient is unlikely to be able to assess.

chronic pain where the experience is literally indescribable. In such cases, they may also suffer hermeneutical injustice if they are unable to convey their experiences in any coherent manner.

Kidd and Carel (2021) also discuss the 'epistemic predicament' that patients can find themselves in. This encompasses the various epistemic challenges that may often confront patients in the healthcare system. These can be complex and ongoing for a period of time, rather than being confined to single instances of epistemic injustice in an otherwise straightforward episode of healthcare. There can be continual attempts, however unintended, by healthcare staff to undermine the testimonial credibility of the patient. The resulting confusion in the patient generated by the experience of not having her testimonies taken seriously can also lead to her questioning her own understanding of her experiences. Her own hermeneutical resources will come to feel marginalised in consequence, such that she may fall victim to hermeneutical injustice in her diminishing ability to make sense of her predicament.

The hermeneutical marginalization that occurs in such cases can be understood as a form of prejudice towards the ill person, in the terms described by Christopher Hookway (2010). He observes that the prejudicial credibility deficit, which Fricker describes as underpinning testimonial injustice, can begin before any epistemic interaction with others takes place. He describes two perspectives through which this can happen. The first of these is the *informational* perspective according to which a judgment is made as to whether the person can provide reliable and trustworthy information. When circumstances are such as to make it less likely that the person will be considered capable of doing this, he will be a potential victim of epistemic injustice. The second is the *participant* perspective in which a judgment is made about whether the person is sufficiently competent to participate in the epistemic activity in question. To the extent that a negative judgment is made about the person's reliability or ability from one or both perspectives, the person will in effect suffer from epistemic silencing. This can be the case with people being treated for illnesses. Kidd and Carel argue that people who are ill are vulnerable in both these ways. In particular, they are liable to suffer prejudice from the participant perspective in that "they are typically regarded as the objects of the epistemic practices of medicine rather than as participants in them" (2016, p. 10). The information sought from them will normally be confined to details about their biographies and their presenting symptoms. This in effect is what Carel and Kidd (2014) describe as the third-person view taken by medicine, where the goal is to focus

directly on the patient's bodily characteristics and to repair or ameliorate physiological processes, rather than giving much space to the first-person perspectives of the individual.<sup>62</sup>

## *6.3.2 Epistemic injustice in psychiatry*

Psychiatry is a branch of medicine. Hence, the factors that contribute to epistemic injustice in somatic medicine can be equally operative in the context of psychiatry. Both forms of epistemic injustice can arise and both can be attributed either to the effects of diagnosis or to the general perception of the patient as being emotionally or cognitively disturbed. Thus, Abdi Sanati and Michalis Kyratsous (2015) explain how patients identified with delusional states, one of the symptoms of schizophrenia, can be subject to testimonial injustice. They cite two case examples of patients whose testimony on matters they knew about was discounted because of negative identity stereotypes formed on the basis of their diagnoses. In one case (Ms. J.N.), the patient reported that her husband was being unfaithful to her, but this was disbelieved, despite subsequently being shown to be true, because she had previously been displaying evidence of persecutory delusions. In the second case (Mr. M.G.), a young man with a diagnosis of schizoaffective disorder was picked up by the police and brought to the hospital following his threat to attack somebody. He was found to be suffering from various delusions. He claimed that the person he threatened to attack had abused a relative of his. His claim was disbelieved, but further investigation showed it to be true.

In each of these cases, the patients' reports turned out to be truthful, but were disbelieved because of the stereotypes associated with a diagnosis of mental disorder, and as such both were victims of testimonial injustice. Fricker (2017), commenting on these examples, observes that there is a prejudicial stereotype of delusional people operative in such cases which can lead to an over-generalisation. An assumption is made that the delusional state of such patients affects all their cognitive processes, rather than just specific instances. This prejudicial over-generalisation can lead to secondary disadvantages, such as compulsory detention in hospital and stigma. Fricker notes that "the effect of the prejudicial over-generalisation fits exactly the theoretical structure of testimonial injustice: the intrinsic injustice of being judged as epistemically lesser owing to prejudice, plus a secondary associated disadvantage" (2017, p.58).

-

<sup>&</sup>lt;sup>62</sup> This, of course, is not to deny the importance of the third-person view in medicine, which is essential if the patient is to receive optimal care, particularly in urgent and emergency situations when there may be insufficient time to take full account of the patient's individual perspective.

Another example of testimonial injustice is cited by Crichton, Carel and Kidd (2017). They report:

When one of the authors... was a medical student in Munich, Germany, he saw a young man on an acute psychiatric ward who said he was a relative of the then Soviet leader. The responsible consultant took this to be a grandiose delusion, and therefore as evidence of a psychotic delusion; it later turned out to be true (p.66).

Again, the manner in which the patient's report was dismissed on account of his diagnosis of mental disorder shows this to be a case of testimonial injustice. What seemed like an unlikely claim was disregarded, because it was assumed that his condition rendered him liable to entertain such delusions. As such, he suffered from a prejudicial credibility deficit in the eyes of those supervising his care.

The same set of factors can also contribute to patients becoming victims of hermeneutical injustice. As Fricker (2016) notes, the two kinds of epistemic injustice are related and stem from the same contextual circumstances. Testimonial injustice can create hermeneutical marginalisation by virtue of the barriers placed on acknowledging the reports and comments of the speaker. As such, the shared hermeneutical resources of the community will be diminished in a structurally prejudicial manner. The speaker's experience of being unable to have their statements acknowledged and recognised as valuable contributions to the dominant discourse is what constitutes hermeneutical marginalisation, and this gives rise to hermeneutical injustice. Where the circumstances are such that the marginalisation is a regular occurrence, the injustice becomes systematic.

Kidd and Carel (2021) observe that the epistemic predicament of patients can be particularly acute and entrenched in psychiatric settings. They note in particular that "one aspect of the epistemic predicament of ill persons is getting others to grasp the complexity and the particularity of their experiences" (2021, p.72). This is not an easy predicament for patients to escape from, in view of the length of time that patients may be under the care of psychiatric services, not just when they are in hospital, but also under the care of community mental health services as well. Among other issues, there will be pressure on patients to cooperate with treatment regimes which very often involve taking psychiatric drugs over a prolonged period. These can have significant adverse effects on the patient (Moncrieff, 2020). The patient's reports about these effects and their reluctance to continue taking the medication may be easily dismissed by the psychiatrist or other clinician, with the patient left feeling that she is unable to have the distinctive nature of her experiences acknowledged by

the professionals involved. In this way, the patient may be a victim of both testimonial and hermeneutical injustice.

It is plausible that the occurrence of both forms of injustice can be systematic in mental health contexts. As discussed in Chapter 5, the typical style of diagnostic interviewing can marginalise the patient's own narrative by not allowing sufficient space for the patient to elaborate it in depth and detail. This is a consequence of the process of interviewing a patient to obtain a DSM-based diagnosis, which focuses on the symptoms that are likely to fit specific diagnostic categories, rather than the idiosyncratic meanings that may be important to the patient (Stanghellini, 2004). To the extent that the patient is not given the space to communicate such meanings, any information she might have to offer will either not be sought or not be considered valuable enough to influence the hermeneutical resources employed in the dominant discourse. When this happens, the patient will be subject to epistemic silencing from the participant perspective in the manner described by Hookway (2010). In such cases, the speaker is judged by the hearer to be incapable, for some reason, of offering useful observations due to prejudice on the hearer's part, and this effect can occur before the speaker has offered any reports. Therefore, this kind of silencing can take place before any distinct occurrence of testimonial injustice. 63 In the context of a diagnostic assessment, the information sought by the clinician, particularly in an under-resourced mental health context, will be largely confined to those items that are relevant to reaching a diagnosis, with other features of the patient's narrative about her experience either being disregarded or not sought at all. The priority placed on deciding on a diagnosis in most cases is what can lead to this kind of silencing being systematic.

Scrutton (2017) observes how, in mental health services, the ability to interpret the patient's experience correctly is assumed to lie with the qualified clinician. Medicine, and by extension psychiatry, is the recognised authority in modern society and, as such, the medical perspective is assumed to be superior to other perspectives in this context. Scrutton suggests that the diagnosis "effectively constitutes a monopoly on how the experience is interpreted" (p.349). While it seems too strong to describe the diagnosis as constituting a monopoly in all clinician-patient interactions, it typically has a dominating effect on the way that the patient's experience is understood and places considerable pressure on the patient to understand her experience in that way, rather than in terms of her own self-narrative. Instead, the diagnosis

-

<sup>&</sup>lt;sup>63</sup> Fricker (2016) argues that the kind of epistemic silencing described by Hookway represents a special kind of testimonial injustice, in that the consequences for the speaker (or potential speaker) are broadly equivalent.

conveys the implication that the patient's condition is better understood within a biomedical narrative. Insofar as it does this, the patient is not enabled to come to an understanding of her condition in any terms other than biomedical ones and, as such, is liable to be a victim of hermeneutical injustice.

Moreover, inasmuch as the qualified clinician determines how the patient's experience is to be interpreted and disregards the patient's own perspective, he may be displaying what Gaile Pohlhaus (2012) describes as 'willful hermeneutical ignorance'. As she describes this, the clinician is using epistemic resources that do not fully allow for the patient's testimony to be regarded as intelligible. The clinician may do this wilfully, in the sense of deliberately deciding to disregard certain elements in the patient's story, or alternatively his psychiatric training may have conditioned him not to think that any kind of epistemic resources, other than those deriving from the biomedical model, are deserving of consideration. Either way, he will be placed in a state of hermeneutical ignorance regarding the patient's predicament.

In summary, people can become victims of both testimonial and hermeneutical injustice in healthcare settings due to the negative stereotypes of ill people that can arise. These cases may be particularly likely to occur when patients have difficulty in articulating the nature of their symptoms, due to the hermeneutical marginalisation they experience. For similar reasons, both kinds of epistemic injustice may also occur in mental health contexts.

#### 6.4 Effects of a diagnosis on intellectual self-trust

One consequence of the epistemic silencing that can follow for people being given psychiatric diagnoses is that their intellectual self-trust can be damaged when, for example, patients are not judged to be capable of participating adequately in the ongoing epistemic activity about their condition. Karen Jones (2012) argues that intellectual self-trust "is a stance that an agent takes towards her own cognitive methods and mechanisms, comprising both cognitive and affective elements" (p.237). Trust in other people entails, among other things, the stance that what they tell us is reliable. Moreover, as Katherine Hawley (2019) explains, it involves something more than mere reliability. We may regard inanimate objects, such as the kettle in which we boil water, as reliable, but when they malfunction for whatever reason we don't think any trust has been broken. Therefore, there is an important extra dimension involved in trusting a person rather than an object. Hawley argues that this extra dimension must involve an element of commitment on the part of the trustee such that he will be trustworthy in some specified sense. This means that there is an expectation that he will

behave in such a way as to justify the trust placed in him. As such, there is a normative dimension involved in trust which is absent from mere reliability. This can be seen particularly in the normal situation where patients generally trust their doctor. It is well-known that doctors have a set of ethical standards with which they are expected to comply. These include the requirements to exercise their medical skills to the best of their ability for the benefit of the patient and not to exploit the patient for personal gain or gratification. Understanding what is expected of doctors encourages patients to put their trust in them.

Jones argues that self-trust can legitimately be thought of as trust. When applied to oneself, it is a broad feeling, typically unconscious though it may become conscious where unusual and difficult circumstances arise, about the extent to which one can trust one's own judgments about things. However, for self-trust, the normative dimension of trust does not seem so salient, since nobody else is directly affected if I lack trust in myself. Nevertheless, in optimal circumstances, I still have an expectation that I can trust myself to carry out certain intellectual and practical tasks, those which I have reason to believe I should be capable of performing. This, of course, is another way of describing the sense of self-efficacy one may feel. Intellectual self-trust, therefore, is the feeling that one can trust one's own abilities to make judgments about one's present and future needs.

Jones notes that self-trust develops in social interactions and is maintained therein. It is consequently influenced by social power. She says: "Social relations of dominance and subordination affect our intellectual self-trust because they affect both the way others respond to us as inquirers and shape our own understandings of our cognitive abilities" (2012, p.245). She argues that intellectual self-trust can be damaged when one is a victim of epistemic injustice and that this damage can be cumulative when there are repeated instances of this. A key element generating testimonial injustice is the credibility deficit accorded to negatively stereotyped people and groups. Where an individual is not accorded credibility because of a stereotype, this is likely to impact on her self-trust.

Hermeneutical injustice can also damage intellectual self-trust. This was the case, described by Fricker (2007), for the women who lacked the concept of sexual harassment and thus could not make adequate sense of their experiences when this happened to them. In such situations, they could be inclined to believe that this was a difficulty in their own understanding of the kind of behaviours that might normally be expected in the workplace. This kind of case is one where conceptual gaps are responsible for leading to hermeneutical injustice. However, the latter can also be produced when dominant groups effectively control the collective hermeneutical resources, such that the victim's own hermeneutical resources

become suppressed, thereby damaging her trust in her own ability to make sense of her circumstances. Jones also observes that in extreme cases "hermeneutical injustice is so corrosive of self-trust that its victim can come to doubt their sanity" (ibid, p246). In such cases, the injustice is likely to have powerful and long-lasting effects.

There may often be conflicts arising between one's own judgment and that of others in one's social environment. Alternatively, one may willingly defer to the judgments of other people whom one generally has good reasons to trust. Thus there are circumstances where we may be strongly inclined to trust the judgments of other agents, when we believe that their expertise in a given area is far superior to our own. A prime example of such a situation is when we are unwell and need to seek a medical opinion about our condition and effective treatment for it. In such circumstances, we place a lot of trust in the doctor's opinion, recognising their epistemic authority, and we do not normally see any reason to trust our own judgment over that of the doctor.<sup>64</sup> For a similar reason, therefore, there will be a strong inclination on the part of the patient receiving care from mental health services to trust the judgment of the psychiatrist about the nature of their condition. The patient is unlikely to find any reason to distrust the diagnostic verdict, at least initially, whatever reactions they may experience subsequently. A diagnosis, by its very nature as a perlocutionary speech act delivered by a dominant and trusted epistemic authority<sup>65</sup>, is likely to have a powerful and long-lasting effect on the patient. Inasmuch as it does so, it may reduce the patient's intellectual self-trust by emphasising that her distress is caused primarily by a biomedical condition, such that her own attempts to understand her experiences may be seen by her as likely to be less fruitful. For the same reason, she may become a victim of hermeneutical injustice. She may be led by the epistemic authority of the clinician to conceptualise her condition as being a function of a mysterious process in her brain, rather than being helped to formulate her predicament in terms of her own feelings and experiences which can make more sense to herself. Her own hermeneutical resources may become correspondingly marginalised if she becomes less inclined to trust her own judgments about herself.

-

<sup>&</sup>lt;sup>64</sup> Interestingly, it is sometimes joked that doctors make the worst patients. Presumably this is because, when they find themselves in the role of patient, they may be too tempted to trust their own judgment, rather than that of the clinician treating them.

<sup>&</sup>lt;sup>65</sup> As discussed in Chapter 4.

# 6.5 Evidence of hermeneutical injustice in testimonies of service-users and former service-users

I claim here that the receipt of a psychiatric diagnosis can render the recipient liable to becoming a victim of hermeneutical injustice. What characterises this is being in the position in which a significant portion of one's personal experience is marginalised or is rendered unintelligible to dominant groups in the community, due to a structural identity prejudice. This may occur, as in the case of Carmita Woods cited by Fricker, when one lacks the necessary conceptual resources to understand one's experiences. Alternatively, it can occur when members of oppressed groups have hermeneutical resources of their own with which to understand their experiences, but these are marginalised by members of dominant groups. Hermeneutical injustice can occur in many ways and circumstances. However, as Fricker (2017) states, what in general causes it is "a background inequality of hermeneutical opportunity – specifically, hermeneutical marginalisation in relation to some area of social experience" (p.53). The wrong that this causes the victim is the unfair disadvantage they find in understanding their experience or in getting others to understand it. Users of mental health services are particularly at risk of becoming victims of this for two reasons: first, because they are inevitably in a state of some mental distress, and second, because of the unequal power relationship with the clinicians assessing and treating them. For both reasons, they will be under a great deal of pressure to accept the judgment that is given to them about their condition.

The manner in which psychiatric diagnoses are given, and in particular the frequent implication that they represent some kind of biomedical condition, can have the effect of marginalising the patient's own meanings concerning his condition. To the extent that the patient is misled into believing that the diagnosis represents a biomedical abnormality which explains his condition, and that consequently his own narrative about himself is of little or no relevance to explaining it, then he becomes a victim of hermeneutical injustice. Since DSM-based diagnoses do not explain why the patient developed his condition, he will be liable to suffer harm if he mistakenly believes they do. His feelings of agency and hopes for recovery are likely to be diminished if he believes he has no control over his mental states, as a consequence of abnormal biochemical changes in his brain which are presumed to be responsible for his current condition. This is particularly the case if he is also led to believe

his condition is a chronic one, when there is little evidence to show that this is necessarily the case.<sup>66</sup>

Kurs and Grinshpoon (2018) note that psychiatrists can be expected to use a psychiatric vocabulary to describe the experiences of patients. To the patient himself, however, this is liable to feel alienating and disempowering, such that the patient may be silenced by it. Kurs and Grinshpoon note: "The patient might seem to lose his or her ability to speak confidently, except when his or her language conforms to the standard medical discourse" (2018, p.340). Inasmuch as this is the case, it would seem to reflect a loss of intellectual self-trust by the patient, as described by Jones (2012). The concentration on a biomedical description of the patient's condition might also deflect attention from the social and environmental factors that led to his present condition, with the consequence that he may not be helped to address the salient issues in his life in a way that would be important for his recovery. In not being given an adequate opportunity to address such issues and to make sense of his experiences in non-medical terms, he may become a victim of hermeneutical injustice.

Some evidence of the impact of a diagnosis on patient's narratives which are suggestive of hermeneutical injustice can be found in the reports of patients and former service-users of mental health services. There have been a number of surveys of the experiences of patients and several former service-users have published reports in various forms of their experiences. The methodologies used in such surveys are very varied and it is difficult to summarise their findings succinctly, particularly because the nature of the findings in these surveys precludes any systematic quantitative analysis. Nevertheless, a systematic meta-review by Perkins and co-workers (2018) of the many published surveys into service-users' experiences demonstrated a range of reactions to receiving a diagnosis, including both negative and positive reactions. In what follows, I concentrate on those expressing more negative reactions (though in section 6.6 below I note some positive reactions also), since I claim that these are the ones most likely to become victims of hermeneutical injustice.

A frequent response from service-users was that the disclosure of a diagnosis was an important moment for them. In some cases, they welcomed the diagnosis, reporting that they felt their experiences were validated by it and that it gave them a greater self-understanding. This might appear to conflict with my claim that a psychiatric diagnosis renders the recipient vulnerable to being a victim of hermeneutical injustice. However, this seemed to depend in part on the particular diagnosis being given. Perkins and co-workers (2018) noted in their

-

<sup>&</sup>lt;sup>66</sup> As discussed in Chapter 2.

meta-review that diagnoses of psychosis (including schizophrenia) and personality disorder were particularly likely to be received negatively by the patient – these were most likely to have negative effects on sense of identity and hopes for recovery. In addition, many respondents reported that a diagnosis could be experienced as labelling which could have stigmatizing effects.

Given the qualitative nature of such surveys, it is not possible to quantify precisely how many service-users had negative experiences of their diagnosis and how such experiences can be summarised. These surveys do not report any statistical data. Nevertheless, it is clear from them that negative experiences of diagnosis, though not universal, are at any rate quite common. Thus, in one survey a respondent reports:

Schizophrenic is the worst diagnosis because I've heard it in the newspapers and on TV, that they are really mad schizophrenic people, they are very dangerous to society, they've got no control. So obviously I came under that category (Dinos, et al, 2004). In this particular case, the diagnosis has a meaning for the recipient derived from popular conceptions of what it means. However, it would appear that the psychiatrist who gave the diagnosis failed to make clear to the patient that such people are not "mad" or "dangerous".

A survey by Liz Pitt and co-workers (2009) of eight people who had received a diagnosis of psychosis, including schizophrenia, schizoaffective disorder, bipolar disorder, and personality disorder,<sup>67</sup> showed that service users could have a similar reaction to the negative labelling they perceived in the diagnosis. One of them states:

I just thought schizophrenic people could go around murdering and raping people... I didn't know nothing (sic) properly about schizophrenia at that time so that's my initial thought. I can remember actually being told... I wasn't well at the time. I went absolutely bananas, yeah, throwing the bloody furniture everywhere. They pinned me down, give me injection... because they were trying to tell me I got schizophrenia and... I'm not schizophrenic (p.421).

Pitt and her co-workers note that there was a mixture of positive and negative reactions to their diagnosis among the respondents. Given the small numbers of respondents, they noted the difficulty of generalising from these results. However, for those voicing negative reactions, the diagnosis was experienced as disempowering. It was seen as a "prognosis of

\_

<sup>&</sup>lt;sup>67</sup> Personality disorder is not normally thought of as a form of psychosis. The authors of this survey aimed it largely at people with a diagnosis of some form of psychosis, but some people with a personality disorder were also included. They note that several of the participants had multiple diagnoses.

doom" (p.421), which in some cases was associated with the lack of information accompanying the diagnosis. The authors also commented that the predominance of the biomedical model and the heavy reliance on medication as the treatment of choice could have contributed to the sense of disempowerment experienced by respondents.

In another survey by Bonnington and Rose (2014) of 46 service users, one of them talking about their diagnosis of borderline personality disorder states:

It [being diagnosed] was such a shock... It really was an insult actually. [The psychiatrist] invested no time in me whatsoever, and it was just like I was a naughty dirty person... it was like I should be ashamed of myself... it's made me very insecure about my worth as a person, who I am, because I used to be so capable and now I'm a nothing, a nobody. It's taken everything away from me (p11).

Another respondent in the same survey talking about borderline personality disorder states:

I can't seem to get anywhere with that diagnosis... [the diagnosis] feels like a bit of trapping... whatever I ask for or need doesn't seem to be dealt with (ibid, p13).

Bonnington and Rose comment that when the diagnosis was conveyed to the patient, it was rarely explained. Several respondents stated that the particular diagnosis of borderline personality disorder seemed also to exclude them from any helpful treatment.

In a study using a questionnaire and a semi-structured interview by McCormack and Thomson (2017) of five people who had developed severe psychological difficulties after experiencing trauma in childhood, one respondent talks about their response to a diagnosis (the actual diagnosis is not stated in this case):

I do not see myself as someone with a mental health [diagnosis]... It's because of a traumatic childhood that was out of my control. I was not born that way (p.160).

The diagnosis was seen as an unwanted label which was felt to be obscuring that person's natural and ongoing psychological reaction to traumatic early experiences.

An on-line survey by Thomas et al (2013) of service-users and other members of the public concentrated specifically on respondents' views about the diagnosis of schizophrenia. A specific question in the survey about individuals' experiences following receipt of that or a more general diagnosis of psychosis received 97 responses.<sup>68</sup> The experiences reported were very varied, but the majority were negative in some way. There were several comments (25%)

<sup>&</sup>lt;sup>68</sup> The question asked was: "If you have been diagnosed with 'schizophrenia' or 'psychosis' yourself', please tell us what happened to you that resulted in you getting the diagnosis".

of the sample) about the harm that recipients felt the diagnosis inflicted on them, with the perceived stigma being a prominent issue. One anonymous respondent reported as follows:

The humiliation of being labelled schizophrenic threatened to become a self-fulfilling prophecy. In the hospital, shelters, group homes and programs I was put in, I was socialised into being a mental patient. I was encouraged to see myself as a broken invalid, to forget my strengths, and instead focus on my weaknesses and vulnerability (p.136).

Other respondents perceived the diagnosis as implying a life sentence from which they would not recover. One respondent reported that any attempt to challenge the diagnosis would be interpreted by the psychiatrist as a symptom of the illness. Another described how all aspects of his life, including his reading interests and political beliefs, would be seen as symptoms of the condition. Several (14%) were concerned about the process followed in making the diagnosis. For example, one felt that the psychiatrist's belief in the genetic basis of schizophrenia led to the importance of that patient's difficult life experiences failing to be acknowledged. Thomas et al state that the lives and beliefs of these patients "were devalued by stigmatising medical assumptions" (p137). In contrast, only a small number of respondents (6%) reported finding the outcome of the diagnosis helpful.

Some former service-users have published detailed accounts of their own experiences of mental health services and the effects that the diagnosis had on them. For example, Jacqui Dillon, who now campaigns on behalf of mental health service-users in the UK, talks about her experience in the following terms:

The clear message I received.... was that I was ill. *Everything* that I said and did was caused by my illness. The abuse never happened – even thinking it did was part of my illness.... The fact that I didn't want to take medication was because I was ill. If I wanted to get better, I must accept my diagnosis and take medication... I would always have this illness. I wouldn't be able to work. I didn't know what was best for me. I lacked insight (2011, italics in original, pp.144-5).

Another former service-user, Elyn Saks, who is now a law professor in the USA, has written a book about her experiences of her psychological disturbance and her involvement with mental health services. She writes about her diagnosis of schizophrenia:

The Diagnosis. What did it mean? Schizophrenia is a brain disease which entails a profound loss of connection to reality. It is often accompanied by delusions... and hallucinations.... Often speech and reason can be disorganized to the point of incoherence. The prognosis: I would largely lose the capacity to take care of myself. I

wasn't expected to have a career, or even a job that might bring in a pay check. I wouldn't be able to form attachments, or keep friendships, or find someone to love me, or have a family of my own – in short I'd never have a *life* (2007, p.168, italics in original).

At this point in the passage, she talks about the uncertainty about whether available treatments would help her, the "terrible side-effects" of any medication, and the belief that there could be no cure. She goes on to describe her reaction to this.

I'd always been optimistic that when and if the mystery of me was solved, it could be fixed; now I was being told that whatever had gone wrong inside my head was permanent, and from all indications, unfixable. Repeatedly, I ran up against words like "debilitating", "baffling", "chronic", "catastrophic", "devastating" and "loss". For the rest of my life. *The rest of my life*. It felt more like a death sentence than a medical diagnosis (ibid, p.168, italics in original).

Another former mental health service user, Patricia Deegan, who is now a clinical psychologist and disability rights advocate in the USA, talks about some of her experiences as follows:

My psychiatrist told me I had chronic schizophrenia... He said I would be sick for the rest of my life and the best I could do was avoid stress and cope (from Deegan, 2004, <a href="https://www.patdeegan.com/pat-deegan/lectures/silence">www.patdeegan.com/pat-deegan/lectures/silence</a>; quoted in Phillips, 2013, p.16).

She goes on to say that she fought against what she saw as a "prognosis of doom" and subsequently went on to achieve qualifications in clinical psychology, thus rejecting the prediction of chronicity which she had been given.

What these writers indicate here is the pessimism and hopelessness that is often associated with a diagnosis of schizophrenia. In both Dillon's and Deegan's cases the diagnosis was explicitly communicated with the message that the condition would be chronic. In Saks' case, it isn't quite clear whether she was told this directly, but in any case this is evidently what she understood by it.

The quotes above, while expressing generally negative reactions, can be interpreted in various ways, and it may be difficult to infer directly from each one that the individual has necessarily been a victim of hermeneutical justice. This would depend on how in each individual case their overall self-narratives were affected by the diagnosis and how each person responds to this. However, I claim that such testimonies reveal a significant impact by the diagnosis on the self-narratives of most of the people concerned, and that as a

consequence they are likely to be victims of hermeneutical injustice.<sup>69</sup> An almost universal theme in service-user reports is that their diagnosis was seen as particularly important, whether they viewed it positively or negatively. A strong theme that seems to emerge from these reports is that of confusion about what the diagnosis meant for the individuals concerned and their sense of identity. In many of the cases where respondents express a negative view, the concern expressed is one of not knowing what the diagnosis meant for them and to what extent they would have control over their lives subsequently. Being able to control our own lives depends on having a clear sense of who we are and what constraints there may be on our shaping of our futures. In other words, our sense of what we can do to control our future lives depends in part on the self-narratives we have. As the quotes above indicate, the message often given to the respondents was that the diagnosis represents a chronic condition from which they could not hope to recover completely, although the empirical evidence does not generally support such a view. This message might not necessarily have been stated explicitly by the clinician (though in some cases it was, as Dillon's and Deegan's reports show), but rather be inferred from popular conceptions of what the diagnosis means, despite the fact that such conditions are frequently not chronic.

I claim therefore, following Scrutton (2017) and Kurs and Grinshpoon (2018), that the receipt of a DSM-based diagnosis can cause the recipient to become a victim of epistemic, and particularly hermeneutical, injustice. I argue in addition that this can come about because of the effect the diagnosis has on the recipient's self-narrative. As Tekin (2011) argues, the biomedical narrative associated with a DSM-based diagnosis implies that the patient's condition is a medical one that has somehow taken over her emotions and cognitive processes. As such, it implies that the patient does not have the same degree of control over her thinking and emotions as she might expect to have if she were fully healthy. It might also seem to her that her feelings of agency (i.e. her sense of self-efficacy) have been diminished, in that she lacks confidence in her ability to influence the course of her life in the future. She may have difficulty knowing to what extent her problems are due to her presumed medical condition or are something inherent in her – what Sadler (2004, 2007) refers to as self-illness ambiguity. Inasmuch as the diagnosis, and the process followed by the clinician in reaching the diagnosis, obscures or denies the reality of her own meanings, she can become a victim of hermeneutical injustice. This kind of injustice occurs, not just where the victims suffer from

-

<sup>&</sup>lt;sup>69</sup> Dillon's statement is also clear evidence of testimonial injustice, in that her report of systematic sexual abuse in her childhood was dismissed by the psychiatrist on the basis of a negative identity prejudice due to her psychological disturbance.

damaging absences in their hermeneutical resources, but also where harmful presences disrupt their own resources. A further element that seems to emerge from these testimonies is that many people felt their intellectual self-trust had been impaired. To the extent that their hermeneutical resources were reduced, there would likely have been a corresponding reduction in their level of intellectual self-trust.

#### 6.6 Possible objection: can a diagnosis enhance the patient's hermeneutical resources?

While I have argued that a psychiatric diagnosis can impair the patient's hermeneutical resources, and hence lead to hermeneutical injustice, it might be objected that the receipt of a psychiatric diagnosis might actually *add* something to the patient's hermeneutical resources. Similarly, on this objection, the *failure* to assign a diagnosis might contribute to hermeneutical injustice by depriving the patient of an important concept to help her make sense of her experiences. As I explained in the conclusion to Chapter 5, clinicians may prefer to arrive at a formulation of the patient's difficulties collaboratively with her, without offering a diagnosis. However, it might still be objected that the patient is being denied access to an important hermeneutical resource if no diagnosis is given to them. In support of this objection, some service-users report that they have found their diagnosis helpful in various ways.<sup>70</sup> Thus, for example, in a short video posted on the BBC website, in which users of mental health services describe their reaction to their diagnosis, one former serviceuser talks positively about a diagnosis of bipolar disorder, saying: "people suddenly realised I wasn't doing things for attention... It was because I have a brain disorder" (BBC, 2018). Other similar comments are reported in the various surveys reviewed by Perkins et al (2018). In one survey, a service-user reported: "It is good to put a name on somethings, because I knew there was something wrong... there must be a reason as to why I am like I am" (Lovell & Hardy, 2014).

\_

<sup>&</sup>lt;sup>70</sup> One way in which service-users find the diagnosis to be not only helpful, but also essential, is when it is necessary for claiming welfare benefits in the UK. In some cases, benefits are only provided when the individual has a formal diagnosis. Anecdotally, I have heard that even when psychiatrists prefer not to assign a diagnosis, they feel obliged to do so purely for this reason. Also, in the USA and other insurance-funded health systems, service-users need to have a diagnosis to ensure that their treatment costs are paid by the insurer. Such requirements, of course, apply regardless of whether the individual has a positive or negative reaction to the diagnosis. In these cases, a diagnosis is given because of stipulative pressures arising outside the mental health services, rather than because the clinician needs to assign it for clinical purposes. Consequently, I do not regard this reason for assigning a diagnosis as a relevant element in the objection discussed here.

There are several ways in which this objection can be answered. The first point to note is that the approach to case-formulation described at the end of Chapter 5 does not preclude the assignment of a diagnosis (Johnstone, 2022; Johnstone & Dallos, 2014). While a diagnosis is not essential in this approach, there is no reason not to give one if the patient requests it. Since, in this kind of situation of co-constructing a formulation, the focus of the clinician will be to help the patient gain an understanding of her situation with which she can feel satisfied. The stated diagnosis can be framed within the broader formulation. Precisely how this is done will vary with each individual case, but much will depend on how the statement of diagnosis is made and whether this leads to the kinds of reactions noted by the service-user reports above. Moreover, since the process of formulation has a different focus than the symptom oriented approach to a traditional diagnostic assessment, as Stanghellini (2004) explains, the patient will be encouraged to articulate her own feelings, which can be incorporated into the formulation in a way that often does not happen when the assessment is purely focussed on establishing a diagnosis.

A second, and related point, is whether the patient is led to believe, either unwittingly or deliberately on the part of the clinician, that the diagnosis somehow *explains* her condition. As explained in Chapter 2, psychiatric diagnoses do not usually give any explanatory information, unlike diagnoses of somatic diseases. Nevertheless, they can be understood by the patient as being explanatorily important. To the extent that the patient understands the diagnosis this way, she may be impeded from gaining an alternative, and therapeutically more helpful, way of understanding her thoughts and emotions. In such cases, her own hermeneutical resources may become diminished or at least may receive very little support, if she is led to believe that her psychological difficulties are caused by a biomedical condition, such as a brain disorder of some unspecified kind. When a diagnosis is communicated in this way, therefore, I claim that, rather than increasing the patient's hermeneutical resources, it can have the effect of diminishing the confidence she has in her existing resources and in some cases obscuring them altogether.

Thirdly, as many of the service-user testimonies indicate, patients are often told that the diagnosis means that their condition is a chronic one with little hope for recovery. As discussed in the case of schizophrenia in Chapter 2, the outcomes for people with the diagnosis are very variable, and there is generally no good reason for the patient to be told that her condition is chronic. Therefore, when the patient believes her condition is chronic, whether or not the clinician has explicitly stated this, she has in most cases been misinformed in some way. It is often stated that, where an individual has periods with no symptoms

followed by a recurrence of the symptoms, she has just had periods of remission from what is actually a continuing illness or disease. However, it is not clear why the recurrence of similar symptoms after a symptom-free period should mean that she has had the same illness all along. We would not, for example, say we have a single long-term disease called 'influenza' in which we have periodic outbreaks of influenza symptoms followed by long periods of remission from influenza when we feel well. As the service-user testimonies show, the belief that the condition is a long-term one can be a very devastating one. If this belief is not challenged in some way, it is likely to have a damaging effect on the patient's feelings of agency and limit her hopes of recovery. Consequently, the message that the condition is a chronic one is liable to undermine the patient's hermeneutical resources, rather than enhance them. I claim that such situations are particularly likely to lead to hermeneutical injustice.

More generally, there is some ambiguity about what the term 'diagnosis' actually means in practice in the context of psychiatry. As discussed in Chapter 4, a diagnosis delivered in the context of a specialist medical service typically conveys some explanatory information about the patient's condition. In everyday life, such as when the car or the washing-machine fails to work properly, we expect the technician to diagnose what the fault is, not simply to re-describe the malfunction. In common parlance, we expect a diagnosis to indicate an explanation of some kind. However, DSM-based diagnoses are purely symptom based. In effect, therefore, they function as a re-description in shorthand form of the patient's symptoms.<sup>71</sup> Nevertheless, this is frequently not understood by the patient who may believe that the diagnosis signifies an explanation and the existence of a brain disease of some kind.

Given its lack of explanatory value and the looseness with which the word *diagnosis* is often used in this context, a psychiatric diagnosis can be understood simply as a label for a defined cluster of symptoms. In some cases, the descriptive nature of the label is quite clear, as for example where a diagnosis of depression is made. Nevertheless, this can still lead to reification of the supposed underlying condition signified by the diagnosis (Hyman, 2010), on the basis that people can be tempted to assume that there must be some entity, however recondite, to which the word refers. This can particularly be the case when the presumed underlying condition is thought to be a chronic one from which the patient is unlikely to recover completely. Consequently, if the diagnosis is communicated in such a way as to make

\_

<sup>&</sup>lt;sup>71</sup> Moreover, they do not necessarily re-describe the symptoms adequately either. As explained in Chapter 2, the diagnosis of schizophrenia is based on polythetic criteria, such that two patients with the same diagnosis can have quite different symptoms. Therefore, the diagnosis often does not convey much information about the patient's actual symptoms.

clear what it does and does not mean, the patient is less likely to be misled by it. Therefore, careful communication can reduce or eliminate the likelihood of the diagnosis causing the patient to become a victim of hermeneutical injustice.

In ideal circumstances, therefore, a psychiatric diagnosis need not be a cause of hermeneutical injustice, and I am not claiming that a diagnosis leads to this in all cases. However, ideal circumstances generally do not prevail in mental health services and particularly not in those that are chronically under-resourced, such as in publicly funded services. In such cases, a shortage of resources almost invariably results in heavy pressures on staff time. As Carel and Kidd (2014) note, incidents of epistemic injustice in healthcare environments are particularly likely to occur when clinicians are working under pressure and when, therefore, they have insufficient time to pay attention to everything that the patient might want to say or to consider whether what the patient says is credible. Consequently, the claim that receiving a psychiatric diagnosis can lead to hermeneutical injustice for the recipient is a plausible one for many typical mental health services.

#### 6.7 Conclusion

Psychiatric diagnostic categories do not generally name distinct diseases in the way that diagnoses in general medicine typically do. Nevertheless, their use in a medicalised context can lead to the belief that they do, even though the supposed disease process, which can be assumed to be somewhere in the brain, is quite mysterious. Moreover, the frequent use of psychiatric medication for treating people with mental disorders can reinforce such a belief in the patient. As such, the belief that the patient's diagnosis represents a brain disease of some kind reflects the biomedical model of mental disorders and presupposes a biomedical narrative regarding the nature of the patient's condition. To the extent that the patient accepts this kind of narrative about his condition, other more personal self-narratives about why he feels the way he does will be relatively marginalised.

As discussed above, a diagnosis may be given to the patient in the context of a more elaborate formulation co-constructed with him. Perkins and co-workers (2018) comment, based on the results of their meta-review, that a case-formulation approach is a suitable alternative to a purely diagnosis-based approach to treatment and that it might improve the experience of diagnosis for those patients who find this a negative experience. Where such an approach is used, the biomedical narrative that might otherwise be conveyed by the diagnosis can be rendered less salient by the self-narrative of the patient that emerges from the

formulation, which by its very nature is individualised to him. However, diagnoses are frequently given to the patient without any such co-constructed formulation. When this happens, misunderstandings about what the diagnosis means are much more likely to occur, with the consequence that the patient can become a victim of hermeneutical injustice if his own hermeneutical resources are marginalised. Again, the time pressures on hard-pressed clinicians in most mental health services make this a more likely occurrence. As Romdenh-Romluc (2017) notes, the meanings assumed by the patient will be those of the dominant cultural group in this context – i.e. the medical profession and associated clinicians. When clinicians do not have the time or the necessary training to engage in a detailed formulation with the patient, but instead confine their assessment to the standard DSM-based diagnostic process, then it is more likely that the dominant biomedical narrative associated with the diagnosis will predominate over any personal narrative that the patient might have about his condition. It is particularly in such circumstances that I claim the patient can become a victim of hermeneutical injustice.

#### Chapter 7

## **Conclusion**

In this thesis, I have argued that people receiving a psychiatric diagnosis may become victims of hermeneutical injustice by virtue of their diagnosis and the manner in which it is communicated to them. This can happen because of the biomedical narrative associated with the diagnosis: the medicalization implicit in a psychiatric diagnosis conveys a particular kind of narrative about the nature of the patient's condition which may conflict with and radically change her previous self-narrative. It can convey the message that her condition is in some way caused by a mysterious disorder in her brain, one that not even the psychiatrist can explain to her. To the extent that it does this, the various events in her life and the ways in which she might have reacted to these can come to be seen as secondary in causal terms to the primary pathology presumed to be located in her brain. In this way, the biomedical narrative can become dominant in the patient's appraisal of her predicament. It can induce a sense of powerlessness in her – a sense that the abnormalities in her brain are beyond her control. This may reduce her sense of agency and induce feelings of hopelessness about recovery, which may then limit the prospect of a positive outcome for her.

In contrast to most medical diagnoses, psychiatric diagnoses generally lack validity and explanatory value. The conditions represented by the DSM diagnostic categories are not in most cases analogous to somatic diseases. They cannot in general be said to constitute natural kinds in medical or psychiatric science. However, these diagnostic categories continue to be used, because they are still judged by many psychiatrists to have utility in the clinical context (e.g. Jablensky, 2016). Moreover, patients are often encouraged to believe that their condition can be understood as "an illness like any other" which is represented by the diagnosis they are given. Such beliefs are reinforced by the epistemic authority vested in the medical profession.

There are alternative approaches that can mitigate or avoid such a consequence. These make greater use of case formulation, without the imposition of non-explanatory diagnoses that convey a biomedical narrative (Johnstone & Dallos, 2014). Such formulations are co-constructed with the patient with the aim of reaching an agreed understanding of the patient's difficulties – agreed, that is, between the patient and therapist. A formulation can be regarded as a narrative about how the patient came to be in the situation in which she finds herself. The role of the therapist is to work collaboratively with the patient in constructing the

formulation. This is intended to reinforce the individual's psychological resources to help her develop a more positive self-narrative, such that she feels more empowered to confront her difficulties. The agreed formulation can then be used as basis for appropriate psychological therapy.

A similar approach to conducting clinical encounters is suggested by Rosa Ritunnano (2022) who proposes that a critical phenomenology stance can help to avoid or minimise hermeneutical injustice. Such an approach should place more emphasis on engaging with the patient's subjective experience, particularly in cases where the patient is experiencing delusions that appear to be unintelligible to the clinicians involved. Rather than giving up on the attempt to understand the patient's statements, the clinician should attend to the patient's own descriptions of her state of mind and the manner in which she is trying to find meaning for herself. The clinician can thereby display an active interest in the particularities of her experience, rather than seeking to fit that experience into an inflexible medicalised category. In this way, the hermeneutical marginalisation that she might otherwise experience should be minimised.

However, such approaches are frequently not used in clinical practice, either because clinicians are not trained in them or because they appear too time-consuming to implement in highly pressured service contexts where resources are severely limited and individual clinicians have too many patients on their caseloads. The tendency in such circumstances will often be for the clinicians to resort to the use of assessment and diagnostic practices that come most readily to them. In the case of psychiatrists, this will typically be the diagnostic system, based on DSM-5 or ICD-11, on which they have been trained. The biomedical narrative associated with these diagnostic categories will therefore continue to influence the way that patients understand their condition. To the extent that patients are avoidably led to accept such a narrative, which does not explain their condition and which diminishes their ability to make sense of their experiences in more empowering terms, they can become victims of hermeneutical injustice.

If the claim that psychiatric patients can be wronged in this way is accepted, one might ask what individual clinicians could do to ameliorate it. This might be a difficult challenge for clinicians to address. In Fricker's view, hermeneutical justice is caused by social circumstances, not by individuals acting in unjust ways, and can arise in the sort of service contexts described above without any subject deliberately perpetrating it. She says: "No agent *perpetrates* hermeneutical injustice – it is a purely structural notion" (2007, p.159, italics in original). It is, she argues, fundamentally a consequence of the victim's hermeneutical

marginalisation. However, Medina (2013) argues for a slightly different view. He emphasises the variable dynamics in the many situations that can generate hermeneutical injustice, saying that in many situations the agent has some ability, and consequently some responsibility, to minimise or avoid such injustice occurring. He states that agents may perpetrate the injustice unknowingly and despite their best intentions. They may fail to help speakers communicate their experiences adequately by being insufficiently responsive to the speaker's attempts to do so. While Medina agrees with Fricker that such agents do not produce hermeneutical injustice on their own, he observes: "the communicative dynamics they participate in do help to reproduce them and to keep them in place" (2013, p.113). Such injustices are not generated by individuals or small groups, because they require frequent and consistent patterns of communication in a multiplicity of settings which constitute the background against which specific occurrences happen. However, he argues that agents who are actively involved in those patterns have some ability, which may be quite limited in some prevailing circumstances, to either repeat and reinforce the hermeneutical gap or to ameliorate it in the particular situation they find themselves in. Agents always have some freedom in what they can say and how they say it, whatever constraints may be operative in their situation.

Similar considerations apply in mental health services, however much resource constraints may limit what individual clinicians can do. It should not take up much time when communicating a psychiatric diagnosis to a patient to explain clearly what it means. It could be explained that it functions simply as a descriptive label for the patient's current condition: it could be made clear that it does not signify a biomedical disease and therefore does not explain how the condition arose. It could also be stated that the diagnosis does not imply that the condition is a long-term one and that there is no necessary reason why the patient cannot recover fully in due course. Such explanations need not be time consuming, though they may need to be repeated a few times. In such ways, as Medina suggests, clinicians can contribute to the gradual erosion of the entrenched communicative patterns in mental health services that have often given patients quite misleading messages about their prospects for recovery. The more that clinicians can manage to do this, the more likely it will be that the culture in the whole system can change correspondingly.

## References

- Aftab, A., Joshi, Y. & Sewell, D. (2020). Conceptualizations of mental disorder at a US academic medical center. *Journal of Nervous and Mental Disease*, 208, 848-856.
- Ahn, W-K. & Kim, N.S. (2008). Causal theories of mental disorder concepts. *Psychological Science Agenda*, 22, 3-8.
- American Psychiatric Association. (1980). *Diagnostic and Statistical Manual of Mental Disorders*. 3<sup>rd</sup> edn. Washington, DC: American Psychiatric Association (APA).
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders*. 4<sup>th</sup> edn. Washington, DC: American Psychiatric Association (APA).
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. 5<sup>th</sup> edn. Washington, DC: American Psychiatric Association (APA).
- Andersen, H. (2017). Reductionism in the biomedical sciences. In M. Solomon, J.R. Simon & H. Kincaid (Eds.) *The Routledge Companion to Philosophy of Medicine* (pp. 81-89). New York: Routledge.
- Andreasen, N.C. (1997). Linking mind and brain in the study of mental illnesses: a project for a scientific psychopathology. *Science*, *275*, 1586-1593.
- Angermeyer, M.C., Holzinger, A., Carta, M.G. & Schomerus, G. (2011). Biogenetic explanations and public acceptance of mental illness: systematic review of population studies. *British Journal of Psychiatry*, 199, 367-372.
- Aragona, M. (2015). Rethinking received views on the history of psychiatric nosology: minor shifts, major continuities. In P. Zachar, D.St. Stoyanov, M. Aragona & A. Jablensky (Eds). *Alternative perspectives on psychiatric validation: DSM, ICD, RDoC, and beyond.* Oxford: Oxford University Press. pp. 27-46.
- Arpaly, N. (2005). How it is not "just like diabetes": Mental disorders and the moral psychologist. *Philosophical Issues*, 15, 282-298.
- Austin, J. L. (1962). How to do things with words. Oxford: Oxford University Press.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*, 122-147.
- Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1–45). Cambridge: Cambridge University Press.
- Bandura, A. (1997). Self-efficacy: the exercise of control. New York: W.H. Freeman.
- Baumeister, R. (1999). Self-concept, self-esteem, and identity. In: V. J. Derlega, B. A.

- Winstead, & W. H. Jones (Eds.), *Nelson-Hall series in psychology. Personality: Contemporary theory and research* (p.339–375). Nelson-Hall Publishers.
- BBC (2018). <a href="http://www.bbc.co.uk/news/av/stories-44217657/">http://www.bbc.co.uk/news/av/stories-44217657/</a> how-does-your-diagnosis-make-you-feel. BBC. Accessed 7 May 2020.
- Beebee, H. & Sabbarton-Leary, N. (2010). Are psychiatric kinds "real"? European Journal of Analytical Philosophy, 6, 11-27.
- Bentall, R. (2009). Doctoring the Mind. London: Penguin.
- Bentall, R. (2017). Six myths about schizophrenia: a paradigm well beyond its use-by date? In J. Poland & Ş. Tekin (Eds). *Extraordinary Science and Psychiatry*. Cambridge, MA: MIT Press. pp221-248.
- Ben-Zeev, D., Young, M.A., & Corrigan, P.W. (2010). DSM-V and the stigma of mental illness. *Journal of Mental Health*, 19, 318-327.
- Bonnington, O. & Rose, D. (2014). Exploring stigmatisation among people diagnosed with either bipolar disorder or borderline personality disorder: a critical realist analysis. *Social Science & Medicine*, 123, 7-17.
- Boorse, C. (1975). On the distinction between disease and illness. *Philosophy and Public Affairs*, *5*, 49-68.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44, 542-573.
- Borsboom, D. (2017). Mental disorders, network models, and dynamical systems. In K.S. Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*. Oxford: Oxford University Press. pp81-97.
- Borsboom, D. & Cramer, A.O.J. (2013). Networks: an integrative to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91-121.
- Botterill, G. & Carruthers, P. (1999). *The Philosophy of Psychology*. Cambridge: Cambridge University Press.
- Boyd, R. (1979). Metaphor and theory change. In A. Ortony (ed.), *Metaphor and Thought*, pp. 356–408. Cambridge, UK: Cambridge University Press.
- Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, *61*, 127-148
- Boyd, R. (1999a). Homeostasis, species, and higher taxa. In R.A. Wilson (Ed). *Species: New interdisciplinary essays*. Cambridge, MA: MIT Press.
- Boyd, R. (1999b). Kinds, complexity and multiple realization. *Philosophical Studies*, 95, 67-98
- Boyd, R. (2021). Rethinking natural kinds, reference and truth: towards more

- correspondence with reality, not less. Synthese, 198 (Suppl 12), S2863-S2903.
- Brigandt, I. (2009). Natural kinds in evolution and systematics: metaphysical and epistemological considerations. *Acta Biotheoretica*, *57*, 77-97.
- British Psychological Society (2017). *Understanding psychosis and schizophrenia*. *Revised edition*. Leicester: British Psychological Society.
- Brückl, T.M., Spoormaker, V.I., Sämann, P.G., et al. (2020). The biological classification of mental disorders (BeCOME) study: a protocol for an observational deep-phenotyping study for the identification of biological sub-types. *BMC Psychiatry*, 20(213), 1-25.
- Bruner, J.S. (1990). Acts of Meaning. Cambridge, MA; Harvard University Press.
- Campbell, J. (2017a). Validity and the causal structure of a disorder. In K.S. Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*. Oxford: Oxford University Press. pp257-273.
- Campbell, J. (2017b). Why we should be realists about psychiatric disorders reply to Peter Zachar. In K.S. Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*. Oxford: Oxford University Press. pp303-310.
- Carel, H. & Kidd, I.J. (2014). Epistemic injustice in healthcare: a philosophical analysis. *Medicine Health Care and Philosophy*, 17, 529-540.
- Chakravartty, A. (2007). *A Metaphysics for Scientific Realism*. Cambridge: Cambridge University Press.
- Chakravartty, A. (2013). Dispositions for scientific realism. In R. Groff & J. Greco (Eds). *Powers and Capacities in Philosophy: The new Aristotelianism*. New York: Routledge.
- Chang, H. (2017). Epistemic iteration and natural kinds: realism and pluralism in taxonomy. In K.S. Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*. Oxford: Oxford University Press. pp230-245.
- Charney, D.S., Nestler, E.J., Sklar, P. & Buxbaum, J.D. (Eds.) (2017). *Charney and Nestler's Neurobiology of Mental Illness (5<sup>th</sup> edition)*. Oxford: Oxford University Press.
- Charon, R. (2006). *Narrative Medicine: Honoring the stories of illness*. New York: Oxford University Press.
- Christman, J. (2004). Narrative unity as a condition of personhood. *Metaphilosophy*, *35*, 695-713.
- Clare, A. (1976). *Psychiatry in dissent: Controversial issues in thought and practice*. London: Tavistock Publications.
- Cooper, R. (2002). Disease. Studies in History and Philosophy of the Biological and Biomedical Sciences, 33, 263-282.

- Cooper, R. (2007). *Psychiatry and Philosophy of Science*. Stocksfield: Acumen Publishing.
- Cooper, R. (2015). Why is the Diagnostic and Statistical Manual of Mental Disorders so hard to revise? Path-dependence and "lock-in" in classification. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 51, 1-10.
- Cooper, R. (2018). Understanding the DSM-5: stasis and change. *History of Psychiatry*, 29, 49-65.
- Cournoyea, M. & Kennedy, A.G. (2014). Causal explanatory pluralism and medically unexplained physical symptoms. *Journal of Evaluation in Clinical Practice*, 20, 928-933.
- Craddock, N., Antebi, D., Attenburrow, M., et al. (2008). Wake-up call for British psychiatry. *British Journal of Psychiatry*, 193, 6-9.
- Craddock, N. & Mynors-Wallis, L. (2014). Psychiatric diagnosis: Impersonal, imperfect and important. *British Journal of Psychiatry*, 204, 93-95.
- Craver, C. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, 22, 575-594.
- Crichton, P., Carel, H. & Kidd, I.J. (2017). Epistemic injustice in psychiatry. *BJPsych Bulletin*. 41, 65-70.
- Cuttle, G. & Burn, W. (2020). Neuroscience: the way forward. *BJPsych Advances*, 26, 318-319.
- Davies, W. (2016). Externalist psychiatry. *Analysis*, 76, 290-296.
- Dennett, D. (1992a). Consciousness explained. London: Viking.
- Dennett, D. (1992b). The self as a center of narrative gravity. In: F. Kessel, P. Cole, & Johnson, D. (Eds.) *Self and Consciousness: Multiple perspectives*. Hillsdale, NJ; Erlbaum.
- Dillon, J. (2011). The personal *is* the political. In M. Rapley, J. Moncrieff & J. Dillon (Eds.) *De-medicalizing misery*. (pp.141-157). Basingstoke: Palgrave Macmillan.
- Dinos, S., Stevens, S., Serfaty, M. et al. (2004). Stigma: the feelings and experiences of 46 people with mental illness. *British Journal of Psychiatry*, 184, 176-181.
- Division of Clinical Psychology (2013). Classification of behaviour and experience in relation to functional psychiatric diagnoses: time for a paradigm shift. Position statement. Leicester: British Psychological Society.
- Dragulinescu, S. (2010). Diseases as natural kinds. *Theoretical Medicine and Bioethics*, *31*, 347-369.

- Dupré, J. (1993). The Disorder of Things. Cambridge, MA: Harvard University Press.
- Ereshefsky, M. (2009). Defining 'health' and 'disease'. Studies in History and Philosophy of Biological and Biomedical sciences, 40, 221-227.
- First, M.B. (2012). The development of DSM-III from a historical/conceptual perspective. In K.S. Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry II: Psychiatric Nosology*. Oxford: Oxford University Press. pp127-140.
- Flanagan, O. (1994). Multiple identity, character transformation, and self-reclamation. In: G. Graham & G.L. Stevens (Eds.) *Philosophical Psychopathology*. (p.135-162). Cambridge, MA; MIT Press.
- Fodor, J. (1987). Psychosemantics. Cambridge, MA: MIT Press.
- Fricker, M. (2007). *Epistemic Injustice: power and the ethics of knowing*. Oxford: Oxford University Press.
- Fricker, M. (2016). Epistemic injustice and the preservation of ignorance. In R. Peels & M. Blaauw (Eds.). *Dimensions of Ignorance*. Cambridge: Cambridge University Press.
- Fricker, M. (2017). Evolving concepts of epistemic injustice. In I.J. Kidd, J. Medina & G. Pohlhaus Jr. (Eds) *Routledge Handbook of Epistemic Injustice*, pp.53-60. Abingdon, Oxon: Routledge.
- Fuchs, T. (2010). Subjectivity and intersubjectivity in psychiatric diagnosis. *Psychopathology*, 43, 268-274.
- Giere, R. (1999). Science without laws. Chicago: University of Chicago Press.
- Giere, R. (2004). How models are used to represent reality. *Philosophy of Science*, 71, 742-752.
- Gifford, F. (2017). The biomedical model and the biopsychosocial model in medicine. In M. Solomon, J.R. Simon & H. Kincaid (Eds.) *The Routledge Companion to Philosophy of Medicine* (pp. 445-454). New York: Routledge.
- Goffman, E. (1961). Asylums: Essays on the social situations of mental patients and other inmates. Harmondsworth: Penguin.
- Griffiths, P.E. (1999). Squaring the Circle: Natural kinds with historical essences. In R.A. Wilson (Ed). *Species: New interdisciplinary essays*. Cambridge, MA: MIT Press.
- Hacking, I. (2007). Natural kinds: rosy dawn, scholastic twilight. In A. O'Hear (Ed.) *Philosophy of Science (Philosophy – Royal Institute of Philosophy Supplement 61)*. Cambridge: Cambridge University Press.
- Hagen, R., Turkington, D., Berge, T. & Gråwe, R.W. (2011). *CBT for Psychosis: a symptom-based approach*. Hove, East Sussex: Routledge.

- Harper, D. & Spellman, D. (2014). Formulation and narrative therapy: Telling a different story. In L. Johnstone & R. Dallos (Eds.) (2014). *Formulation in psychology and psychotherapy: Making sense of people's problems.* (2<sup>nd</sup> edn). (pp. 96-120). Hove: Routledge.
- Harrison, G., Hopper, K., Craig, T., Laska, E., Siegel, C., Wanderling, J. et al. (2001). Recovery from psychotic illness. *British Journal of Psychiatry*, 178, 506-517.
- Haslam, N. (2014). Natural kinds in psychiatry: Conceptually implausible, empirically questionable, and stigmatizing. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying Psychopathology: Mental kinds and natural kinds* (pp.11-28). Cambridge, MA; MIT Press.
- Hawley, K. (2019). *How to be Trustworthy*. Oxford: Oxford University Press.
- Hawley, K. & Bird, A. (2011). What are natural kinds? *Philosophical Perspectives*, 25, 205-221.
- Ho, B.C., Andreason, N.C., Nopoulos, P., Arndt, S. et al. (2003). Progressive structural brain abnormalities and their relationship with to clinical outcome: a longitudinal magnetic resonance imaging study early in schizophrenia. *Archives of General Psychiatry*, 60, 585-594.
- Hookway, C. (2010). Some varieties of epistemic injustice: reflections on Fricker. *Episteme*, 7, 151-163.
- Hornsby, J. (1994). Illocution and its significance. In S.L. Tsohatzidis (Ed.)

  Foundations of speech act theory: philosophical and linguistic perspectives. London: Routledge.
- Horwitz, A.V. (2015). The DSM-5 and the continuing transformation of normal sadness into depressive disorder. *Emotion Review*, 7, 209-215.
- Hucklenbroich, P. (2014). "Disease Entity" as the key theoretical concept in medicine. *Journal of Medicine and Philosophy*, 39, 609-633.
- Huda, A.S. (2019). *The Medical Model in Mental Health: An explanation and evaluation*. Oxford: Oxford University Press.
- Huda, A.S. (2021). The medical model and its application in mental health. *International Review of Psychiatry*, *33*, 463-470.
- Hyman, S. (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology*, *6*, 155-179.
- Insel, T. (2013). Director's blog: Transforming diagnosis. April 29. http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml.
- Jablensky, A. (2016). Psychiatric classifications: validity and utility. World

- Psychiatry, 15, 26-31.
- James, H. (1864-1915/1999). *Henry James: a life in letters*. Edited by P. Horne. London: Penguin.
- Johnstone, L. (2022). *A straight talking introduction to psychiatric diagnosis.* 2<sup>nd</sup> edition. Monmouth: PCCS Books.
- Johnstone, L. & Dallos, R. (Eds.) (2014). *Formulation in psychology and psychotherapy: Making sense of people's problems.* (2<sup>nd</sup> edn). Hove: Routledge.
- Jones, K. (2012). The politics of intellectual self-trust. Social Epistemology, 26, 237-251.
- Jutel, A. (2009). Sociology of diagnosis: A preliminary review. *Sociology of Health and Illness*, 31, 278-299.
- Jutel, A. (2017). Diagnostic categories. In M. Solomon, J.R. Simon & H. Kincaid (Eds.). *The Routledge Companion to Philosophy of Medicine*. Abingdon, Oxon: Routledge.
- Kasanin, J. (1933). The acute schizoaffective psychoses. *American Journal of Psychiatry*, 90, 97-126.
- Kendell, R.E. (1991). The major functional psychoses: are they independent entities or part of a continuum? In A. Kerr & H. McClelland (Eds). *Concepts of Mental Disorder: a continuing debate*. London: Gaskell. Pp1-16.
- Kendell, R.E. (2001). The distinction between mental and physical illness. *British Journal of Psychiatry*, 178, 490-493.
- Kendell, R.E., Cooper, J., Gourlay, A., Copeland, J. et al. (1971). Diagnostic criteria of American and British psychiatrists. *Archives of General Psychiatry*, 25, 123-130.
- Kendell, R.E. & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, 160, 4-12.
- Kendler, K.S. (2015). A joint history of the nature of genetic variation and the nature of schizophrenia. *Molecular Psychiatry*, 20, 77-83.
- Kendler, K.S., Kupfer, D., Narrow, W. et al. (2009). Guidelines for making changes to DSM-V. Washington, DC: American Psychiatric Association. Unpublished paper.
- Keshavan, M.S., Morris, D.W., Sweeney, J.A., Pearlson, G. et al. (2011). A dimensional analysis to the psychosis spectrum between bipolar disorder and schizophrenia: the Schizo-Bipolar Scale. *Schizophrenia Research*, 133, 250-254.
- Khalidi, M.A. (2013). *Natural Categories and Human Kinds*. Cambridge: Cambridge University Press.
- Khalidi, M.A. (2018). Natural kinds as nodes in causal networks. *Synthese*, 195, 1379-1396.

- Kidd, I.J. & Carel, H. (2017). Epistemic injustice and illness. *Journal of Applied Philosophy*, 34, 172-190.
- Kidd, I.J. & Carel, H. (2018). Healthcare practice, epistemic injustice, and naturalism. *Royal Institute of Philosophy Supplement*, 84, 211-233.
- Kidd, I.J. & Carel, H. (2021). The predicament of patients. *Royal Institute of Philosophy Supplement*, 89, 65-84.
- Kiesler, D.J. (1999). *Beyond the disease model of mental disorders*. Westport, CT: Praeger Publishers.
- Kim, J. (1998). *Mind in a physical world: an essay on the mind-body problem and mental causation*. Cambridge, MA: MIT Press.
- Kinderman, P., Read, J., Moncrieff, J. & Bentall, R. (2013). Drop the language of disorder. *Evidence-Based Mental Health*, 16, 2-3.
- Kirmayer, L.J., Groleau, D., Looper, K.J. & Dao, M.D. (2004). Explaining medically unexplained symptoms. *Canadian Journal of Psychiatry*, 49, 663-671.
- Kleinman, A. (1988). *Illness Narratives: Suffering, healing & the human condition*. New York; Basic Books.
- Kornblith, H. (1993). *Inductive Inference and its Natural Ground*. Cambridge, MA: MIT Press.
- Kripke, S. (1980). Naming and Necessity. Oxford: Blackwell.
- Kurs, R. & Grinshpoon, A. (2018). Vulnerability of individuals with mental disorders to epistemic injustice in both clinical and social domains. *Ethics & Behavior*, 28, 336-346.
- Kyle, T.K., Dhurandhar, E.J. & Allison, D.B. (2016). Regarding obesity as a disease: Evolving policies and their implications. *Endocrinology & Metabolism Clinics of North America*, 45, 511-520.
- Laing, R.D. (1969). *The politics of the family and other essays*. London: Tavistock Publishers.
- Lancet Psychiatry (2015). This year's model. Lancet Psychiatry, 2, 477.
- Lange, M. (2007). The end of diseases. *Philosophical Topics*, 35, 265-292.
- Langton, R. (2010). Epistemic injustice: power and the ethics of knowing. By Miranda Fricker. *Hypatia*, 25, 459-464.
- Lewis, D (1986). Causal explanation. In *Philosophical papers* (Vol. 2, pp. 214-240). Oxford: Oxford University Press.
- Li, Y & He, Q. (2021). Is mental illness like any other medical illness? Causal attributions,

- supportive communications and the social withdrawal inclinations of people with chronic mental illnesses in China. *Health Communication*, *36*, 1949-1960.
- Lichtenstein, P., Yip, B.H., Bjork, C., Pawitan, Y. et al. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*, *373*, 234-239.
- Lieberman, J.A. & First, M.B. (2007). Renaming schizophrenia: diagnosis and treatment are more important than semantics. *British Medical Journal*, 334, 108.
- Locke, J. (1706/1997). *An Essay concerning Human Understanding (5<sup>th</sup> ed)*. London: Penguin.
- Lovell, L.J. & Hardy, G. (2014). Having a diagnosis of borderline personality disorder in a forensic setting: a qualitative exploration. *Journal of Forensic Practice*, 16, 228-240.
- Lumsden, D. (2013). Whole life narratives and the self. *Philosophy, Psychiatry, & Psychology, 20*, 1-10.
- MacIntyre, A. (2007). After Virtue. (3rd edn). London: Bloomsbury.
- MacLeod, M. & Reydon, T.A.C. (2013). Natural kinds in philosophy and in the life sciences: scholastic twilight or new dawn? *Biological Theory*, 7, 89-99.
- Magnus, P.D. (2018). Taxonomy, ontology, and natural kinds. *Synthese*, 195, 1427-1439.
- Malla, A., Joober, R., & Garcia, A. (2015). "Mental illness is like any other medical illness": a critical examination of the statement and its impact on patient care and society. *Journal of Psychiatry and Neuroscience*, 40, 147-150.
- Mason, R. (2011). Two kinds of unknowing. *Hypatia*, 26, 295-307.
- Maung, H.H. (2016). Diagnosis and causal explanation in psychiatry. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 60, 15-24.
- Maung, H.H. (2017). The causal explanatory functions of medical diagnoses. *Theoretical Medicine and Bioethics*, *38*, 41-59.
- Maung, H. (2019). The functions of diagnoses in medicine and psychiatry. In Ş. Tekin & R. Bluhm (Eds.), *The Bloomsbury Companion to Philosophy of Psychiatry* (pp. 507-524). London: Bloomsbury.
- McCormack, L. & Thomson, S. (2017). Complex trauma in childhood, a psychiatric diagnosis in adulthood: making meaning of a double-edged phenomenon. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*, 156-165.
- McGann, P.J. (2011). Troubling diagnoses. *Sociology of Diagnosis: Advances in Medical Sociology*, 12, 331-362.

- Medina, J. (2013). The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination. Oxford: Oxford University Press.
- Medina, J. (2017). Varieties of hermeneutical injustice. In I.J. Kidd, J. Medina, & G. Pohlhaus, Jr. (Eds.) *The Routledge Handbook of Epistemic Injustice*, pp. 41-52. Abingdon, Oxon: Routledge.
- Mill, J.S. (1843). *A system of logic: ratiocinative and inductive*. Toronto: University of Toronto Press.
- Mills, C. (2007). White ignorance. In S. Sullivan & N. Tuana (Eds.) *Race and Epistemologies of Ignorance*, pp. 11-38. Albany, NY: SUNY Press.
- Moncrieff, J. (2014). The nature of mental disorder: Disease, distress, or personal tendency? *Philosophy, Psychiatry, and Psychology, 21*, 257-260.
- Moncrieff, J. (2020). A straight talking introduction to psychiatric drugs. 2<sup>nd</sup> edition. Monmouth: PCCS Books.
- Moncrieff, J. & Crawford, M.J. (2001). British psychiatry in the 20<sup>th</sup> century observations from a psychiatric journal. *Social Science and Medicine*, 53, 349-356.
- Moncrieff, J. & Middleton, H. (2015). Schizophrenia: a critical psychiatry perspective. *Current Opinion in Psychiatry*, 28, 264-268.
- Mulheirn, I. (2013). The myth of the "welfare scrounger". New Statesman, 15 March 2013.
- Murphy, D. (2010). Explanation in psychiatry. *Philosophy Compass*, 5/7, 602-610.
- Murphy, D. (2013). The medical model and the philosophy of science. In K.W.M. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini & T. Thornton (Eds.), *The Oxford Handbook of Philosophy and Psychiatry* (pp. 966-986). Oxford: Oxford University Press.
- Murphy, D. (2014). Natural kinds in folk psychology and in psychiatry. In H. Kincaid & J.A. Sullivan (Eds.), *Classifying Psychopathology: Mental kinds and natural kinds* (pp105-122). Cambridge, MA: MIT Press.
- Murray, R.M. (2017). Mistakes I have made in my research career. *Schizophrenia Bulletin*, 43, 253-256.
- Olson, E. & Witt, K. (2019). Narrative and persistence. *Canadian Journal of Philosophy*, 49, 419-434.
- Parsons, T. (1951). The Social System. New York, NY: The Free Press.
- Perkins, A., Ridler, J., Browes, D. et al. (2018). Experiencing mental health diagnosis: a systematic review of service user, clinician, and carer perspectives across clinical settings. *The LancetPsychiatry*, *Sep*; 5(9): 747-764.

- Pescolido, B.A., Martin, J.K., Long, J.S. et al. (2010). "A disease like any other"? A decade of change in public reactions to schizophrenia, depression, and alcohol dependence. *American Journal of Psychiatry*, 167, 1321-1330.
- Phillips, J. (2013). On narrative: Psychopathology informing philosophy. *Philosophy, Psychiatry & Psychology, 20,* 11-23.
- Pitt, L., Kilbride, M., Welford, M., et al. (2009). Impact of a diagnosis of psychosis: user-led qualitative study. *Psychiatric Bulletin*, *33*, 419-423.
- Pohlhaus, G. (2012). Relational knowing and epistemic injustice: Towards a theory of willful hermeneutical ignorance. *Hypatia*, 27, 715-735.
- Poland, J. (2014). Deeply rooted sources of error and bias in psychiatric classification. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying Psychopathology: Mental kinds and natural kinds* (pp. 29-63). Cambridge, MA: MIT Press.
- Putnam, H. (1975). The meaning of 'meaning'. In H. Putnam (Ed.) *Mind, Language and Reality: Philosophical papers*. Cambridge: Cambridge University Press.
- Quine, W.V. (1969). Natural kinds. In W.V. Quine (Ed.) *Ontological Relativity and other Essays*. New York: Columbia University Press.
- Read, J. (2013). Does 'schizophrenia' exist? Reliability and validity. In J. Read & J. Dillon (Eds.), *Models of Madness: Psychological, social and biological approaches to psychosis* (2nd ed.) (pp. 47-61). London: Routledge.
- Read, J., Haslam, N., Sayce, L. & Davies, E. (2006). Prejudice and schizophrenia: A review of the 'mental illness is like any other, approach. *Acta Psychiatrica Scandicavica*, 114, 303-318.
- Read, J., Perry, B.D., Moskowitz, A. & Connolly, J. (2001). A traumagenic neurodevelopmental model of schizophrenia. *Psychiatry*, 64, 319-345.
- Regier, D.A. (2012). Diagnostic threshold considerations for DSM-5. In K.S.Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry II: Nosology*. 2012, pp. 285-297. Oxford: Oxford University Press.
- Regier, D.A., Narrow, W.E., Kuhl, E.A. & Kupfer, D.J. (2009) The conceptual development of DSM-V. *American Journal of Psychiatry*, 166, 645-650.
- Reiss, J. & Ankeny, R.A. (2016). Philosophy of Medicine. *Stanford Encylopedia of Philosophy*.
- Repnikov, G. & Murphy, D. (2017). Saving the explananda. In K.S. Kendler & J. Parnas (Eds). *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*. Oxford: Oxford University Press. pp274-281.
- Reznek, L. (1995). Dis-ease about kinds: reply to D'Amico. *Journal of Medicine and Philosophy*, 20, 571-584.

- Rieppel, O. (2005). Monophyly, paraphyly, and natural kinds. *Biology and Philosophy*, 20, 465-487.
- Ritunnano, R. (2022). Overcoming hermeneutical injustice in mental health: a role for critical phenomenology. *Journal of the British Society for Phenomenology*, *53*, 243-260.
- Robins, E. & Guze, S. (1970). Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry*, 126, 983-987.
- Romdenh-Romluc, K. (2017). Hermeneutical injustice and the problem of authority. *Feminist Philosophy Quarterly*, 3 (3), 1-23.
- Rose, N. (2013). What is diagnosis for? Lecture given at the Institute of Psychiatry, London, 4 June 2013. Downloaded from nikolasrose.com, 1 October 2020.
- Rosenberg, C. (2002). The tyranny of diagnosis: Specific entities and individual experience. *Milbank Quarterly*, 80, 237-260.
- Rosenberg, C. (2006). Contested boundaries: Psychiatry, disease, and diagnosis. *Perspectives in Biology and Medicine, 49*, 407-424.
- Rudd, A. (2007). In defence of narrative. European Journal of Philosophy, 17, 60-75.
- Rüsch, N., Angermeyer, M.C. & Corrigan, P.W. (2005). Mental illness stigma: Concepts, consequences, and initiatives to reduce stigma. *European Psychiatry*, 20, 529-539.
- Sadegh-Zadeh, K (2011). The logic of diagnosis. In F. Gifford (Ed.) *Handbook of philosophy of science. Vol. 16. Philosophy of medicine.* Oxford: Elsevier.
- Sadler, J.Z. (2004). Values and Psychiatric Diagnosis. Oxford: Oxford University Press.
- Sadler, J.Z. (2007). The psychiatric significance of the personal self. *Psychiatry*, 70, 113-129.
- Saks, E. (2007). *The Center Cannot Hold: My journey through madness*. New York: Hyperion Books.
- Samuels, R. (2009). Delusion as a natural kind. In M. Broome & L. Bortolotti (Eds). *Psychiatry as Cognitive Neuroscience: Philosophical perspectives* (pp49-79). Oxford: Oxford University Press.
- Sanati, A. & Kyratsous, M. (2015). Epistemic injustice in assessment of delusions. *Journal of evaluation in clinical practice*, 21, 479-485.
- Schaffner, K.F. (2012). A philosophical overview of the problems of validity for psychiatric disorders. In K.S. Kendler & J. Parnas (Eds.) *Philosophical issues in psychiatry II: Nosology.* pp. 169-189. Oxford: Oxford University Press.
- Schechtman, M. (1996). The Constitution of Selves. Ithaca: Cornell University Press.

- Schechtman, M. (2007). Stories, lives, and basic survival: a refinement and defense of the narrative view. *Royal Institute of Philosophy Supplement*, 82(60), 1550-178.
- Scrutton, A.P. (2017). Epistemic injustice and mental illness. In I.J. Kidd, J. Medina & G. Pohlhaus Jr. (Eds) *Routledge Handbook of Epistemic Injustice*, pp.347-355. Abingdon, Oxon: Routledge.
- Searle, J. (1995). The construction of social reality. London: Penguin.
- Searle, J. (2010). *Making the social world: the structure of human civilization*. Oxford: Oxford University Press.
- Simon, J. (2008). Constructive realism: an approach to medical ontology. *Perspectives in Biology and Medicine, 51*, 353-366.
- Simon, J. (2010). Advertisement for the ontology of medicine. *Theoretical Medicine & Bioethics*, 31, 333-346.
- Sisti, D. & Caplan, A.L. (2017). The concept of disease. In M. Solomon, J.R. Simon & H. Kincaid (Eds.) *The Routledge Companion to Philosophy of Medicine* (pp. 1-15). New York: Routledge.
- Sitko, K., Bentall, R.P., Shevlin, N. et al. (2014). Associations between specific psychotic symptoms and specific childhood adversities are mediated by attachment styles: an analysis of the National Comorbidity Survey. *Psychiatry Research*, 217, 202-209.
- Slater, M.H. (2015). Natural kindness. *British Journal for the Philosophy of Science*, 66, 375-411.
- Slater, E. & Roth, M. (1977). *Clinical Psychiatry*. *3<sup>rd</sup> edition, Revised reprint*. London: Ballière Tindall.
- Solomon, M. (2015). Making Medical Knowledge. Oxford: Oxford University Press.
- Stanghellini, G. (2004). The puzzle of the psychiatric interview. *Journal of Phenomenological Psychology*, 35, 173-195.
- Stegenga, J. (2015). Effectiveness of medical interventions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *54*, 34-44.
- Stegenga, J. (2018). *Care and cure: An introduction to the philosophy of medicine*. Chicago: University of Chicago Press.
- Stein, D.J. (2014). Psychopharmacology and natural kinds: A conceptual framework. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying Psychopathology: Mental kinds and natural kinds* (pp. 65-73). Cambridge, MA: MIT Press.
- Stempsey, W.E. (2000). Disease and Diagnosis: value-dependent realism. Dordrecht:

#### Kluwer.

- Stoner, L. & Cornwall, J. (2014). Did the American Medical Association make the correct decision classifying obesity as a disease? *Australasian Medical Journal*, 7(11), 462-464.
- Strawson, G. (2004). Against narrativity. Ratio, 17, 429-452.
- Sulmasy, D.P. (2005). Diseases and natural kinds. *Theoretical Medicine and Bioethics*, 26, 487-513.
- Tabb, K. (2015). Psychiatric progress and the assumption of diagnostic discrimination. *Philosophy of Science*, 82, 1047-1058.
- Taylor, C. (1989). Sources of the Self: The making of the modern identity. Cambridge, MA; Harvard University Press.
- Tekin, Ş. (2011). Self-concept through the diagnostic looking glass: narratives and mental disorder. *Philosophical Psychology*, 24, 357-380.
- Tekin, Ş. (2014). Self-insight in the time of mood disorders: after the diagnosis, beyond the treatment. *Philosophy, Psychiatry, & Psychology, 21*, 139-155.
- Thagard, P. (1999). *How scientists explain disease*. Princeton: Princeton University Press.
- Thomas, P., Seebohm, P., Wallcraft, J. et al. (2013). Personal consequences of the diagnosis of schizophrenia: a preliminary report from the inquiry into the schizophrenia label. *Mental Health and Social Inclusion*, 17, 135-139.
- Thornton, T. (2010). Narrative rather than idiographic approaches as counterpart to the nomothetic approach to assessment. *Psychopathology*, 43, 252-261.
- Tikkinen, K.A.O., Rutanen, J., Frances, A. et al. (2019). Public, health professional and legislator perspectives on the concept of psychiatric disease: a population-based survey. *BMJ Open*, *9*, 1-8.
- Tsou, J.Y. (2013). Depression and suicide are natural kinds: Implications for physician-assisted suicide. *International Journal of Law and Psychiatry*, *5-6*, 461-470.
- Tyrer, P. (2014). A comparison of the DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment*, 20, 280-285.
- Ure, J.A., Corral, R. & Wainwright, E. (2018). Schizophrenia and brain networks. *Neurology and Neurosurgery*, 1(1), DOI: 10.15761/NNS.1000102
- Valles, S. (2020). Philosophy of biomedicine. *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <a href="https://plato.stanford.edu/archives/sum2020/entries/biomedicine/">https://plato.stanford.edu/archives/sum2020/entries/biomedicine/</a>>.

- Wakefield, J. (1992). The concept of mental disorder: on the boundary between biological facts and social values. *American Psychologist*, 47, 373-388.
- Wakefield, J. (2007). The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6, 149-156.
- Whitbeck, C. (1977). Causation in medicine: The disease entity model. *Philosophy of Science*, 44, 619-637.
- Whitbeck, C. (1978). Four basic concepts of medical science. In P.D. Asquith & I. Hacking (Eds.) *PSA: Proceedings of the 1978 biennial meeting of the Philosophy of Science Association*, pp. 210-222. Philosophy of Science Association, East Lancing, MI.
- Whitbeck, C. (1981). What is diagnosis? Some critical reflections. *Metamedicine*, 2, 319-329.
- Williams, N.E. (2011). Arthritis and nature's joints. In J.K. Campbell, M. O'Rourke & M.H. Slater (Eds.) *Carving Nature at its Joints*. Cambridge, MA: MIT Press.
- Wilson, R.A., Barker, M.J. & Brigandt, I. (2007). When traditional essentialism fails: biological natural kinds. *Philosophical Topics*, *35*, 189-215.
- Witt, K. (2020). Narrative and characterization. Erkenntnis, 85, 45-63.
- Wittgenstein, L. (1958). *Philosophical Investigations*. Trans. G.E.M. Anscombe. Oxford: Blackwell.
- Woodward, J. (2003). *Making Things Happen: A theory of causal explanation*. Oxford: Oxford University Press.
- World Health Organisation (2019/2021). *International statistical classification of diseases* and related health problems. Eleventh revision. (ICD-11) Geneva: WHO.
- Yanos, P.T., Roe, D. & Lysaker, P.H. (2010). The impact of illness identity on recovery from severe mental illness. *American Journal of Psychiatric Rehabilitation*, 13, 73-93.
- Zachar, P. (2000). Psychiatric disorders are not natural kinds. *Philosophy, Psychiatry, and Psychology, 7*, 167-182.
- Zipursky, R.B., Reilly, T.J. & Murray, R.M. (2013). The myth of schizophrenia as a progressive brain disease. *Schizophrenia Bulletin*, *39*, 1363-1372.