IN DEFENSE OF MENTAL DISORDERS AS HOMEOSTATIC PROPERTY CLUSTER KINDS

By

Zinet Ritschel

Submitted to

Central European University

Department of Philosophy

In partial fulfillment of the requirements for the degree of Masters of Art

Supervisor: Maria Kronfeldner

Vienna, Austria

2021

Abstract

In this thesis I argue that mental disorders are best understood as symptomatic property clusters supported by homeostatic biological and non-biological causal mechanisms. This conception draws on Richard Boyd's homeostatic property cluster theory. I reject two rival non-causal kind theories, arguing that they are only successful to the extent that they depend on tacit causal assumptions, and that they cannot otherwise adequately account for the case of mental disorders. In response to the problem of multifactorial causation for the homeostatic property cluster view, I propose a distinction between proximal and distal causes as the basis for a non-pragmatic causal hierarchy which can justify prioritising proximal causes in causal selection. Finally, I argue that other non-epistemic aims and values can legitimately undermine the priority of proximal causes in causal selection. I conclude that the homeostatic property cluster theory provides an ideal model of mental disorders. Because this model is grounded in genuine causal relations, it maintains the possibility of epistemic value for mental disorder categories and resists collapse into pure constructionism. Since the model is equally equipped to accommodate the role of normative influences in causal selection it can also account for the interest-relative nature of the resulting classifications and explanations.

Acknowledgements

I am grateful to have had the good fortune of writing this thesis under the kind and competent tutelage of Maria Kronfeldner. Being her student has been a privilege as well as a pleasure.

Table of contents

Introduction

Recent years have seen an upsurge of philosophical interest in questions pertaining to the nature of mental disorders. Some philosophers argue that mental disorders are natural kinds which reflect real patterns and divisions in the world (Cooper, 2007; Kendler et al, 2011; Tsou, 2008, 2016). This reflects a realist view according to which mental disorders are something we can discover by means of empirical investigation similarly to how chemical elements have been discovered by the physical sciences. For example, rather than having been merely invented as a category, PTSD can be said to have been discovered in the wake of the Vietnam War largely due to the specific symptoms suffered by a large number of traumatised veterans (Crocq, 2000).

In contrast, other authors have argued in favour of a purely constructivist position according to which mental disorders are merely reflections of our classificatory practices and nothing more (Szasz, 1962; Foucault, 1961). On this view, mental disorders as such are a myth and psychiatric categories are more akin to culturally embedded terms like 'criminal', which do not reflect anything invariable in the world, but simply mirrors the normative structure of the relevant cultural context. According to the constructivists, mental disorders are not discovered, but invented.

Most recent philosophers who oppose the natural kind-view of mental disorders agree with the constructivists that mental disorders are too deeply rooted in shifting cultural values and therefore too unstable to be considered natural kinds like chemical elements (Hacking, 1992, 1995, 2007; Hyman, 2010; Zachar, 2000, 2002). But instead of rejecting the reality of mental disorders all together, they instead propose third-way theories which acknowledge the reality of specific kinds of mental suffering that can be studied and intervened on, while nevertheless insisting that these kinds reflect social and classificatory conventions to an extent that makes them unsuitable for being considered natural kinds akin to those found in the physical sciences (Murphy 2006; Tsou 2008; Zachar 2002; Kender et. al 2011). It is with such third-way theories

that this thesis is mainly concerned.

Contemporary psychiatry is predominantly characterised by what has been termed 'the biomedical model' which conceptualises mental disorders along the lines of natural kinds (Luhrmann, 2001; Boorsboom et al. 2019; Olbert & Gala, 2015; Insel & Cuthbert, 2015). This discourse draws on the heritage of the "disease entity" model in medicine which takes medical diagnoses to correspond to particular and repeatable causal types (Hucklenbroich, 2014). On this view, tuberculosis is for example simply identical to an infection with the bacterium Mycobacterium tuberculosis. This bacterial infection is the cause of the associated symptoms such as coughing.

According to the biomedical model of psychiatry, mental disorders are symptoms of underlying neurological dysfunctions which are relatively uniform across patients. It follows that in principle mental disorders can be treated with no reference to their psychological content or social aspects. Because mental disorders are understood as brain dysfunctions, much psychiatric research is focused on the development of neuro-pharmaceutical medication aimed to treat the proposed neurological disturbances associated with the disorders.

Especially popular is the 'monoamine hypothesis' according to which the so-called major depressive disorder (MDD) is due to a deficiency in monoamine neurotransmitters such as noradrenaline, serotonin, and dopamine (Delgado 2000; Hirschfeld 2000; Owens 2004). The currently most widely prescribed anti-depressant drugs function by elevating the neural levels of one or several of these neurotransmitters to compensate for this alleged deficiency (Harmer et al, 2017). The current recommended first-line medical treatment for major depressive disorders is "selective serotonin reuptake inhibitors" (SSRIs) which block neural re-absorption of serotonin and thus raises the overall neural serotonin levels (Koenig and Thase, 2009). In 2018, 4.8 million English patients - 10% of the adult population of England - took antidepressant medication,

predominantly SSRIs, prescribed by the public health system. Of these patients, 1.1 million (23%) had received the medication continuously for three years, while the median length of treatment was 2 years (Kendrick, 2020). The most comprehensive recent meta-analysis of antidepressant efficiency, including 522 clinical trials, 116477 patients, and 21 different antidepressant drugs, concluded that anti-depressant drugs show a moderate effect on the symptoms of severely depressed patients over an 8-week period relative to placebo (Cipriani et al, 2018).

The conception of MDD as identical with a neurochemical imbalance is not only broadly accepted within the realm of psychiatry and medicine, but has also come to dominate the general cultural narrative about mental disorders (Frances 2015; Jackson 2005; Whitaker, 2011). Boorsboom et al. (2019) point out that the information about mental disorders which is communicated to the general public by government-funded scientific organisations is often articulated in biomedical terms and therefore likely to be at least partly responsible for the prevalence of this popularity.

Apart from medication several other empirically supported treatments of MDD are proven to have at least a moderate effect. Such approaches include conversational and mindfulness-based therapy (Cujipers et al, 2008; Chi et al., 2018) as well as more somatic approaches such as yoga (Bridges & Sharma, 2017; Prathikanti et al, 2017), and massage (Arnold et al, 2020). Hence, neuro-pharmaceutical interventions are not the only - or necessarily the best – way of treating MDD (Kendler, 2012; Lilienfeld, 2014; Miller, 2010).

It is my impression that much recent philosophical resistance to the natural-kind conception of mental disorders is motivated by a desire to reject the monopoly of the biomedical model in psychiatry, protest what is perceived as psychiatry's pathologisation of normal emotional responses, and oppose the role of neuro-pharmaceutical medicine as a primary means of

psychiatric treatment. I sympathise with these sentiments and agree with the importance of such opposition. However, it appears to me that in the process of defending these claims, the baby is often thrown out with the bathwater. If psychiatry is to have any legitimacy as a science, we must ensure that its objects of study are sufficiently stable to serve as such.

In this thesis, I will argue that Richard Boyd's homeostatic property cluster (HPC) theory (1999, 1999b) is an ideal model for genuine mental disorders since it can equally acknowledge both their biological and social aspects and can admit causation an adequate role in accounting for the stability of mental disorders as kinds. For this purpose, I will use major depressive disorder as my primary example. The structure of the thesis is as follows:

In chapter 1, I introduce and explain the essentialist natural kind theory and show how the biomedical model draws on this essentialist tradition in the way it conceptualises mental disorders. I explain what the epistemic motivations might be for adhering to the essentialist model in psychiatry. I go on to argue that an essentialist conception of MDD is not viable. I then introduce the view I will defend throughout this thesis, namely the conception of MDD as an HPC kind, drawing on the writing of Richard Boyd (ibid) and Kenneth Kendler et al. (2011).

In chapter 2, I first show how DSM categories can contain causal information despite being formally defined by their symptoms. Specifically, these categories possess negative and probabilistic causal information. I then go on to question whether an appeal to causation is even necessary for a useful conception of MDD. For this purpose, I assess two non-causal, pragmatic theories of kinds. I will argue that these theories hinge on implicit causal assumptions and cannot adequately account for MDD in the absence of causal reasoning. I further argue that the HPC model is especially well-suited to accommodate the negative and probabilistic causal information characteristic of MDD.

In chapter 3, I respond to an objection against the HPC model on the grounds of multifactorial

causation. I advance a distinction between proximal and distal causes, arguing that proximal causes are more closely connected to the effect, and that this justifies their non-pragmatic priority in causal selection. I then show how neurobiology can be identified as a proximal cause in the case of MDD. I suggest this partly justifies - and perhaps has partly motivated - the prevalent focus on neurobiology in psychiatry.

In chapter 4, I maintain a commitment to the distal/proximal distinction as grounding a non-pragmatic causal hierarchy. However, I further argue in favour of the legitimate role of pragmatic and normative influences in causal selection and assert that such considerations can trump the prima facie superiority of proximal causes. I argue that such non-epistemic influences are both inevitable and potentially legitimate in our classificatory practices, but that their influence ought to be well-reflected and transparent.

I finally conclude that the homeostatic property cluster theory is well-suited for accounting for the nature of MDD. It constitutes a golden middle way between essentialism and mere pragmatism insofar as it allows for pragmatic and normative considerations to play a role in causal selection, but nevertheless constrain these by its fundamental dependence on real causal patterns. For this reason, the resulting kinds do not collapse into constructionism. Hence, we are equipped to argue against the biomedical model on normative grounds without categorically rejecting the potential legitimacy of psychiatric medical treatment and research.

Disclaimers

I have chosen major depressive disorder as my main example because it is a well-researched, highly prevalent, and highly medicalised mental disorder which is empirically correlated with

both biological and non-biological factors. The condition is especially interesting because it highlights the context-sensitivity of psychiatric classification insofar as certain causes exclude such a diagnosis, suggesting that the symptoms of MDD exist on a continuum spanning from pathology to non-pathology. That is, the symptoms of MDD are considered an appropriate and healthy response to some events in a person's life such as bereavement. In contrast, it is not obvious that the characteristic symptoms of e.g., schizophrenia, such as hallucinations, are ever regarded as a similarly appropriate response to life events.

For the sake of the argument, I assume that the proposed symptoms of MDD found in the DSM in fact cluster sufficiently stably to constitute a valid kind. Several authors argue that this is not the case (Kutchins and Kirk 1997; Poland 2014; Hyman 2010). I do not intend to challenge these claims, nor do I intend to argue that all current DSM categories are valid kinds. My thesis is concerned with establishing an adequate model of mental disorders given the assumption that they instantiate relatively stable symptomatic property clusters. I will occasionally substitute the term "MDD" for "mental disorder(s)" since my model is equally applicable to both.

Throughout this thesis, I assume that mental and social causation is possible. I also assume indeterminism, and in extension the possibility of probabilistic causal explanations. As will be especially relevant chapter 3 and 4, I stress that I do not make any claims about the metaphysics of causation. I am only concerned with the epistemological role of causation for the purpose of psychiatric classification and explanation. Hence, I assume a neutral stance towards questions about the nature of causation itself as well as on the exact definition of a causal mechanism.

Chapter 1
Psychiatry, Essentialism, and Homeostatic Property Cluster Theory

1. Traditional Essentialism

1.1 Essentialist natural kinds and their role in science

Traditional essentialists hold that natural kinds must possess essences "[…] that define them in terms of necessary and sufficient, intrinsic, unchanging, ahistorical properties" (Boyd 1999, p. 146). Such essences play three distinct epistemic roles. Firstly, they are classificatory; they demarcate kinds from each other. Secondly, they are explanatory; they can adequately account for the typical traits of a kind. Thirdly, they are predictive; knowing that something belongs to a kind will allow me to infer that it possesses the typical properties associated with the kind, possesses the relevant essence of the kind, and will react in predictable ways, etc. Throughout this thesis, I will refer to the totality of these three epistemic roles as 'epistemic value'. I will refer specifically to predictive power as inductive potential.

Essentialist natural kinds are stable entities. For example, the chemical element 'gold' can be understood as a natural kind whose essence consists in having 79 protons. This essence is classificatory insofar as it is a necessary and sufficient condition for being gold. Whether some material is correctly classified as gold then depends on its essential internal structure, namely its number of protons. Something which lacks this number of protons might resemble gold in other respects and yet not fall under the same kind. The proposed essence of gold is also explanatory and predictive insofar as the typical properties possessed by gold – such as its melting point or appearance - can be explained and predicted with reference to its essence. Gold has a melting

point of 1064 degrees Celsius because it has 79 protons, and because having this number of protons disposes it to react in a particular way to certain temperatures. The essence causes the relevant properties. These qualities are assumed to be independent of social or cultural norms. Gold exists 'out there' in the world and our classificatory and explanatory practices are shaped to accord with its nature.

1.2. The commitments of Psychiatry

1.2.1 The current system of psychiatric classification

Classification is integral to psychiatric practice given the crucial role of diagnostic categories in diagnosing, researching, and treating mental disorders. The influential publication The Diagnostic and Statistical Manual of Mental Disorders (DSM), published by American Psychiatric Association, provides an authoritative classification system and reliable diagnostic criteria for psychiatric practitioners[1]. From the third edition, published in 1980, and onwards the publication has used an atheoretical and symptomatically descriptive (i.e. non-causal) method of classification (Tsou, 2016). This is also true for the most recent edition, the DSM-5, which is agnostic on the cause and nature of mental disorders. The manual's strategy for classifying disorders is instead atheoretical and employs descriptive polythetic classification criteria. Rather than dictating necessary and sufficient criteria, the polythetic classification only stipulates that a patient must, in order to qualify for a diagnosis, experience some sufficient number out of a larger set of associated symptoms. For example, a patient must during the same 2-week period

[1] Other classification systems of mental disorders exist, such as the International Classification of Diseases published by WHO (ICD-11). For simplicity, and due to prevalent similarities between the two systems, I have here chosen to focus exclusively on the DSM.

experience a minimum of 5 of the 9 symptoms which DSM-5 outlines as criteria for major depressive disorder [2].

By observing the symptoms of a patient, a psychiatric practitioner is then equipped to assign an appropriate diagnosis in accordance with the prescriptions of the manual. As shown, this does not entail any explicit commitment to the aetiology of the disorder, i.e., the causes which brought it about. Nevertheless, the DSM excludes the possibility of a diagnosis if the symptoms are deemed a "normal reaction" to some event or are caused by drugs or other physiological disorders (APA 2013). I shall expand further on the consequences of such exclusion criteria in chapter 2. For now, it suffices to say that the (relatively) acausal psychiatric system of classification contrasts against the disease entity model as presented in the introduction. A medical diagnosis such as 'tuberculosis' is aetiological since it entails that the patient's symptoms must be caused by the specific bacterium mycobacterium tuberculosis. According to the criteria outlined in the DSM-5, an MDD diagnosis entails no commitment to any such underlying cause, but merely describes the patient's symptoms.

1.2.2. The usefulness of natural kinds in psychiatry and society

In spite of the non-causal classification system outlined in the DSM-5, contemporary psychiatry is characterised by its adherence to the biomedical model presented in the introduction. By identifying mental disorders with neurochemical imbalances, the biomedical model treats mental disorders like essentialist natural kinds. By understanding mental disorders as natural kinds with a

[2] The nine criteria for depression in the DSM are as follows: (1) persistent low mood, (2) anhedonia (the inability to take pleasure in activities which were previously considered pleasurable), (3) weight loss or weight gain, (4) (continued) insomnia or hypersomnia, (5) psychomotor agitation or retardation, (6) fatigue, (7) feelings of worthlessness or excessive guilt, (8) difficulty thinking or making decisions, and (9) recurrent thoughts of death or suicidal ideation. At least one of the patient's symptoms must be low mood or loss of pleasure. (APA 2013, p. 160-161).

decisive neurochemical causal structure, we are equipped to treat them like other objects of scientific study: We can postulate a biochemical essence for a mental disorder such as depression (e.g. serotonin deficiency), and claim that this essence is directly responsible for, and therefore explain, the symptoms (loss of pleasure, guilt, etc.). We can then establish generalisations which can guide both scientific research and clinical treatment (e.g., that drugs which act as serotonin protagonists are likely to alleviate the symptoms of MDD). These generalisations then ground predictions pertaining to specific instances (this MDD patient is likely to benefit from SSRI medication). The biomedical conception of mental disorders further entails an ethical dimension which conceivably accounts for some of its popularity in the broad population insofar as it can contribute to reducing the moral responsibility imposed on patients for their condition by likening it to an involuntary somatic pathology like tuberculosis rather than a moral failing (see Anonymous, 2013). Further, a biomedical conception of mental disorders can aid in bureaucratic interventions for the benefits of patients, such as by making psychiatric patients eligible for insurance payments or other forms of private or government funded disability support (Rosenberg, 2006).

1.2.3. Short summary of the state of psychiatry

As shown, the commitments of psychiatry are somewhat paradoxical. On one hand, the authoritative classification system which is utilised in clinical practice is entirely descriptive and agnostic about the causes and nature of mental disorders (with the exception of the few exclusion criteria outlined in section 2.1.) On the other hand, psychiatric research is dominated by the biomedical model which treats mental disorders as if they were essentialist natural kinds whose essence consist in particular neurobiological imbalances. This in turn reflects on psychiatric practice insofar as neuro-pharmaceutical medication is unambiguously the first-line treatment of MDD (Koenig and Thase, 2009). In the following section, I will briefly show why the essentialist natural kind model is inadequate to account for the nature of MDD.

1.3. The Philosophical Inadequacy of Essentialist Natural Kinds in Psychiatry

The majority of philosophers agree that psychiatric categories do not pick out essentialist natural kinds (Craver 2009; Haslam 2000, 2002; Zachar 2000) and thereby reject the assumptions underlying the biomedical model of psychiatry. Importantly, not every instance of the disorder instantiates fixed symptomatic properties (Olbert & Gala, 2015). As per the DSM criteria, MDD is characterised by symptomatic heterogeneity insofar as its symptomatic expressions can and do vary between patients. For this reason alone, MDD is an unsuitable subject for essentialist analysis.

MDD is also characterised by causal heterogeneity. No empirical evidence for a causal essence of MDD has yet been found (Slothouber, 2019). As for the popular monoamine hypothesis, the relationship between monoamine neurotransmitters and MDD is far from sufficiently stable to warrant an essentialist position. Studies have shown that decreased monoamine levels are neither necessary (Meyer et al., 2004; Parsey et al., 2006) nor sufficient (Reivich et al. 2004; Kring et al. 2007; Delgado, 2011) for MDD. Unless a particular type of neurochemical imbalance is instantiated by all and only patients who suffer from MDD, then such an imbalance cannot be postulated as an essence in the strict and traditional sense specified above. Even though abnormal levels of monoamine neurotransmitters appear to be in some way associated with MDD, there simply is no one invariable biological causal structure which is instantiated by every case of the disorder.

In the next section, I will present Richard Boyd's post-essentialist homeostatic property cluster theory of natural kinds which is meant to serve as an alternative to the unsuccessful traditional essentialist conception described in this section. I will then go on to show how the HPC theory can plausibly account for the case of MDD.

1.4. Richard Boyd's post-essentialist Homeostatic Property Clusters

Richard Boyd's Homeostatic Property Cluster theory of natural kinds proposes an alternative to the essentialist conception of natural kinds outlined in the previous section. Boyd specifies the key characteristics of HPC kinds as follows:

"(1) There is a family (F) of properties that are contingently clustered in nature.
(2) Their co-occurrence is the result of what may be described as homeostasis: either the presence of some properties tends to favour the presence of others, or there are underlying mechanisms that tend to maintain the properties in F, or both.
(3) There is a kind term (t) that is applied to things in which the homeostatic clustering of most of the properties in F occurs" (Boyd, 1999, p. 143-144)

The HPC theory defines natural kinds not in virtue of static essences leading to almost as 'fixed' static properties, but rather conceives of such kinds as homeostatic property clusters which are generated and sustained by underlying causal mechanisms. Some variation of typical properties between particular instances of the same natural kind can occur. Still, HPC kinds do display an overarching stability which is explained by the fact that the property clusters are caused by the similar underlying causal mechanisms. Hence, HPCs can be regarded as having essences, but in a much looser sense than that proposed by the traditional essentialists. Nevertheless, HPCs retain much of the epistemic values associated with natural kinds. They maintain their classificatory, explanatory, and predictive power while still being able to accommodate naturally occurring variation.

The property clusters are not strictly delimited necessary and sufficient sets of properties, nor must they necessarily be caused by one single mechanism. Instead, HPCs have vague boundaries since the theory is flexible with respect to both the clustered properties and the relevant causal mechanisms. One example of a natural kind that is affected by several causal mechanisms is

blood pressure. A stable blood pressure can be conceived of as the 'typical property' which is produced by multiple cardiovascular, neuromuscular and hormonal underlying mechanisms. A change in one of these mechanisms (e.g. an increase in the stress-hormone cortisol) will then change the blood pressure (e.g. increase it). However, neither of these mechanisms alone are necessary and sufficient for maintaining a stable blood pressure. Changes in the blood pressure can be explained with reference to the underlying causal mechanisms underlying them (e.g., a medical practitioner can diagnose a patient with chronic stress partly on the background of these observations).

Further, the typical properties and the causal mechanisms need not be physical. Social properties and mechanisms are equally plausible candidates for this role. Insofar as a reliable causal relationship can be established between some social or relational mechanisms and some set of properties, nothing prevents this relationship from being conceived of as an HPC. However, this also means that the 'naturalness' of Boyd's natural kinds take on a radically different form than how natural kinds were conceived on the essentialist understanding. HPCs need not be found in nature, entirely independent from human practice and conceptualisation. They just need to be found. Insofar as an appropriate causal relationship exists between some mechanisms and some set of (variable) properties, this relationship is sufficient to constitute an HPC. Hence, HPCs can be highly social and are not "natural" kinds in the conservative meaning of the term which understands such kinds as existing independently of human conceptualisation.

In the following section I will show how such this view can be used to establish a post-essentialist conception of mental disorders generally and MDD specifically.

1.5. A post-essentialist conception of MDD

1.5.1. What are the causes of MDD?

What, then, are the causes[3] of major depressive disorder? Compared to many somatic diseases such as tuberculosis, the causal trajectories of mental disorders are much less clear. In fact, empirical evidence suggests that the causal structure of MDD is characterised by significant heterogeneity and complexity at multiple levels of analysis which include psychological and social causal pathways as well as biological ones (Poland et al., 1994). Such causes include, but are not limited to, genetics, neurochemistry, brain circuitry (i.e., abnormal activity at the level of brain centres), psychology (personality, negative biases) and social and environmental factors (stress, adverse life experiences) (Muang, 2016).

Unlike in the case of bacterial infections, none of the above are invariably associated with MDD. Instead, they are better understood as probabilistically increasing a person's risk of having or developing the disorder. Going forward, whenever I refer to the causes of MDD, I will be referring to such probabilistic risk factors unless otherwise specified.

1.5.2. MDD as an HPC

Kendler et al. (2011) adapt Boyd's theory to the realm of psychiatry and argue that mental disorders are best understood as mechanistic property clusters. The fundamental structure of MPCs is similar to that of Boyd's HPCs, but the authors change the name from 'homeostatic' to 'mechanical' property clusters since they believe that the former descriptor entails misleading physiological associations (I shall continue to use the term HPC when not referring directly to this article). The authors want to emphasise the importance of the variable causal mechanisms that produce, underlie, and sustain mental disorders. They posit the relevant underlying

[3] As mentioned in the introduction, I do not mean to propose any claims about the metaphysics of causation. My discussion aims at neutrality pertaining to the different philosophical positions concerning this question.

mechanisms as being multilevel - e.g., molecular, developmental, neurobiological, psychological, and social – and responsible for producing the symptoms of the mental disorder in question. They write that: "The MPC view encourages the thought that there are robust explanatory structures to be discovered underlying most psychiatric disorders. However, it cautions us to expect that they will be messy and that it will take hard work and some degree of idealisation and abstraction to bring them into focus." (2011, p. 4). Nevertheless, this account captures the epistemic value of psychiatric categories: we can explain depressive symptoms with reference to their causes, and we can classify them with reference to both the causes and the symptom-clusters. Yet it does not commit its authors to a solely neurobiological view of the nature of mental disorders. Given the multilevel nature of the causal mechanisms which underlie the symptoms, social and psychological causes are potentially equally relevant as neurobiological causes. This is further compatible with the idea that symptoms can mutually enhance and cause each other. As Cramer et al. put it: "fatigue may lead to a lack of concentration, which may lead to thoughts of inferiority and worry, which may in turn lead to sleepless nights, thereby reinforcing fatigue" (Cramer et al., 2010 p. 140–141)

We see how this mirrors the polythetic DSM-5 classification system where MDD is defined as a cluster of unfixed but stably co-occurring symptoms such as depressive mood, sleep disturbances, anhedonia, and fatigue. The HPC theory adds the commitment to some number of underlying causes which explain the stability of the symptoms. The DSM differs from the HPC insofar as it for pragmatic purposes imposes categorical boundaries on the diagnostic categories by enforcing a threshold specifying an exact number of symptoms which must be present before a patient can receive a diagnosis. The HPC itself is likely to have much vaguer boundaries.
The HPC view of MDD has a range of advantages. Firstly, it maintains the epistemic values associated with traditional natural kinds. Secondly, unlike the essentialist view, the HPC theory can accommodate the symptomatic and causal heterogeneity characteristic of MDD. Thirdly, the

HPC view can account for the vague boundaries characteristic of MDD. And importantly, the theory can prioritise non-biological causes on an equal level with biological causes and hence has the potential to challenge the neuro-centric assumptions of the biomedical model.

1.6. Summary of chapter 1

In this chapter, I have explained the notion of essentialist natural kinds and shown their relevance in psychiatry. I have explained why the epistemic value associated with essentialist natural kinds makes it attractive to consider mental disorders such natural kinds. Namely, because it allows for psychiatry to establish explanations, predictions, and generalisations pertaining to the relevant disorders.

I have argued that an essentialist conception of MDD is not viable. I have instead proposed a conception of MDD as a homeostatic property cluster kind where the known risk factors, biological and non-biological, function as the causal homeostatic mechanisms. In the following chapter, I will question whether the appeal to causation is in fact necessary for a successful model of MDD.

Chapter 2
Is Appeal to Causation Superfluous?

In this chapter, I will first show that DSM categories, despite being defined by their symptoms, do contain some causal information from which the explanatory value of the MDD category derives. This causal information is respectively negative and probabilistic.

I will then assess two non-causal, pragmatic accounts of kinds, namely Peter Zachar's practical kind theory and Matthew Slater's stable property cluster theory. I will argue that both theories hinge on implicit causal assumptions and cannot adequately account for MDD. I conclude that some appeal to causation is necessary for defining mental disorder categories.

2.1. The explanatory value of DSM categories

2.1.1 Mental disorder categories and explanatory circularity

Recall that the DSM defines MDD in virtue of its associated symptoms rather than its causal background. But if major depressive disorder is defined by some number of specific symptoms, then 100% of all cases of major depressive disorder will necessarily exhibit these symptoms – simply because these instances would not in the first place be classified as cases of MDD if they did not! By defining a disorder in virtue of certain symptoms which constitute it (rather than in virtue of e.g., an underlying bacterial infection), we severely compromise its explanatory value. 'Having MDD' thus is no more explanatory as to why somebody is fatigued than someone's being a bachelor is explanatory of his being unmarried. It seems then that the diagnostic category can never explain the symptoms but only describe them.

2.1.2. Implicit causation

However, while the DSM-5 explicitly classifies disorders via non-causal criteria, it nevertheless makes an appeal to aetiology in its exclusion criteria. The DSM-5 states that a patient is only a candidate for a diagnosis on the condition that: "the [symptoms are] not attributable to the physiological effects of a substance or to another medical condition" (p. 161). The manual further states (albeit discretely placed in the introduction) that "an expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder" (p. 20). While these causal factors are negative, they still entail the conceptual relevance of aetiology for psychiatric classification.

Previous editions of the DSM included a 'bereavement exclusion' stating that, apart from in especially severe cases, being recently bereaved would exclude a patient from qualifying for an MDD diagnosis. The DSM-5 controversially removed the bereavement exclusion and hence broadened the scope of the diagnosis (Iglewicz et al, 2013). As per the quote stating that the symptoms must not be due to a loss, I shall throughout this thesis assume that recently bereaved and grieving people are, ceteris paribus, not suitable candidates for an MDD diagnosis.

This points to an equally clinically useful and intuitively attractive distinction between pathological and non-pathological responses by appealing to *what they are responses to.* Are they largely considered appropriate responses or not? In this sense, MDD appears to be an aetiological kind, meaning that it is a kind which is partly defined in virtue of its causal background (Khalidi 2021). It is further deeply relational since the aetiology is judged relative to cultural expectations. However, the aetiological demands on psychiatric categories remain minimal insofar as no specific causal trajectory is considered essential for MDD. The DSM criteria employ negative causal reasoning only insofar as some causes exclude a patient who exhibits the relevant symptoms from obtaining the diagnosis. No one aetiology can positively *qualify* someone for a diagnosis. Thus, whether someone suffers from MDD or not is still

predominantly decided on the background of symptoms.

Further, recall that I listed a number of empirically validated biological, psychological, and social risk factors of MDD in chapter 1. While the DSM does not refer to these factors in the definition of MDD, I assume that the diagnosis can be explanatory with reference to them since scientific research continues to aim at uncovering causal dependencies between MDD and associated risk factors. If these risk factors were rejected or ignored in psychiatry, then research would be pointless and treatment – especially medical treatment - would be entirely unjustified. Still, these risk factors only provide probabilistic causal information and hence fall short of the solid explanatory value associated with essentialist natural kinds (Muang, 2016).

Thus, while psychiatric categories are formally defined in virtue of their symptoms, they maintain a twofold explanatory value. Firstly, given the negative causal reasoning implicit in the formal definitions, by knowing a diagnosis we know a few causes which have *not* caused the symptoms. They can therefore be understood as partly aetiological kinds. Secondly, we can by appealing to the information derived from empirical studies obtain a probabilistic explanation of how these symptoms might in fact have been caused.

In what follows, I will examine whether appealing to such causal factors is at even necessary for a successful theory of the nature of MDD.

2.2. A Pragmatic Approach

2.2.1 Mental disorders as practical kinds

Peter Zachar (2000) argues in favour of a conception of mental disorders as 'practical kinds'. Such kinds are not justified with reference to any causal foundation, but simply in virtue of being stable patterns which have practical utility. These categories are merely tools we can use to grasp and manipulate the world, but they do not claim to reflect its actual structure. Practical kinds are

flexible relative to our purposes and unconstrained by any demand to 'fit' the world. They nevertheless remain confined by their *real* practical utility which is what prevents them from being arbitrary. A genuine practical kind is Body Mass Index or BMI (Haslam, 2002). While weight is distributed on a continuum from healthy to unhealthy, the BMI (height/weight^2) has a cut-off point of 25 for being overweight. The BMI serves the purpose of demarcating what members of the population are particularly vulnerable to certain health risk associated with obesity. Under other circumstances, such as if the majority of the populations were super-muscular bodybuilders, the BMI would have to be revised given that it is currently unequipped to distinguish between 50 kgs of fat or 50 kgs of muscle. Still, under our current circumstances it is a useful tool and therefore a genuine practical kind.

According to Zachar, psychiatric categories function in an analogous way. On this view, human suffering is conceived of as a continuum, upon which pathological/non-pathological distinctions are imposed simply for the purpose of treatment. Such a distinction does not hinge on any genuine underlying dichotomy but is nevertheless still constrained by pragmatic considerations such as successfully identifying those patients who require treatment. What is practical about our current mental-disorder categories? They guide bureaucratic purposes, pick out people who need treatment, guide research, etc. According to Zachar, the legitimacy of such kinds ought to be judged by how useful they are for these purposes, not whether they instantiate an essence or reflect real divisions or causal structures. In this sense, his position is both anti-essentialist (categories are not meant to identify necessary and sufficient conditions for membership) and anti-realist (categories are only tools which facilitate our understanding and manipulation of the world – they do not describe how it really is). The main difference between this view and the HPC view is that practical kinds lack any real boundaries while HPCs have real but indistinct boundaries.

2.2.2. An Objection: Where does the practicality come from?

But what grounds the practicality of practical kinds?

There seems to be two possible answers to this question. Firstly, the practicality of such kinds might be grounded in real boundaries in nature - what one might call "objective facts". Perhaps it really is the case that those with a BMI over 25 are statistically more likely to suffer from certain diseases due to their weight. This again might be explained by reference to how this number is calculated (body weight relative to height) and how an excessive amount of body fat effects the human body. That is, whether a certain percentage of body fat is likely to cause certain health problems. While the cut-off point of 25 is somewhat arbitrary, the utility it has seems derive from the actual constellation of the world. In this case, practical kinds cannot maintain their anti-realism. They stop being mere tools and become instead reflections of real causal processes and "how things really are".

In order to avoid this, Zachar can also claim that practical kinds are still useful in the absence of any objective facts. But this position seems to collapse his view into pure constructivism. What is usefulness if not something which hinges on objective facts (albeit relative to a purpose)? On this account, it seems that one might as well classify overweight as a BMI of 18 as well as 28. Further, such a commitment leaves Zachar no way to explain how models can be pragmatically improved or how they can guide future research. It appears we need to anchor practicality in reality to establish a convincing theory.

2.3. A New Strategy: Stable Property Clusters

2.3.1. Slater's SPCs

Several philosophers have defended simple cluster theories of natural kinds (Chakravartty; 2007; Häggqvist, 2005; Slater, 2015). On this view the epistemic value of "natural kind" terms does need to be justified, but not by appeal to neither causation nor pure pragmatism.

Matthew Slater (2015) proposes a 'stable property cluster' theory which defines natural kinds solely in virtue of the non-causal grounds of their epistemic value. Slater claims that stable property cluster kinds (call these SPCs) are defined in virtue of their counterfactually stable inductive potential. All it requires for something to be a SPC is that some properties cluster together in a relatively stable way which remains stable even under relevant counterfactual circumstances. This counterfactual stability itself – and not any underlying cause of it – is the reason for the kind's inductive and explanatory power (I shall assess this claim in the next section). The SPC theory would then conceptualise MDD in a manner resembling the official definition in the DSM-5. Namely, as a cluster of (relatively) stably co-occurring symptoms. The demand for relevant counterfactual dependence would then conceivably play the role of exclusion criteria as not to include, e.g., normal grieving reactions. On one hand, the theory is not as rigid as traditional essentialism since it allows that symptomatic clusters can be more or less stable. On the other hand, it is less radically unconstrained than Zachar's practical kinds since it demands that the properties must *really* cluster. SPCs are clearly well-suited for the purpose of prediction. If we know that some set of properties cluster together in a stable way, we can predict by observing some number of the properties that the others are likely to be present. But what about their other epistemic qualities?

2.3.2. Problems with the SPC theory

Ideally, epistemic value entails not just prediction, but also explanatory and classificatory power. Can SPCs provide this as well? Slater seems to think so when he writes that: "[…] an account of natural kinds would do better to focus on the special sort of stability a cluster of properties might possess in virtue of which it is apt for induction and explanation rather than focusing on the something causing that stability" (2015, p. 396)

Slater appears to assume a slender understanding of explanation which relies on induction on the basis of a stable property cluster. Knowing that an animal has trotters, says mooh, has horns, and

produces milk, will allow me to infer that it also eats grass insofar as these properties cluster stably in the kind cow. In this sense, the animal's being a cow allegedly explains why it eats grass, why it has horns, etc. – because it is a cow, and that is what cows tend to do!

Even if we accept that this can constitute an explanation, it is clearly rather impoverished and non-informative relative to other causal explanations which pertain to evolution, natural selection, physiology, etc. For the purposes of this thesis, it suffices to say that the SPC theory is *especially* unsuited to account for mental disorders which are formally defined in virtue of their symptoms. Given Slater's explicit rejection of causal reasoning, he will face the explanatory circularity outlined in section 1.1. of this chapter when trying make the SPC account explanatory for the case of such mental disorders.

As for classification, take two people, Stephen and Lucy, who display the same collection of depressive symptoms. However, Stephen is recently bereaved, whereas Lucy has suffered no such losses and has a fulfilling life. As per the bereavement exclusion – or the common-sense response that grief is an appropriate reaction to loss – it seems that it is justified to diagnose Lucy with MDD but not Stephen.

The SPC theory cannot discriminate on the basis of causation, and hence seems unable to distinguish between these two cases. Analogously, when diagnosing a physical disease, it seems that reference to the underlying causal properties of the clustering symptoms is crucial for correct classification – it is very important to know whether the patient's cough and fatigue is caused by tuberculosis or by having sung along at a concert all night!

However, Slater writes that the relevant property clusters must remain stable under relevant counterfactual circumstances. To counter the above criticisms, he will likely argue that the relevant counterfactual circumstances can aid in appropriately demarcating kinds. What counts as relevant counterfactual circumstances is a matter of judgement. For example, say that in

counterfactual circumstances where Lucy had not been recently bereaved, she would not exhibit depressive symptoms. Presumably, this will exclude her from an MDD diagnosis. Similarly, one might propose a particular genetic mutation as being relevant for Huntington's disease, but not for Alzheimer's even though both disorders have the same symptoms. In that case, in a counterfactual world where the patient lacked this genetic mutation, then his symptoms would not cluster stably. Therefore, he has Huntington's and not Alzheimer's.

Other counterfactual circumstances might instead affirm the kind-status. For example, it might be the case that if Stephen's natural serotonin levels were higher, then he would not suffer depressive symptoms. This appears to affirm that Stephen does indeed suffer from MDD. Different scientific disciplines will have different judgements pertaining to what counterfactual circumstances are relevant for an SPC kind. That is, what qualifies as a relevant counterfactual condition depends in part on who is asking the question. For the evolutionary psychologist investigating MDD, it might be a relevant counterfactual that Lucy would not instantiate depressive symptoms if her prehistoric ancestors had followed a different evolutionary trajectory. But this is irrelevant for the clinical psychiatrist who is more concerned with counterfactuals pertaining to bereavement or serotonin levels.

Given that the aim of the SPC theory is to avoid reference to underlying causation, I remain unconvinced by these claims. Slater's appeal to counterfactual stability is an especially implausible move if he wishes to bypass causal reasoning given that a number of philosophers argue that causal relations simply amount to counterfactual dependence[4.] Slater himself makes no explicit assumptions about the nature of causation and neither will I be able to provide more than the epistemological considerations directly relevant for the topic of this thesis. Yet it can safely be assumed that counterfactual dependence is intimately connected to causation.

[4] See Woodward 2003; Lewis 1979, 2004; Hitchcock 2001, 2007

Why, then, is the fact that the property cluster disperses in counterfactual situations where the patient has higher serotonin levels relevant for the clinical psychiatrist? It appears to me that such judgements hinge on tacit causal assumptions. I believe that the connection between counterfactual dependence and causation gives rise to the different judgments about relevant and irrelevant counterfactual situations. Some circumstances are relevant for demarcating kind-membership because they suggest that the depressive symptoms are not connected to their cause in the 'right' – i.e. the MDD-typical - way *and therefore* unsuitable for falling under the kind MDD. Slater might object by insisting that the relevant counterfactual scenarios are simply those which influence the epistemic value of the SPC relative to the interests of the scientific domain conducting the inquiry. That Stephen and Lucy would not be depressed if their ancestors had been different does not reveal a great deal about their current depressive states to a practicing psychiatrist. *That* – and not any causal relation between the elements featuring in counterfactual scenarios – is what justifies this scenario's irrelevance. In this sense, SPCs remain constrained by the structure of the world, namely the structure that results in the kind's inductive potential. But this argument is viciously circular: a kind is natural only insofar as it has inductive potential. It has inductive potential if it is a relevantly counterfactually stable cluster. What counts as relevant counterfactual stability is that which results in the kind's inductive potential. Full circle.

The problem is that Slater, in the absence of causal reasoning, fails to provide adequate justification for the counterfactual criteria according to which these categories are distinguished. It is clear how this issue mirrors the problem we encountered when examining Zachar's practical kinds. While practical kinds appear to be practical only insofar as they mirror the causal structure of the world, the inductive potential of SPCs is equally likely to be parasitic on certain causal assumptions that support which counterfactual conditions we identify as relevant or irrelevant. Conclusively, either the judgements we make about relevance of counterfactuals simply reflect random inclinations (given that there in theory is no limitations on what we might find relevant),

or we must provide further justification in favour of them. If we justify these judgements in terms of the resulting inductive potential, then our argument is circular. If, like Zachar, we justify them in terms of practicality, then we are back where we started.

I think that Slater has good reason to emphasise the role of normative influences in scientific classification and general practice by introducing the notion of *relevance* of counterfactual criteria relative to purposes of the inquirer. I sympathise with his desire to legitimise explanatory pluralism in this way. Nevertheless, I think he is mistaken in refusing to acknowledge the conceptual importance of causation for the epistemic value of kinds. Without reference to such causation, his theory is ultimately unconvincing. I will suggest how the HPC model can reconcile causal and normative influences in chapter 4.

2.4. Summary of Chapter 2

In this chapter, I have argued that DSM categories do contain some causal information, namely negative and probabilistic causal information. I have further argued that while psychiatric classification is largely non-causal compared to other areas of medical science, some appeal to causation remains necessary. Considering mental disorders merely practical kinds is unjustified insofar as nothing can explain the practicality of these kinds except by appealing to the causal structure of the world. Relying instead on the inductive potential of counterfactual stable property clusters is similarly inadequate insofar as such SPCs rely on judgements about relevant counterfactuals which are likely to be either parasitic on causal assumptions or simply unjustified. However, appealing to causation admittedly entails a more complicated picture than simpler models like those examined in this chapter. In the following chapter, I will examine what problems the appeal to causation poses for the HPC model.

Chapter 3
The Problems of Causal Classification

In this chapter, I tackle Peter Zachar's objection that respectively causal heterogeneity and multifactorial causation are a problem for the HPC model of mental disorders. I argue that the HPC account can accommodate some amount of causal heterogeneity. I then respond to the problem of picking out primary causal factors given multifactorial causation. By analysing a simple contrast case, I will advance a distinction between proximal and distal causes. I argue that proximal causes are comparatively more closely connected to the relevant effect and are therefore more likely to ensure stability in an HPC and thus prima facie better suited for being picked out as primary causal factors. I will then show how this understanding can – albeit with less certainty - be transferred to the case of MDD where neurobiology can be identified as a proximal cause. I claim that this partly justifies, and perhaps has motivated, the prevalent focus on neurobiology in psychiatry.

3.1. How do we Identify Primary Causal Factors?

3.1.1. Causes in MDD

To remind the reader what kind of causation is at play in HPC kinds as opposed to in essentialist kinds: In essentialist discourse about kinds, the proposed intrinsic essence of a kind directly and invariably *causes* and hence explains its typical properties in a rigid manner. Where the essence is present, the properties will – ceteris paribus - follow.

In HPC discourse, a cluster of typical properties is caused by one or several homeostatic mechanisms which maintain these in a stable but variable way. On this view, the relevant kinds need not possess the exact same causal foundation or typical properties but can have various degrees of causally supported similarities to each other. They possess largely similar properties

due to being supported by largely similar causal mechanisms.

MDD does not have any invariable causes but is instead associated with a range of risk factors.

3.1.2. Causal Heterogeneity and Multifactorial Causation

Peter Zachar (2000) identifies the problem of what he calls "causal overdetermination" for psychiatric kinds which are defined (partly) in virtue of their causal mechanisms. Zachar does not expound further on this objection but given his comments on the same topic in a later publication (Zachar, 2002), I take him *not* to refer to the argument of causal overdetermination which is often raised in the literature as an objection against counterfactual theories of causation and pertains to the difficulty in determining priority between two simultaneously occurring and independently sufficient causes (see e.g., O'Connor, 1976).

Rather, it seems Zachar is concerned with problems pertaining either to causal heterogeneity or multifactorial causation. In what follows, I will develop and respond to these two objections.

1. Causal heterogeneity

The causal trajectories behind various patients' MDD are quite heterogenous. If the disorder is partly classified in virtue of its causal trajectory, how can it accommodate such variation?

On my reading, Boyd's theory leaves room for some variability on both the level of causation and properties. The causal mechanisms underlying individual instances of MDD need not be identical. Even if one case of MDD has a somewhat different causal trajectory than another otherwise identical case, this does not mean that we must reject one of these cases as an instance of MDD as long as they both fall under the vague cluster of causal trajectories associated with the kind. It is consistent with Boyd's theory to insist that similarity in properties trumps similarity in causal history in the classification of disorders – even if both are relevant. Boyd writes, "I do not for better or worse, hold that HPC kinds are defined by historical relations rather than

shared properties" (1999b, p. 80). Hence, when a conflict arises between classifying in virtue of similarity versus on the background of causal mechanisms, we can choose to classify by similarity. Given the importance of symptomatic expressions for mental disorders, I think it is justified to prioritise these in our classificatory purposes pertaining to them. Admittedly, this priority is not given by the HPC itself, but by pragmatic decisions about how best to interpret the HPC. I defend the legitimacy of pragmatic causal selection in chapter 4.

2. Multifactorial Causation

It is a broadly accepted view among philosophers that most events are not monocausal, but in fact have an enormous number of causes (Lewis, 1973, 1986; Beebee, 2004). Each of these causes are themselves caused by something else. Considering this, classifying disorders with reference to their causal mechanisms seems especially difficult. The multitude of causes (simultaneous and extending in time) appear to obscure any explanatory and predictive value the theory might have had. If we accept *all* the causes of MDD as equals, we encounter two problems.

The first pertains to classification: The conceptual coherence of HPC theory depends on being able to distinguish some limited number of causes as the homeostatic mechanisms underlying the HPC kind. If *all* causal factors are treated as equal, causation will likely be irrelevant for classification due to the extent to which the totality of causal trajectories shared between people resemble each other (in a sense, we are all caused by the big bang, having been born, etc.). Hence the HPC theory will be unable to properly demarcate kinds.

The second problem pertains to explanation: Too many causes hinder rather than aid the explanatory aim. While it is true that having a heartbeat is a causal factor of MDD (insofar as dead people are not afflicted by this condition), it is not a good explanation of the aetiology of MDD.

Evidently, this is a problem that afflicts *all* causal statements and explanations, not merely those underlying the HPC theory. I shall attempt to tackle it nevertheless as I think the discussion might illuminate some motivations behind the preference for biomedical explanations in psychiatric research.

Clearly, we normally resolve this difficulty by resorting to causal selection. Namely, we pick out some causes as primary and bracket the others. Often this is relatively simple, especially when the causal process is spatiotemporally extended and visible: think of a ball smashing a window. It is intuitively obvious that it was *the ball* that broke the window. But the heterogeneity of risk factors in MDD means that there is a large and non-transparent range of different ways in which an individual can acquire the disorder. Given the multiplicity and opacity which characterises the causal trajectories of MDD, how can we justifiably identify any primary causal factors for this disorder?

In the following section, I will analyse a contrast case in order to show one way in which we can conceivably select some primary causes.

3.2. Introducing the proximal-distal distinction

3.2.1 A contrast case: the broken window

Take the following scenario:

You own a basketball hall. One of your players, Peter, throws the basketball and accidentally breaks a window. What is the cause of the window breaking? Is it:

(1) The ball hitting the window?

(2) The fragile molecular structure of the glass from which the window was made?

(3) The hardness of the ball?

(4) The force with which the ball was thrown?

(5) The placement of the window?

(6) The clumsiness of Peter?

(7) The nervousness of Peter which caused his clumsiness?

(8) The fact that Peter was born?

I have chosen this example because window being broken by a ball is a paradigmatic case of a causal process. It seems strange to claim that our classifications in this respect are *merely* practical constructs influenced only by pragmatic concerns – something really seems to happen out there in the world which we attempt to describe. The relevant causal factors and consequences, (a ball thrown, a ball hitting the window, the window breaking, the window being broken) are further *relatively* monocausal insofar as we would seldom hesitate to point out the ball as the primary cause. Nevertheless, as shown above, the apparent primary cause becomes a set of causes if looked at more closely. In fact, there is a multitude of factors contributing to the breaking of the window[5]. The precise set of causes can differ somewhat. The ball can be thrown not by Peter but by James, a different ball can be thrown, or a different window is broken. Hence, while I take as my main example the one particular scenario outlined above, the event can be understood as an instance of a repeatable (but not invariable!) causal process which can take place in basketball halls all over the world. Hence, the cause-effect relationship can vary in individual instances (*this* window broken by *this* ball under *these* circumstances in *this* basketball hall) but these individual instances still all fall under a general kind (windows broken by balls in basketball halls).
Given that there is a large multitude of causal factors interacting to produce the effect, how do we in this case choose which of these factors are primary?

[5] The process is more multi-factorial than I have made explicit given that many more factors interact than I have specified: E.g., the friction as the ball travels through the air, Peter's parents getting together back in 1970, etc.)

3.2.2. Proximal and Distal Causation

In identifying a primary causal factor in the above scenario, we might benefit from an appeal to the distinction between proximal and distal causes[6].

A proximal cause is a cause which is proximal to the relevant effect. In this case, (1) – (4) would likely be considered the proximal causes of the window breaking.

Similarly, a distal cause is a cause which is relatively more distal to the effect, and which leads to the proximal causes. In this case, if (3) the hardness of the basketball had been exaggerated by filling it with lead due to a mistake made during its production in the factory, then this mistake in manufacturing could be cited as a distal cause of the window breaking. Similarly, insofar as (5) – (8) are factors which lead to the relevant interactions between (1) – (4), these might be cited as distal causes. Further, there seems to be a hierarchy between distal causes which allows (8) to be singled out as being the *most* distal among these.

One characteristic feature of proximal causes is their relatively greater adherence to lawlike generalisations[8]. For example, it can be predicted that most, if not all, molecular structures of this particular fragility will break when subjected to such-and-such an impact within the relevant set of background conditions. The distal causes of the window breaking (Peter's nervousness and clumsiness, the placement of the window, etc.) are less likely to feature in such lawlike generalisations though they might partake in some generalisations pertaining to probability (nervous players are most likely to miss their shots, windows located just behind the hoop in a basketball hall are more likely to be broken). Hence, insofar as we knew that the window had a certain constitution and was subjected to a certain force within a particular set of circumstances, we could predict that it would be broken. So, the proximal causal elements in the broken-window case have stable *inductive potential* which help predict the outcome of the event with

[6] I use these terms in accordance with the loose definition specified in this section irrespectively of how they might be used by other authors elsewhere in the literature.

[8] As previously stated, I assume indeterminism throughout this thesis

significant certainty, and consequently is likely to feature in an explanation of the effect.

It appears, then, that epistemic value is correlated firstly with proximal causes, and secondly with distal causes. What explains the superior epistemic value of proximal causes?

The potential causal trajectory of an event can be illustrated as an upside-down tree which branches at every point of variation. For the window to break it *must* be the case that the molecules are caused to act in such-and-such a way. They must be brought to act in this manner, and that can happen in a variety of ways.

First and foremost, they must be impacted by a sufficient force[9]. But this impact can be imposed by a broad variety of objects – say, a hammer, a ball, a stone, etc, as well as a corresponding variety of velocities since a harder object will not need to be thrown as forcefully to obtain the same effect.

If we take another step back in the causal chain, the object can be thrown by many different people, or might even be propelled forward by an explosion, a machine, or similar purely physical events.

If we take one step further back the causal trajectory and asked why *that* happened we will be able to cite an even broader variety of potential reasons: someone might have acted in a particular way for a particular reason, the machine might be programmed in such-and-such a way, the explosion might have been the result of a bomb placed in the hall in a terrorist attack, etc. The more distal the causes, the more branching we find. Proximal causes tend to be more invariable, homogenous, and predicable.

If there *were* a situation in nature where such variation was not possible, then the distal causes would be as tightly connected to the inductive potential as the proximal causes.

[9] I use the example of physical impact although high temperatures can result in the same effect

For example, imagine a possible world in which deer and tigers are built exactly as the following states:

The deer runs iff the tiger is hungry

The tiger is hungry iff it has not been fed at noon

Then, the tiger not being fed at noon is entirely predictive of the deer running – it is no less predictive than the tiger being hungry.

However, it is difficult to think of such cases occurring in real life where only one thing alone can cause some event. Usually, events can be caused by a complex network of causal factors. Each factor (from proximal to distal) is multicausal, and for that reason it is the case that the more distal you go, the more potential variation is possible, and the less reliable your predictions will generally be.

Still, some distal causes are more likely bring about the proximal causes (and hence the effect) than others. For example, say that uncontrolled growth of cells and inadequate immune response towards this growth are the proximal causes of cancerous tumours, and that respectively smoking and asbestos exposure are distal causes of such tumours. It is estimated that 90% of lung cancer in men is caused by smoking whereas only 10% is caused by other factors such as exposure to asbestos (Walser et al 2008). Hence, by knowing that some man is diagnosed with lung cancer, we can predict that he most likely has a history of smoking. Therefore, it is worth noting that under certain circumstances some distal causes can still be highly predictive of the effect, and that knowing the effect can in some cases give us good reason to assume the influence of such a distal cause. Still, this is contingent on circumstance – in societies where tobacco smoking is not practiced, the correlation does not hold. Smoking remains a *potential* distal cause (insofar as if someone were to smoke, this would be likely to cause cancer), but the probability that a case of lung cancer is actually caused by smoking within in a tobacco-less

society is negligible. In contrast, all cancerous tumours are caused by uncontrolled cell growth (and inadequate immune response).

3.2.3. Short summary:

So far in this chapter, I have shown that causal factors are usually not singular but produce their effects in interaction with a range of other causally relevant factors. I have argued that the distinction between proximal and distal causation can nevertheless help in establishing a non-pragmatic causal hierarchy which aids in identifying primary causal factors, namely those which are most likely involved in maintaining stability to the clustering of the relevant properties in an HPC. Given that distal causes are relatively variable, heterogenous, and non-predictive, this might give us reason to consider the proximal causes primary in producing the effect relatively to distal causes. However, distal causes do possess epistemic value corresponding to the probabilistic extent to which they are connected to the relevant proximal causes. Some distal causes will be quite explanatorily irrelevant if they exist on a low level in the causal hierarchy and do not exhibit any significant probabilistic correspondence to the relevant effect.

These features suggest that proximal causes are especially apt to sustain the homeostasis in a homeostatic property cluster. This suggests that such proximal causes are the most suitable candidates for the role of primary causes. Proximal causes are non-pragmatic insofar as they are not prioritised because they serve our purposes, but because they *really are* more closely connected to the effect in the manner described above. In what follows, I will examine in which way the proximal/distal distinction might help selecting the primary causes for MDD.

3.3. Proximal and distal causation in MDD

The previous sections provide some useful criteria for narrowing down the selection of potentially relevant causal factors for MDD. Certainly, very distal causes – such as the big bang, or a patient's having a heartbeat – are inappropriate candidates for being included in the

homeostatic property cluster insofar as these could serve equally well as an explanation of why someone does *not* suffer from MDD – they might in some sense have caused the effect, but they are only very contingently connected to it. For this reason, we can rule these out as primary causal factors. In part, this solves the problem of multifactorial causation by limiting the number of potential primary causes in the HPC.

However, the complex and heterogenous empirical evidence for the causal risk factors of MDD makes it more difficult to positively identify its proximal causes. Recall that it is possible to predict that the glass window will regularly break when subjected to such-and-such impact given certain background conditions. In light of our current evidence, we must conclude that a similar correspondence does not hold for MDD. This aligns with the previously established understanding that MDD is a less robustly epistemically valuable kind than e.g., gold.

Still, insofar as empirical evidence goes, we have good reason to consider neurochemistry a proximal cause of MDD. This does not amount to the essentialist stance because the HPC model continues to allow for some neurochemical and symptomatic variation between patients. It also does not necessarily amount to reductionism: the ball is not identical to the broken window, and neurochemistry need not be identical to the mental disorder in order to be identified as a proximal cause of it.

Scientific evidence testifies in favour of a strong relationship between brain states and mental states since the latter can be predictably impacted by systematically manipulating the former. This is especially evident in the case of the impact of traumatic brain injuries (Boake and Diller, 2005) and the influence of psychoactive drugs (Breithaupt & Weighmann, 2004) as well as targeted brain stimulation techniques such as transcranial magnetic stimulation (Walsh et al., 2006). While neurochemistry is conceivably not the only proximal cause of MDD, other such causes are difficult to pick out. A patient's environment *broadly understood* can likely be a second proximal cause, but the specific environmental factors that contribute to MDD are highly variable and difficult to study.

Given that we have good reason to consider neurochemistry a proximal cause (or at least a strongly correlated but ultimately contingent distal cause) of mental states more generally, this might give us good reason to consider neurochemistry a proximal cause of MDD as well, even if the precise neurochemical structures responsible for the disorder have not been identified. Given that proximal causes contribute with stability and hence epistemic value to the HPC, and since the interests of scientific researchers are with epistemically valuable entities, this gives us some

reason to consider neurochemistry a primary cause of MDD.

Tsou (2008) seems to follow the same line of reasoning when he like Kendler et. al (2011) defends a view of mental disorders as HPC kinds, but in contrast to these authors Tsou heavily prioritises neurobiological structures as the fundamental causal processes underlying such disorders. His reason to prioritise the neurochemical evidence above social og psychological risk factors is exactly the assumption that only biological structure can ensure the stability of the mental disorder categories. However, Tsou does not provide an explanation for the stability of MDD. He merely notes that this stability is confirmed by several lines of research, and that such stability justifies inferring by inference to the best explanation that mental disorders are real. The proximal/distal distinction seems capable of contributing to this account by illuminating the grounds for this stability and justifying the decision to consider neurochemical causes as primary.

3.4. Summary of Chapter 3

In this chapter I have attempted to redeem the HPC account of mental disorders from the problems of causal heterogeneity and multifactorial causation. Firstly, I have argued that the HPC account allows some variation on the level of causation, and hence allows some causal heterogeneity between individual instances of MDD.

Secondly, I have argued that it is possible to justifiably prioritise some causes among the great number of relevant causal factors by appealing to the proximal/distal distinction. Proximal causes are more closely related to the effect than distal causes, and hence are better suited for the role of primary causes in the HPC. I have shown why neurochemistry is a likely candidate for being a proximal (though not therefore singular!) cause of MDD.

In the following chapter I will go on to show the extent to which other considerations - pragmatic and normative - can, do, and ought to play justified roles in causal selection and hence in the resulting HPC kind categories.

Chapter 4
The Role of Normativity in Causal Selection

The attentive reader will have noticed that while the proximal/distal distinction may aid in non-pragmatically identifying one or several primary causal factors, this distinction alone is not enough to establish the kinds of decisive explanations we are often interested in. It still seems obvious that *the ball* broke the window even if other proximal factors were equally causally relevant to the outcome. In this chapter, I will show how pragmatic and normative considerations can undermine the apparent priority of proximal over distal causes. I then argue in favour of the inevitability and potential legitimacy of such influences in classificatory practice. Finally, I conclude that the HPC theory provides an ideal model for conceptualising MDD.

4.1. Normative influence in causal selection

4.1.1. Pragmatic value influences: Ability to intervene

In the previous chapter I have shown why proximal causes can be understood as having non-pragmatic priority over distal causes. But proximal causes are not the whole story. The purpose of pointing out a primary cause is often entwined with the aim of manipulating the effect in some way: If Peter did not want to break the window, he ought not to have thrown a ball at it! Such pragmatic considerations often influence the process of causal selection among all causal factors (Hart and Honoré, 1985). This is especially true for practical sciences such as medicine whose purpose is to prevent pathologies and produce health (ibid). Such sciences are likely to ignore causes which it is beyond their power to impact.

To relate back to the basketball scenario, the force of gravity might be another proximal cause of

the window breaking. If gravity were more intense, then the players would not have the strength to throw the ball to the level of the window. As the owner of the hall, you would avoid future accidents if only you could turn up gravity. Unfortunately, this is impossible. Since you cannot intervene in this way, pointing gravity out as a cause within this context has little pragmatic value. When contemplating how to keep your windows intact, you will likely ignore the factors you cannot change and focus on those you can *regardless* of whether these are proximal or distal. It is then not that these primary causes are privileged simpliciter. But you will focus on the manipulable causes of the problem, not those outside of your control.

Different people might have different capacities for intervening on different causes. It follows that they can differ in what causes they consider manipulable. An architect might cite the placement of the window as a cause for its breaking, while a glazier might quote the thickness of the glass. This grounds explanatory pluralism. This emphasis on control is a bias insofar as these causes are not independently privileged. But it is not irrational nor is it arbitrary. It is guided by real pragmatic considerations which can have a rational justification.

For this reason, when we are articulating causal relationships, we often tacitly appealing to a particular manner of solution or intervention. For example, when psychiatrists state that mental disorders are (primarily caused by) neurochemical imbalances, this seems to entail an appeal to treatment which directly targets these imbalances. This implicit emphasis on controllable causes has conceivably contributed to the focus on neurochemistry in psychiatry. It might simply be considered easier to intervene directly on the level of neurochemistry rather than e.g., taking complex measures to improve social function in family homes or equip a patient with a stable social network. In comparison, the production and prescription of drugs is standardised and relatively uncomplicated.

4.1.2. Other value influences: Willingness to intervene

Nevertheless, even if we accept that controllability plays a role for causal selection, still we do not have the full picture. Even within the same domain of research, far more causes *can* be manipulated than are selected as primary. Maria Kronfeldner (2014) introduces an additional normative aspect to the process of causal selection. She argues that causal selection is not merely determined by our ability to intervene on causal factors but equally by our willingness to do so. Imagine that you as the owner of the basketball hall has identified all the interventions that you could conceivably undertake to prevent your windows from breaking in the future. These include exchanging traditional basketballs for soft balls and filling the windows with cement. These are effective solutions. Unfortunately, they will undermine other aims and interests of yours. Given that you want people to continue playing basketball in your hall, you will only want to intervene when doing so does not impact the quality of the facilities. You are *unwilling* to compromise on things that would make your hall a less attractive choice for potential clients. Instead, you can ensure that there are no windows in especially vulnerable positions (such as behind a goalpost) and that the fragile glass windows are replaced by more sturdy polycarbonate windows. Once more, it seems that the proximal/distal distinction does not matter. You can intervene on distal causes as well as proximal ones in order to prevent a similar outcome for the future insofar as this does not compromise other interests of yours.

In the realm of psychiatry, we might have good reasons to be unwilling to prescribe anti-depressant medicine given its prevalent physical and psychological side-effects (Bolling and Kohlenberg, 2004) and withdrawal symptoms (Davies & Read, 2018) measured up against the merely moderate proven effect (Cipriani et al, 2018). On the other hand, we might have reason to prescribe such medicine in spite of these disadvantages if we nevertheless believe that it is the best available option.

4.1.3. Pragmatic and value-laden normativity in psychiatry

Considering the above, the primary causal factors we identify in psychiatry are then likely the factors which are both susceptible to intervention, *and* which we are willing to intervene on. These features depend on the relevant context in which the assessment takes place. In the context of clinical practice, little can be done about past adverse childhood experiences, though an adult patient's depression might conceivably have (partly) been caused by such experiences. Clinical practice focuses on treating the patient's present symptoms, not preventing them from occurring in the first place. In contrast, citing adverse childhood experience as a cause of depression in adulthood can help guide social policy decisions in a manner which aims to prevent future instances of MDD. The same disorder is being assessed in both cases, but the different capacities within each field means that they are likely to pay attention to distinct causes. Normative considerations also entail that while we might identify decisive proximal causes for some problem, it does not follow that intervention must happen directly on the level of these causes. In many cases of somatic medicine, we do successfully intervene directly on a proximal level, for example by treating a bacterial infection with antibiotics. But in other cases, we prefer to intervene elsewhere. For example, in spite of the correlation between smoking and lung cancer, we do not aim to invent medicine which will make lung cells more resistant to the destructive effects of tobacco smoke. This would entail great difficulty relative to other means of prevention, and it is unlikely that we are willing to exert efforts to ensure the continuation of what is largely considered a destructive habit. Instead, medical professionals, public health advocates simply discourage people from smoking tobacco![10] Ideally, this advice then informs public policy decisions, for example by raising taxes on tobacco products and ensuring accessible support for those who wish to quit smoking.

[10] In contrast, the tobacco industry advocates a harm-reduction approach which would not negatively impact the sale of tobacco (Dewhirst, 2020)

Similarly, even if we accept that the proximal causes of MDD are neurochemical, this does not entail that we ought to intervene *directly* on the level of neurochemistry if other treatments targeting distal causes are better suited for the purpose given our aims and values.

Naturally, it is not the case that 'anything goes' when it comes to choosing the relevant distal causes to intervene on. Along with the normative considerations, there must exist sufficiently strong correlation between the distal cause and the effect to justify intervening on the former in clinical practice as well as relating to policy decisions. For example, the link between childhood abuse and adult MDD is well-established (Humphreys et al. 2020). One study found that 75.6% out of 349 chronically depressed patients reported clinically significant histories of childhood trauma caused by emotional neglect and physical neglect, emotional abuse, physical abuse, and/or sexual abuse (Negele et. al, 2015). In contrast, other potential distal causes are less well-established and more speculative such as the hypothesis that excess sugar consumption plays a role in the aetiology of depression (Westover and Marangell, 2002). While normative considerations might cause us to prefer some interventions over others, we are also pragmatically justified in intervention partly due to the relevant probability that a distal cause will ultimately lead to the effect.

Conclusively, the status of neurochemistry as a proximal cause does not necessarily justify our prioritising it as a primary cause of MDD. Instead, we can be justified in picking out causes in accordance with both pragmatic and value concerns which reflect respectively our ability and our willingness to intervene on the causal factor in question. Instead of targeting a proximal cause directly, we can have good reason to intervene on the effect via sufficiently stably correlated distal causes (and hence assigning these the status as primary causes).

4.2. Is interest-relativity a problem for the HPC view?

If normativity impacts what causes we select as primary, and the HPC is a kind which is delineated partly in virtue of its causes, does this not compromise the integrity of the model?

Clearly, it quite some way off from the traditional mind-independent natural kinds presented in chapter 1. But is an entirely objective classification system even possible? In what follows I will examine whether paradigmatic essentialist kinds *actually* constitute interest-free classifications.

4.2.1. Interest-relative essentialist classification

Zachar proposes the following thought experiment: "Assume that we were to make contact with extra-terrestrial life, and the result of that contact was our initiation into a larger galactic community […] it turns out that all planets in the community use a gold-based currency, just like we do on Earth. What we call gold, however, is too common to be the galactic basis of a currency, but a particular isotope of the element at atomic number 79 is less common, and only that isotope is considered to be gold in the larger galaxy. Furthermore, this distinction is not an arbitrary one. Only that isotope of gold is stable over the long term in the energy fields of some planets, so it is the only kind of gold that can serve as an intergalactic standard." (Zachar 2002, p. 220)

Indeed, not all instances of gold-molecules are identical. The kind 'gold' allows for a variety of isotopes. How come? It appears that the category 'gold' ignores these divisions because they are not (currently) relevant for us. Under different circumstances, such as those specified in the thought experiment, we might want to limit the kind gold to apply to a single isotype. It seems then that even in the realm of proposed essentialist natural kinds, our classificatory practices admittedly do partly track the internal structure of the kind, but simultaneously continues to let themselves be guided by pragmatic considerations. Hence, the inability to ground an interest-free classification system is not unique to the HPC theory but instead reflects the general impossibility of interest-free classification. This aligns with the general philosophical consensus that science is not, in fact, a value-free manner of knowledge production, but is deeply entwined with human interests and values in its practices of classification as well as concepts, assumptions,

questions, and choices of scientific methods (Elliott, 2017).

4.2.2. Illegitimate value influences in psychiatry

Naturally, and concerningly, such normative influences can also be illegitimate. Psychiatry has been vehemently accused of being excessively influenced by financial motivations of both the large pharmaceutical industry as well as the American Psychiatric Association which publish the DSM. In 2020, the global market for antidepressant drugs was estimated at a value of 26.25 billion USD (The Business Research Company, 2021). It has been revealed that some pharmaceutical companies have manipulated the results of clinic trials misrepresent their products as more efficient and their side-effects as less severe than the unmanipulated results otherwise suggested (Elliott, 2017). Of the 522 trials included in the previously referenced extensive recent meta-analysis which concluded that antidepressant medication has a moderate effect relative to placebo, 409 were funded by pharmaceutical companies (Cipriani et al, 2018). The APA's main revenue is publishing new editions of the DSM and receiving contributions from the pharmaceutical industry (van der Kolk, 2014). The APA has been accused of encouraging diagnostic inflation by inventing gratuitous new psychiatric categories and broadening existing psychiatric categories, thereby increasing the number of diagnosed patients to whom neuro-pharmaceutical drugs can be marketed and sold (Frances, 2015). 69% of the 160 top researchers in the DSM-5 task force report having individual ties to the pharmaceutical industry (Cosgrove & Drimsky, 2012). The biomedical model of mental disorders might also find political support insofar as the conception of mental disorder as inherent in the individual draws attention away from social factors which conservative political forces are likely unwilling to change.

While I cannot defend my stance in detail here, I do believe that these criticisms are justified, and that such financial interests are in part responsible for the insistence on maintaining the biomedical model of mental disorders. Either way, it appears that where neuro-pharmaceutical

medicine is concerned there is a genuine but moderate ability to *actually* intervene on the effect, but a disproportionately vast willingness to try! This willingness is likely unduly exaggerated by financial and political interests like those described above. Still, I believe that this is not a point against the potential legitimacy of non-epistemic value-influence in psychiatric practice, but rather ought to encourage transparency and reflexivity about these values and their impact.

4.3. A methodological framework for interest-relative classification

Ingo Brigandt (2020) makes a similar point. He proposes a general methodological framework for philosophically studying kinds which emphasises both the inevitability and legitimacy of non-epistemic values and aims which influence our classificatory practices. He argues that *any* kind must be investigated "alongside the human aims and purposes which motivate referring to this kind in scientific theorising and practice." (p. 2) This is especially important given that what counts as the most appropriate boundaries of a kind are often determined by the aims which motivate the inquiry. Such aims are not necessarily epistemic, but can for example also include political, ethical, or environmental values.

Hence, kinds – including HPCs - must be articulated *along with* the specific aims which motivate referring to the kind in scientific practice. This is important firstly because these aims can justify whether a kind category is relevant, and secondly because these aims set the standard for what counts as the most appropriate account of the nature and boundaries of the kind in question. Importantly, Brigandt argues that epistemic aims have been considered the most apt for serving these purposes, but he wants to emphasise the extent to which non-epistemic aims can and do successfully aid in establishing appropriate kind-categories. Naturally, explicitly articulating these values equips us to assess whether their influence is legitimate or not.

This is especially relevant to psychiatry insofar as diagnostic categories are deeply evaluative. Recall that the DSM-5 states that "An expectable […] response to a common stressor […] is not a mental disorder" (p. 20). So, what we are asking when inquiring whether some cluster of behaviour is a disorder is whether it warrants treatment *given what it is a response to?* If we neglect this question, we risk creating diagnostic categories which pathologise and invalidate warranted responses. In doing so, we further risk contributing to diagnostic inflation as well as obscure and invalidate the genuine importance of the social, environmental, and political factors which has caused the apparent symptoms.

Nancy Potter (2014) makes a convincing case that the psychiatric diagnosis "oppositional defiant disorder" (APA, 2013) has been defined and formulated in a manner which disproportionately target young socioeconomically disadvantaged African American boys who act out in school. Potter argues that this behaviour is in fact a rational response to systemic racist oppression rather than indicative of clinical pathology. Maintaining – and articulating – an active awareness of the non-epistemic aim of social justice which ought to underlie psychiatric classification might prevent (or at least resist) the recognition of such a category. Articulating social justice as an aim underlying classificatory practice would motivate closer attention the social causes of perceived symptoms which would likely hinder the formation of such illegitimate diagnostic categories. Similarly, when individual patients present with depressive symptoms, such an awareness ought to guide whether they are liable candidates for a diagnosis (are they recently bereaved? Overworked?) and what kind of treatment they ought to receive (e.g., conversational therapy vs. Medication). Note that this sensitivity to normative influences again makes reference to causes – how did the apparent symptoms originate, not just biologically, but socially? Rather than deny or suppress interest-relativity and normative influence in psychiatric classification, we ought to accept it, make it transparent, and let it guide us towards classifications which serve both our epistemic and non-epistemic aims and values.

4.4. Why is the HPC theory the best model of MDD?

Why, then, does the HPC theory offer an ideal model of MDD? To summarise the findings of the preceding chapters, the HPC model allows for flexibility and explanatory pluralism insofar as its causal mechanisms are determined via causal selection which depends on both the epistemic and non-epistemic commitments of the context of inquiry. The HPC theory can therefore accept the potential importance of neurochemical factors on sustaining the symptoms of MDD while not therefore entailing neither that the disorder is reducible to neurochemical factors nor that interventions ought necessarily to happen directly on the level of neurochemistry.

Because the HPC model transparently incorporates probabilistic causal relations, it invites an answer to the question of what normative considerations underlie our causal selection and hence encourages reflexivity and transparency in this regard, thus paving the way for utilising the ideal methodological framework for studying kinds proposed by Brigandt. Still, the HPC resists collapse into relativism because it remains grounded in real causal structures as well as real symptom clusters. For the same reason, the HPC model maintains the epistemic value associated with traditional essentialist kinds, and it allows that these stably clustering symptoms and causal mechanisms can be discovered empirically.

What does the conception of MDD as an HPC kind entail for psychiatry?

Given the probabilistic and heterogenous nature of the MDD's causal background, diagnostic practice is likely to continue to happen based on observed symptoms in combination with the relevant exclusion criteria rather than proposing one or several aetiologies as diagnostic criteria. Psychiatric categories will for the same reason likely continue to be formally defined predominantly in virtue of their symptoms. As shown, this is not in conflict with the HPC view. Nevertheless, stable correlations between symptom clusters and their risk factors - biological as well as social - ought to be acknowledged in the literature and in psychiatric practice more broadly. Further, the epistemic and non-epistemic values which feature in the causal selection

process ought to be articulated and considered both when new diagnostic categories are established, when deciding whether an individual patient qualifies for a diagnosis, and when deciding on an appropriate course of treatment (medical or non-medical). This will help prevent essentialist assumptions pertaining to the nature of MDD from gaining undue scientific legitimacy, and hopefully aid in expanding the scope of research, diagnosis, and treatment beyond the neuro-centric confines of the biomedical model.

On the HPC view, mental disorders have vague boundaries and are therefore indeterminate rather than categorical. This means that many people will conceivably exist in a diagnostic grey-area rather than being unambiguously afflicted by the disorder. I believe this reflects the genuine complexity of such disorders. Still, clinical diagnostic practice will presumably continue to impose some diagnostic thresholds (such as the DSM-5's minimum number of instantiated symptoms) for purely practical purposes. Thus, the formal definition of our psychiatric categories will not map onto the disorder perfectly, but that is to be expected given its genuinely vague boundaries which conflict with the pragmatic need for decisive demarcation for e.g., treatment and insurance purposes.

Conclusion

In this thesis, I have argued that the homeostatic property cluster theory is an ideal model for capturing the nature of genuine mental disorders. I have used MDD as an example throughout the thesis, but my intention has not been to defend the validity of the current MDD diagnostic criteria as such. I have argued that causation is an indispensable part of conceptualising psychiatric disorders, and simpler models which reject appeal to causation in favour of exclusive focus on pragmatic value and inductive potential are inadequate for the purposes of psychiatry. I suspect that much of the motivation for denying causation in the classification of mental disorders stems from a resistance to the biomedical model. I sympathise and agree with this resistance and hope to have shown that appeal to underlying causal mechanisms does not entail a commitment to a conception of mental disorders as purely physical brain disorders since such mechanisms can also be social, psychological, or environmental. But neither does it entail a rejection of the potential importance of biological factors for mental disorders, nor a categorical rejection of medication as a viable treatment. The HPC model can maintain that MDD can and likely does have a neurochemical component which contributes to its stability and that psychiatry can be a legitimate medical science. Simultaneously, because the selection of primary causal factors among the many homeostatic mechanisms is necessarily guided by both epistemic and non-epistemic aims and values, the HPC model provides a framework which firstly allows social and psychological factors to play a genuine causal role as homeostatic mechanism, secondly allows that such non-biological causes can potentially override the prima facie causal priority of neurochemistry, and thirdly encourages transparency and reflexivity about what epistemic and non-epistemic influences guide our causal selection. Thus, the HPC model has the potential for aiding in minimising diagnostic inflation and resisting the monopoly of the biomedical model insofar as these practices conflict with other important epistemic and non-epistemic aims and values. So, while mental disorder categories are shaped by normative influences and therefore

not natural kinds in the traditional sense, they remain epistemically valuable and legitimate objects of both empirical and theoretical study.

Bibliography

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington.

Anonymous. (2013) *No Dishonour in Depression*. Nature, p. 498.

Arnold M., Müller-Oerlinghausen B., Hemrich N., Bönsch D. (2020): *Effects of Psychoactive Massage in Outpatients with Depressive Disorders: A Randomized Controlled Mixed-Methods Study.* Brain Sci, 10.

Beebee, H. (2004): *Causing and nothingness*. In J. Collins, N. Hall, & L. A. Paul (Eds.), Causation and counterfactuals. MIT Press.

Belmaker, R. H. and Agam, G. (2008): *Major Depressive Disorder*. New England Journal of Medicine, 358.

Boake, C. and Diller, L. (2005): *History of Rehabilitation for Traumatic Brain Injury*. In W. M. High, Jr., A. M. Sander, M. A. Struchen, & K. A. Hart (Eds.), Rehabilitation for traumatic brain injury. Oxford University Press

Bolling M., and Kohlenberg R. (2004): *Reasons for quitting serotonin reuptake inhibitor therapy: paradoxical psychological side effects and patient satisfaction.* Psychother Psychosom 73.

Boorsboom, D., Cramer, A. and Kalis, A. (2019): *Reductionism in Retreat*. Behavioral and Brain Sciences; New York, 42.

Boyd R. (1999); *Homeostasis, species, and higher taxa*. In: Wilson RA (ed) Species: New Interdisciplinary Essays. MIT Press.

Boyd, R (1999): *Kinds, Complexity and Multiple Realization: Comments on Millikan's 'Historical Kinds and the Special Sciences*. Philosophical Studies 95.

Boyd, R. (1999b): *Homeostasis, Species, and Higher Taxa*. in R. Wilson (ed.), Species: New Interdisciplinary Essays. MIT Press.

Breithaupt, H. and Weigmann, K. (2004): *Manipulating your mind*. EMBO reports, 5(3).

Bridges, L. and Sharma, M. (2017): *The Efficacy of Yoga as a Form of Treatment for Depression*. Journal of Evidence-Based Integrative Medicine, 22(4).

Brigandt, I. (2020): *How to Philosophically Tackle Kinds without Talking About 'Natural Kinds'.* Canadian Journal of Philosophy.

Chakravartty, A. (2007): *A metaphysics for scientific realism: Knowing the unobservable*. Cambridge University Press.

Chi X, Bo A, Liu T, Zhang P and Chi I (2018): *Effects of Mindfulness-Based Stress Reduction on Depression in Adolescents and Young Adults: A Systematic Review and Meta-Analysis*. Front. Psychol. 9

Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. (2018): *Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis*. Focus, 16(4).

Clayton, A. H., Croft, H. A., Horrigan, J. P., Wightman, D. S., Krishen, A., Richard, N. E., & Modell, J. G. (2006): *Bupropion extended release compared with escitalopram: effects on sexual functioning and antidepressant efficacy in 2 randomized, double- blind, placebo-controlled studies*. The Journal of Clinical Psychiatry, 67(5).

Cooper, R. (2007): *Psychiatry and the Philosophy of Science*. Routledge.

Cosgrove, L and Drimsky L. (2012): *A comparison of DSM-iv and DSM-5 panel members' financial associations with industry: A pernicious problem persists*. PLOS Medicine. 9(3).

Craver, C. (2009): *Mechanisms and Natural Kinds.* Philosophical Psychology 22(5).

Crocq, M. (2000): *From shell shock and war neurosis to posttraumatic stress disorder: a history of psychotraumatology.* Dialogues Clin. Neurosci, 2(1).

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). *Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies*. Journal of Consulting and Clinical Psychology, 76(6).

Davies J. and Read J. (2018): *A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: are guidelines evidence-based*? Addict Behav.

Delgado, P. (2000): *Depression: The Case for a Monoamine Deficiency*. Journal of Clinical Psychiatry 61(6).

Delgado, P. (2011): *The Tryptophan Depletion Challenge Test in Medical Research: Unresolved Issues and Broader Implications for the Use of Physiological Challenges to Investigate and Categorize Disease.* Biological Psychiatry, 69.

Dewhirst, T. (2020): *Co-optation of harm reduction by Big Tobacco.* Tobacco Control. Accessed online: 12 May 2021.

Elliott, K. (2017): *A Tapestry of Values: An Introduction to Values in Science*. New York: Oxford University Press,

Foucault, M. (1961/2006). *History of Madness.* Routledge.

Frances, A. (2015): *Lessons from the DSM-5.* Psychopathology Review 2(1).

Hacking, I. (1992): *Multiple Personality Disorder and its Hosts*. History of the Human Sciences, 5(2).

Hacking, I. (1995): *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton University Press.

Hacking, I. (2007): *Kinds of People*. Proceedings of the British Academy, 151.

Harmer, C., Duman, R., Cowen, P. (2017): *How do antidepressants work? New perspectives for refining future treatment approaches.* The Lancet Psychiatry 4(5).

Hart H., and Honore´T. (1985): *Causation in the Law*. 2nd edn. Clarendon Press.

Haslam, N (2000): *Psychiatric Categories as Natural Kinds: Essentialist Thinking About Mental Disorder.* Social Research 67.

Haslam, N. (2002): *Kinds of kinds: A conceptual taxonomy of psychiatric categories.* Philosophy, Psychiatry, and Psychology 9(3).

Hirschfeld, R. (2000): *History and Evolution of the Monoamine Hypothesis of Depression.* Journal of Clinical Psychiatry 61(6).

Hitchcock, C. (2001): *The Intransitivity of Causation Revealed in Equations and Graphs*. Journal of Philosophy, 98.

Hitchcock, C. (2007) *Prevention, Preemption, and the Principle of Sufficient Reason*. Philosophical Review, 116.

Hucklenbroich, P. (2014). *Disease Entity' as the Key Theoretical Concept of Medicine*. Journal of Medicine and Philosophy, 39.

Humphreys K, LeMoult J, Wear JG, Piersiak H, Lee A, Gotlib I (2020): *Child maltreatment and depression: A meta-analysis of studies using the Childhood Trauma Questionnaire*. Child Abuse Negl. 102.

Hyman, S. (2010): *The diagnosis of mental disorders: The problem of reification*. Annual Review of Clinical Psychology 6.

Häggqvist, S. (2005): *Kinds, Projectibility and Explanation*. Croatian Journal of Philosophy 5(13).

Insel, T. and Cuthbert, B. (2015). *Brain disorders? Precisely*. Science, 348.

Jackson, G. (2005): *Rethinking Psychiatric Drugs: A Guide for Informed Consent.* Authorhouse.

Kendler K., Zachar P., and Craver, C. (2011): *What kinds of things are psychiatric disorders?* Psychological Medicine 41.

Kendler, K. (2008): *Explanatory Models for Psychiatric Illness*. Am J Psychiatry,165(6).

Kendler, K. (2012): *The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism*. Molecular Psychiatry 17.

Kendrick, T. (2020): *Strategies to Reduce use of Antidepressants.* British Journal of Clinical Pharmacology, 87(1).

Khalidi, A. (2021): *Etiological Kinds.* Philosophy of Science, 88.

Koenig A, and Thase M. (2009): *First-line pharmacotherapies for depression – what is the best choice?* Pol. Arch. Med. Wewn. 119.

Kring, A., Davison, Gerald C, Neale, John M., and Johnson, Sheri L. (2007*). Abnormal Psychology*. Hoboken, NJ: Wiley & Sons.

Kronfeldner, M. (2014): *How Norms Make Causes*. International Journal of Epidemiology, 43.

Kutchins, H., and Kirk, S. (1997): *Making us crazy: DSM, the psychiatric bible and the creation of mental disorders*. New York: Free Press.

Lewis, D. (1973): *Causation*. The Journal of Philosophy, 70.

Lewis, D. (1979): *Counterfactual Dependence and Time's Arrow*. Nous, 13.

Lewis, D. (1986). *Causal explanation.* In idem, Philosophical papers. Oxford: Oxford University Press.

Lilienfeld, S. (2014): *The Research Domain Criteria (RDoC): An analysis of methodological and conceptual challenges.* Behaviour Research and Therapy 62.

Luhrmann, T. (2001): *Of Two Minds: An Anthropologist Looks at American Psychiatry.* Vintage.

Mathews, D. C., & Zarate, C. A. (2013): *Current status of ketamine and related compounds for depression.* The Journal of Clinical Psychiatry, 74(5).

Meyer, J., Houle, S., Sagrati, S., Carella, A., Hussey, D., Ginovart, N., Goulding, V., Kennedy, J., and Wilson, A. (2004): *Brain serotonin transporter binding potential measured with carbon 11- labeled DASB positron emission tomography: effects of major depressive episodes and severity of dysfunctional attitudes*. Archives of General Psychiatry, 61(12).

Miller, G. (2010): *Mistreating psychology in the decades of the brain*. Perspectives on Psychological Science 5(6).

Muang, H. (2016): *Do Psychiatric Diagnoses Explain?* Doctoral thesis. Lancaster University.

Murphy, D. (2006): *Psychiatry in the scientific image*. Cambridge: MIT Press

Negele, A., Kaufhold, J., Kallenbach, L., and Leuzinger-Bohleber, M. (2015): *Childhood Trauma and Its Relation to Chronic Depression in Adulthood.* Depression research and treatment.

O'Connor, J. (1976): *Causal overdetermination and counterfactuals.* Philosophical Studies 29(4).

Olbert, C. and Gala, G. (2015): *Supervenience and psychiatry: Are mental disorders brain disorders?* Journal of Theoretical and Philosophical Psychology, 35(4).

Owens, M. (2004): *Selectivity of Antidepressants: From the Monoamine Hypothesis of Depression to the SSRI Revolution and Beyond.* Journal of Clinical Psychiatry 65(4).

Parsey, R., Oquendo, M., Ogden, R., Olvet, D.,, Simpson, N., Huang, Y., Heertum, R., Arango, V., and Mann, J. (2006): *Altered serotonin 1A binding in major depression: a [carbonyl- C-11]WAY100635 positron emission tomography study*. Biological Psychiatry, 59(2).

Poland J, Von Eckhardt B, Spaulding W. (1994): *Problems with the DSM Approach to Classifying Psychopathology.* In Philosophical Psychopathology, eds. Graham G. and Stephens L. The MIT Press.

Poland, J. (2014): *Deeply Rooted Sources of Error and Bias in Psychiatric Classification.* In Classifying Psychopathology, eds. Sullivan, J. and Kincaid, H. The MIT Press.

Potter, N. (2014): *Oppositional defiant disorder: cultural factors that influence interpretations of defiant behavior and their social and scientific consequences*. In Mental Kinds and Natural Kinds, eds. Kincaid H. and Sullivan J. The MIT Press.

Prathikanti S, Rivera R, Cochran A, Tungol JG, Fayazmanesh N, et al. (2017): *Treating major depression with yoga: A prospective, randomized, controlled pilot trial*. PLOS ONE, 12(3)

Reivich, M., Amsterdam, J., Brunswick, D. J., and Yann Shiue, C. (2004): *PET brain imaging with [11C](+)McN5652 shows increased serotonin transporter availability in major depression.* Journal of Affective Disorders, 82(2).

Rosenberg, C. (2006): *Contested Boundaries: Psychiatry, Disease, and Diagnosis.* Perspectives in Biology and Medicine, 49(3).

Slater, M. (2015: *Natural Kindness*. British Journal for the Philosophy of Science, 66.

Slothouber, N. (2019*). In Search of Psychiatric Kinds: Natural Kinds and Natural Classification in Psychiatry.* Doctoral thesis. The University of Western Ontario

Szasz, T. (1962): The Myth of Mental Illness: Foundations of a Theory of Personal Conduct. Harper & Row.

The Business Research Company (2021): *Antidepressants Global Market Report 2021: COVID-19 Implications And Growth To 2030.* Accessed online May 12th, 2021.

Tsou, J. (2008): *The Reality and Classification of Mental Disorders: Issues in the Philosophy of Psychiatry.* Doctoral dissertation, Chicago University.

Tsou, J. (2016): *Natural Kinds, Psychiatric Classification and the History of the DSM.* History of Psychiatry 27(4).

Van der Kolk, B. (2014): *The Body Keeps the Score. Mind Brain and Body in the Transformation of Trauma.* Penguin Books.

Walser, T., Cui, X., Yanagawa, J., Lee, J., Heinrich, E., Lee, G., Sharma, S. and Dubinett, S. (2008): *Smoking and lung cancer: the role of inflammation*. Proceedings of the American Thoracic Society, 5(8).

Walsh V, Desmond J, Pascual-Leone A. (2006): *Manipulating brains.* Behav Neurol. 17(3-4).

Westover, A and Marangell, L. (2002): *A cross-national relationship between sugar consumption and major depression?* Depression & Anxiety, 16(3)

Whitaker, R. (2011): *Anatomy of an Epidemic: Magic Bullets, Psychiatric Drugs, and the Astonishing Rise of Mental Illness in America.* New York: Random House.

Woodward, J. (2003): *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

Zachar, P. (2002): *The Practical Kinds Model as a Pragmatist Theory of Classification*. Philosophy, Psychiatry, & Psychology, 9(3).