In Search of Psychiatric Kinds: Natural Kinds and Natural Classification in Psychiatry

Nicholas Slothouber
The University of Western Ontario

Supervisor
Sullivan, Jacqueline
The University of Western Ontario
Klimchuk, Dennis
The University of Western Ontario

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Recommended Citation
https://ir.lib.uwo.ca/etd/6627

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.
ABSTRACT

In recent years both philosophers and scientists have asked whether or not our current kinds of mental disorder—e.g., schizophrenia, depression, bipolar disorder—are natural kinds; and, moreover, whether or not the search for natural kinds of mental disorder is a realistic desideratum for psychiatry. In this dissertation I clarify the sense in which a kind can be said to be “natural” or “real” and argue that, despite a few notable exceptions, kinds of mental disorder cannot be considered natural kinds. Furthermore, I contend that psychopathological phenomena do not cluster together into kinds in the way that paradigmatic natural kinds (e.g., chemical kinds and species) do; and, in light of this fact, I conclude that the normative ideal of natural classification—i.e. classifying real or natural kinds—is not appropriate as a general strategy for psychiatry. In the conclusion to this dissertation I propose an alternative way forward. Rather than assuming a priori that psychopathological phenomena cluster together into kinds, a more tractable and theoretically promising approach would be to first explain how particular kinds of experiences and behaviours—e.g, hallucinations, delusions, disorganized thinking, mania, low mood, etc.—are produced.

Keywords: Natural Kinds; Homeostatic Property Cluster (HPC) Kinds; Mechanistic Property Cluster (MPC) Kinds; Mental Disorders; Mental Illnesses; Psychiatric Kinds; DSM Categories; Depression; Schizophrenia; Manic-Depressive Illness (Bipolar Disorder); Psychiatric Classification; Natural Classification; Diagnostic and Statistical Manual of Mental Disorders (DSM)
SUMMARY FOR LAY AUDIENCE

A problem that scientists, psychiatrists, and philosophers have been wrestling with in recent years is how to best classify mental disorders. The primary motivation for this discussion is the belief that current psychiatric classifications have done a poor job of promoting the development of effective therapeutic interventions for mental illness.

Although there is widespread agreement amongst scientists and philosophers with respect to what psychiatric classification has failed to do, it is not yet clear precisely why it has failed and how to go about fixing it. One suggestion that has gained popularity in recent years is that the psychiatric classification has failed because its categories do not really reflect the way the world is. For example, two people diagnosed with schizophrenia can be so different (e.g., one might be suffering from delusions and hallucinations, while another might be suffering from speech and motor problems) that it is more likely than not that their problems are caused by different things, which strongly suggests that providing them with the same treatment will be unhelpful. Nevertheless, there is still a lot of disagreement about whether or not categories like schizophrenia do reflect the way the world is. Some argue that schizophrenia is real, while others argue that it is a made up.

In this dissertation I clarify the sense in which mental disorders can be said to be “real” and argue that, despite a few notable exceptions, mental disorders cannot be said to be real, and are better understood as hypothetical constructs. Furthermore, I argue that hypothetical constructs like schizophrenia and depression are not useful for psychiatry, and a better way forward is to focus on trying to explain and treat particular experiences and behaviours—e.g., hallucinations, delusions, low mood, etc.
ACKNOWLEDGEMENTS

I am grateful to numerous people at Western University and elsewhere who provided critical feedback, comments, and suggestions on various parts of this dissertation. These include Frédéric Banville, Tim Bayne, Susan Berkhout, Ingo Brigandt, Justin Bzovy, Rachel Cooper, Dennis Klimchuk, Derek Mitchell, Derek Oswick, Jamie Shaw, Christopher Slothouber, Vanessa Slothouber, Jackie Sullivan, Kathryn Tabb, Şerife Tekin, Jonathan Tsou, Peter Verveniotis, Adam Woodcox, Jessey Wright, and Peter Zachar.

I owe special thanks to the members of my dissertation committee, Dennis Klimchuk and Jackie Sullivan, for their valuable feedback, encouragement, and patience with this project. I also owe a debt of gratitude to Dawn Good, who I’ve worked with for more than a decade, and who has helped shape my understanding of neurological and psychiatric disorders in particular, and the mind/brain in general.

Earlier drafts of material from this dissertation were presented at the Association for the Advancement of Philosophy and Psychiatry (AAPP) in May 2016 and May 2017, the Canadian Philosophical Association (CPA) in June 2016, and the Canadian Society for the History and Philosophy of Science (CSHPS/SCHPS) in June 2017 and June 2019. I am especially grateful to Susan Berkhout, Ingo Brigandt, Rachel Cooper, Kathryn Tabb, Şerife Tekin, Jonathan Tsou, and Peter Zachar for their helpful feedback at these meetings.

Finally, I would like to express my deep gratitude to my parents, family, and friends for their encouragement during the dissertation writing process. In particular, I am
indebted to both my brother and sister for reading and discussing various drafts of this dissertation, and my wife for her emotional support and patience throughout my studies.
# TABLE OF CONTENTS

Abstract ........................................................................................................... i

Summary for Lay Audience ........................................................................... ii

Acknowledgments ........................................................................................ iii

Introduction ................................................................................................... 1

Chapter 1: Natural Kinds and Natural Classification ................................. 11

Chapter 2: Psychiatric Kinds and Natural Kinds ....................................... 51

Chapter 3: Psychiatric Kinds and Homeostatic Property Cluster Kinds ...... 93

Chapter 4: Psychiatric Kinds and Natural Classification .......................... 130

Conclusion ................................................................................................... 179

References .................................................................................................. 187
INTRODUCTION

Classification, or “nosology” in the technical language of the history of medicine, has always been the sine qua non of psychiatry. There are good historical reasons for this. In the early nineteenth century, when psychiatry was still emerging as a distinct discipline, little was understood about the nature of psychopathological phenomena. Psychiatrists hoped that, by first classifying psychopathological phenomena into different categories (on the basis of similarities in behavioural signs, symptoms, course, and outcome), their causal basis could then be investigated, and real kinds of mental disorder would eventually be discovered (Cf. Goldstein, 1987; Shorter, 1997, chapter 2 & chapter 3; and Harrington, 2019, chapter 1).

By the time the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM)—DSM-III—rolled around in 1980 the situation was much the same. Not enough was known about the causal nature of mental disorders, and so DSM-III nosologists settled on predetermined sets of symptom-based inclusion and exclusion criteria, which could be used consistently from clinician-to-clinician, country-to-country. As both Shorter (1997, 300-301) and Harrington (2019, 134) note, the hope was that the new DSM categories would correspond to real kinds of mental disorder, and the belief was that biological causes would eventually be discovered for all mental disorders.

Initially, DSM-III was hailed as a crowing achievement of psychiatry. The first two editions of the DSM—DSM-I and DSM-II—were heavily influenced by psychoanalytic theory, received very little fanfare, and did not appreciably improve diagnostic consistency (Cf. Shorter, 1997, chapter 8). Copies of DSM-III, however, sold
out immediately and were on backorder for approximately 6 months. As Andreason (2007, 111) puts it: “DSM-III and its successors…became universally and uncritically accepted as the ultimate authority on psychopathology and diagnosis”.

Fast forward to 2013 and Thomas Insel, former director of the National Institute of Mental Health (NIMH), had this to say about the most recent edition of the DSM, DSM-5:

In a few weeks, the American Psychiatric Association will release its new edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). This volume will tweak several current diagnostic categories, from autism spectrum disorders to mood disorders. While many of these changes have been contentious, the final product involves mostly modest alterations of the previous edition, based on new insights emerging from research since 1990 when DSM-IV was published. Sometimes this research recommended new categories (e.g., mood dysregulation disorder) or that previous categories could be dropped (e.g., Asperger’s syndrome).

The goal of this new manual, as with all previous editions, is to provide a common language for describing psychopathology. While DSM has been described as a “Bible” for the field, it is, at best, a dictionary, creating a set of labels and defining each. The strength of each of the editions of DSM has been “reliability” – each edition has ensured that clinicians use the same terms in the same ways. The weakness is its lack of validity. Unlike our definitions of ischemic heart disease, lymphoma, or AIDS, the DSM diagnoses are based on a consensus about clusters of clinical symptoms, not any objective laboratory measure. In the rest of medicine, this would be equivalent to creating diagnostic systems based on the nature of chest pain or the quality of fever. Indeed, symptom-based diagnosis, once common in other areas of medicine, has been largely replaced in the past half century as we have understood that symptoms alone rarely indicate the best choice of treatment.

Patients with mental disorders deserve better (Insel, 2013).

In Insel’s mind, the DSM is at best a “dictionary” rather than a proper medical diagnostic system. Others have been far less kind. For instance, in his book Madness Explained:

*Psychosis and Human Nature* clinical psychologist Richard Bentall draws an analogy
between psychiatric classification and astrology, going so far as to say that current mental disorder categories are not much better than star signs, another persistent and widely accepted diagnostic system (Bentall, 2005).¹

Notwithstanding the vitriolic attacks launched on psychiatry in recent years there is still widespread disagreement amongst philosophers and scientists with respect to what kinds of things current mental disorders are. Although it is true that categories of mental disorder have been, and still are, constructed on the basis of clusters of observable signs and symptoms, does this imply that they are merely convenient pigeonholes and hence not real?

Philosophers and scientists alike have offered different and sometimes contradictory answers to this question. For instance, in 1961 the Hungarian psychiatrist Thomas Szasz published *The Myth of Mental Illness*, arguing that real medicine deals with disease, and there is simply no evidence to suggest that kinds of mental disorder are real kinds of disease.² Later, philosopher Christopher Boorse (1975, 1976) put forth a rebuttal, contending that mental illnesses are diseases underwritten by biological dysfunctions. According to Boorse, the human body and mind is comprised of various subsystems (e.g., organs, the nervous system, different brain modules) which perform various functions (e.g., the heart pumps blood, the visual system receives, transduces, and codes visual information), and a “biological dysfunction” is simply a subsystem which is

---

¹ See also: Bentall (2006, 223).

² By “disease” Szasz meant a medical problem which arises from an observable anatomical abnormality, such as a brain lesion.
no longer working properly.³

More recently, debates concerning the ontological status of psychiatric kinds have moved away from narrow accounts of “disease” and toward a broader discussion of “natural kinds”. In particular, philosophers have asked whether or not our current kinds of mental disorder are “real” or “natural” (irrespective of whether or not they are diseases); and, moreover, whether or not the search for natural kinds of mental disorder is a realistic desideratum for psychiatry.

As with previous discussions concerning the disease-status of mental disorders, the answers have been varied and contradictory. Some philosophers argue that at least some kinds of mental disorder are natural or real and that psychiatric classification ought generally to be developed in terms of a natural kind approach, since it would be an improvement on current psychiatric classification (Cooper, 2005, 2007; Kender, Zachar & Craver, 2011; Tsou, 2007, 2008, 2012, 2013, 2016). Others contend that current mental disorders are not natural kinds at all, and that the natural kind approach to psychiatric classification may not even be possible or desirable (Hacking, 1992, 1995a, 1995b, 1999, 2007c; Hyman, 2010; McGuinn, 1991; Zachar, 2002, 2014a, 2014b).

The view that psychiatric kinds are, or ought to reflect, natural kinds will be the main point of contention throughout this dissertation. Recent work in the philosophy of psychiatry highlights just how important the concept of ‘natural kind’ is thought to be for psychiatric classification. Cooper (2005, 2007), for instance, contends that identifying

³ Wakefield (1992a, 1992b, 1999) provides a similar account, but conceives of the natural function of a sub-systems in an evolutionary sense—namely, a “natural function” is that which has been selected over the course of evolutionary history.
natural kinds of mental disorder is important for ensuring that psychiatric categories support generalizations by picking out more-or-less homogenous kinds. Others argue that identifying natural kinds is necessary for moving from a symptom-based to a causally-based classification system, which would be in line with much of the rest of medicine (Cf. Kendler, Zachar & Craver, 2011; Tsou, 2007, 2008, 2012, 2013, 2016). In short, many philosophers and scientists believe that the solution to the current crisis of psychiatric classification is to revise our categories in such a way as to reflect natural kinds.

Be that as it may, the phrase ‘natural kind’ is very much a philosophical term of art. Although philosophers often draw a contrast between kinds that exist in nature and kinds that are constructed for our own convenience, beyond this intuitive contrast there has been considerable debate in philosophy about what constitutes a natural kind. In chapter 1 I clarify the terms of this debate. In the end, I argue that the central idea running through the history of natural kind theorizing is that natural kinds are, at bottom, collections of individuals who are alike in important respects—i.e. individuals are members of a natural kind in virtue of having in common important properties. Moreover, “important properties” are properties which cluster or “hang together” in virtue of being produced by certain causal mechanisms. In short, natural kinds are clusters of properties which cluster together because they have a common causal structure.

In chapter 2 I contend that, despite a few notable exceptions, there is no reason to think that psychiatric categories or kind-concepts are natural kinds. In particular, I argue that, while nothing in principle prevents mental disorders from being natural kinds, paradigm examples of kinds of mental disorder, such as schizophrenia and depression, are
individuated by superficial rather than important similarities, and so cannot be said to be natural kinds in the sense outlined in chapter 1. More specifically, psychiatric kinds are not natural kinds since we cannot even explain causally how it is that the signs and symptoms that are characteristic of putatively distinct kinds of mental disorder are produced. Without knowing the true causal structure of a particular kind of mental disorder, it is impossible to know whether the properties definitive of a psychiatric kind do indeed cluster together, or whether these groupings are merely a reflection of our predilections to group certain properties together.

In chapter 3 I address philosophical arguments claiming that kinds of mental disorder are natural kinds. Although psychiatric kinds are not natural kinds in the sense outlined in chapter 1, in recent years some philosophers have maintained that there is a sense in which at least some kinds of mental disorder can be considered “natural” or “real”. In particular, philosophers of psychiatry have defended the Homeostatic Property Cluster (HPC) view of natural kinds as a way of conceptualizing mental disorders. In addition to maintaining that some psychiatric kinds are natural kinds in accordance with the HPC theory, some philosophers have advanced the further claim that psychiatric categories which do not at present reflect HPC kinds ought to be revised until they do.

In this chapter I contend that philosophers of psychiatry who claim that mental disorders are HPC kinds, and that the HPC theory is a useful model for psychiatric classification, often equivocate between two very different versions of the HPC theory of natural kinds. In particular, I suggest that there is both a strong and weak reading of the HPC theory, and that mental disorders are only HPC kinds in the weak sense. The upshot
of this, I argue, is that philosophers cannot consistently hold both that psychiatric kinds are HPC kinds, and that HPC kinds are useful models for psychiatric classification. I conclude this chapter by clarifying the sense in which I think kinds of mental disorder can be said to be “natural” or “real”.

In chapter 4 I ask whether or not it is plausible to expect that current psychiatric categories will come to reflect natural kinds. When philosophers suggest that psychiatric categories ought to reflect natural kinds they presuppose a picture of the world in which diverse psychopathological phenomena cluster together in the way that other paradigmatic natural kinds do. They also presuppose that by consulting our best theories about the world we can know which clusters are real and which are bogus. In this chapter I argue that the idea that our best theories about the world are a reliable guide to which psychiatric kinds exist and which don’t is historically false. Furthermore, I present evidence for the conclusion that psychopathological phenomena do not cluster together in ways that would be useful for psychiatric classification.

In light of the above considerations I conclude that the normative ideal of natural classification—i.e. classifying real or natural kinds in the sense described in chapter 1—is not appropriate as a general strategy for psychiatry. In the conclusion to this dissertation I propose an alternative way forward. Rather than assuming a priori that psychopathological phenomena cluster together into kinds, a more tractable and theoretically promising approach would be to first explain how particular kinds of experiences and behaviours—e.g., hallucinations, delusions, disorganized thinking, mania, low mood, etc.—are produced.
A Few Notes On Methodology

In addressing the two questions that motivate this dissertation—namely, “Are mental disorders natural kinds?” and “Should psychiatric categories reflect natural kinds”—I decided to proceed from a general conception of natural kinds. The principal reason for this is that, when we ask whether or not a kind is “natural” or “real”, what we really want to know, I think, is whether the properties definitive of a kind really do cluster together in nature, or whether human beings grouped them together as a matter of convenience. As we will see in chapter 1, there are many different theories of natural kinds on the philosophical market. Some theories are so inclusive that almost any kind can be considered natural. Others are so restrictive that only kinds of physical particles and chemical elements can be plausibly considered natural. Simply adopting one theory of natural kinds to the exclusion of others, or claiming that there is some fast-and-loose sense in which some kind is “natural”, runs the risk of turning natural kind theorizing into a game of rhetoric, as Hacking (2007a) has convincingly argued.

Throughout this dissertation I principally use depression and schizophrenia as paradigmatic examples of psychiatric kinds (though I briefly touch other kinds of mental disorder as well). There are three reasons for this. First, the most recent version of the DSM, DSM-5, has grown to 947 pages, and includes a total of 541 diagnostic categories, 151 of which are defined by predetermined inclusion-exclusion criteria. In an essay of this scope I simply cannot address each and every one of these categories, nor can I accord proper justice to even several of them. Second, when philosophers claim that mental disorders are natural kinds, they almost exclusively use depression and
schizophrenia as paradigmatic examples of natural kinds of mental disorder, and are often skeptical of other kinds of mental disorder such as *hysteria, neurasthenia, dissociative fugue, multiple personality disorder* (*dissociative identity disorder*), *cluster B personality disorders* (including antisocial, borderline, histrionic, and narcissistic types), and *attention-deficit/hyperactivity disorder* (*ADHD*). In addressing depression and schizophrenia I avoid setting up a “straw man” and make it much more likely that my conclusions generalize to other (more controversial) psychiatric kinds. Finally, along with bipolar disorder, depression and schizophrenia are two of the oldest surviving categories in psychiatry today. As we will see in due course, a central assumption of natural kind realism is that our best theories of the world are our best guides as to which kinds are real and which are bogus. Given that depression and schizophrenia have been the focus of more than two centuries of scientific theorizing, we should expect them to be amongst the most scientifically mature kinds in psychiatry.

My approach to assessing depression and schizophrenia is three-pronged: philosophical, empirical, and historical. In chapter 2 and chapter 3 I discuss the most up-to-date empirical literature on depression and schizophrenia and utilize the philosophical account of natural kinds developed in chapter 1 in favour of the conclusion that depression and schizophrenia are not natural kinds. In the first half of chapter 4 I analyze the concept of ‘schizophrenia’ from a historical perspective for the purpose of arguing that our best theories about the world are not necessarily reliable guides to which psychiatric kinds are real and which aren’t. In the second half of the chapter I present empirical evidence for the conclusion that psychopathological phenomena do not cluster
together in the way that paradigmatic natural kinds do.

Notwithstanding the fact that I consider depression and schizophrenia to be paradigmatic kinds of mental disorder, what I have to say about depression and schizophrenia in what follows should not be taken to apply necessarily to each and every category listed in the DSM. While I do believe that my conclusions generalize to most kinds of mental disorder—or so I will argue—there are no doubt exceptions, some of which I point out in various places.
CHAPTER 1

Natural Kinds and Natural Classification

[T]here are in nature distinctions of Kind...Kinds are Classes between which there is an impassible barrier...The problem is, to find a few definite characters...[or]...marks whereby we may determine on which side of the barrier an object takes its place.


1.1 Introduction

There is widespread disagreement amongst philosophers and scientists with respect to whether or not our current psychiatric categories are natural kinds, and, moreover, whether or not the search for natural kinds is a realistic desideratum for psychiatry. But what is a natural kind? Philosophers often draw a contrast between kinds that exist in nature and kinds that are constructed for our own convenience. But beyond this intuitive contrast there has been considerable debate in philosophy about what constitutes a natural kind. The proliferation of theories of natural kinds has even prompted philosopher Ian Hacking to write: “there are so many radically incompatible theories of natural kinds now in circulation that the concept itself has self-destructed...[and we should] acknowledge that the concept of a natural kind, which began in a promising way and has taught us many things, is now obsolete” (Hacking, 2007a, 205).

And so he concludes: “Some classifications are more natural than others, but *there is no such thing as a natural kind*” (ibid., 203; emphasis in original).¹

According to Hacking, there are two possible ways of responding to the murkiness

---

¹ More precisely, Hacking argues that there is no such thing as a natural kind insofar as there is no “definite, human-independent, class of natural kinds” (ibid., 207). He allows that the expression ‘natural kind’ might be used more moderately to describe those classifications “that strike us as natural, as opposed to those that strike us as conventional” (ibid., 208), without there being one well-defined class of natural kinds.
of the natural kind concept and its idiosyncratic usage amongst philosophers. The first is to abandon the concept altogether. Hacking, for instance, prefers instead Goodman’s (1978) alternative phrase ‘relevant kind’. According to Goodman, we recognize a kind or category because of the role it plays in some theoretical or practical enterprise, not because it corresponds with what he considers to be some nebulous and unknowable “kind in nature”.\(^2\) Another possible response, Hacking suggests, is to embrace pluralism about natural kinds, or what John Dupré has formerly called “promiscuous realism”. According to Dupré (1981, 1993, 2001, 2002), the idea of a natural kind is grounded in judgements of similarity (or “sameness relations”).\(^3\) Given that individual members of a putative kind are similar in many respects (Dupré’s “realism”), there are many similarities (none of which are privileged and all of which may or may not be relevant to various concerns) that can serve to distinguish kinds of things (Dupré’s “promiscuity”), and so there may be many cross-cutting categories that merit the label “natural kind”.\(^4\)

The problem with both skepticism and pluralism about natural kinds, however, is that they merely substitute one problem for another—namely, ambiguity for vagueness. Even if we grant that there is a sense in which all of our categories are relevant kinds, or a broad sense in which many of our categories are natural kinds, surely not all of these categories are ontologically on a par. Chemical elements, chemical compounds, and species are types of naturally occurring objects, whereas postal codes and telephone area

\(^2\) See, especially, Goodman (1978, chapter 1).

\(^3\) Quine (1969, chapter 5), too, argues that the notion of ‘natural kind’ is grounded in judgements of similarity. The difference, however, is that Quine thinks that intuitive or innate judgements of similarity will, as a matter of course, be replaced by (but not wholly superseded by) scientifically sophisticated ones, such that the concept of natural kind will be dispensed with altogether.

\(^4\) Cf. Dupré (1981, 82).
codes are not. Presumably race and class, too, are not kinds of naturally occurring objects, even if there is a sense in which both are “more natural” or “more real” than postal codes and telephone area codes (Cf. Root, 2000). But loosening our definition of ‘natural kind’, or attaching the equally vague adjective ‘relevant’ to ‘kind’ only serves to obscure these differences.

Rather than assenting to pluralism or skepticism about natural kinds, in this chapter I give a rough-and-ready characterization of natural kinds that grants to the skeptic that there is no well-defined class of natural kinds, but also denies that natural kinds are merely collections of individuals who are similar in some respects. In my mind, the central idea running through the history of natural kind theorizing is that natural kinds are, at bottom, collections of individuals who are alike in important respects—i.e. individuals are members of a natural kind in virtue of having in common important properties. And we know that a kind is “natural” or “real” when we discover, rather than stipulate, what these important properties are.

But what is an “important property”? As Putnam (1975a, 239) has pointed out, ‘importance’ is an interest-relative notion, and the appeal to important properties would seem to concede to skeptics and pluralists that our categories are formulated relative to particular human interests, not the world. In the end, believers in natural kinds agree that “important properties” are properties which cluster or “hang together” in virtue of being produced by certain causal mechanisms. In short, natural kinds are clusters of properties which cluster together because they have a common causal structure.

Furthermore, natural kind theorists maintain that we can know which clusters are
real and which are bogus by consulting our best theories about the world. In this sense, the natural kind realist does not, and need not, deny that natural kind categories are “relevant” or “conventional” in the trivial sense in which all categories are constructed relative to certain human interests. She simply maintains that, relative to our theoretical interests (e.g., identifying unique species or a chemically pure substance), we are able to recognize what the important properties of a natural kind are.

The foregoing conception of natural kinds has its roots in eighteenth and nineteenth-century debates about scientific classification. Since John Stuart Mill was the first to translate these debates into natural kind talk, I begin by describing the context in which Mill was working. Next, I outline Mill’s account of natural kinds. And, finally, I compare Mill’s account to contemporary alternatives. Although I say very little about psychiatric kinds in this chapter, in the next chapter I argue that psychiatric kinds are not natural kinds in the sense outlined here.

1.2 The Origin of Natural Kinds

1.2.1 The Assumption of Essences

The terms ‘kind’, ‘natural kind, and ‘real kind’ were not introduced into the philosophical lexicon until the nineteenth century, though the basic contrast between kinds existing in nature and kinds invented by us has deep roots in the history of philosophy and science. Aristotle, for instance, posited that the fundamental objects of science are substances, that is, things with universal natures or essences. The task of scientific classification, for Aristotle, was to construct “real definitions” of things in nature by identifying what it is that makes an individual substance distinct from other
individual substances—its essence. The candidate essence for the species *human being*, for example, was rationality. While human beings were subsumed under the genus *animal* insofar as they possess the powers of sensation and self-movement, rationality was considered to be the “difference” that separates human beings from other animals, thereby giving rise to characteristic human properties such as language, laughter, and so on. On this basis all things in nature could be classified in an absolute hierarchy according to the method of genus and difference, such that each thing had its proper place in nature.\(^5\) “Pseudo-definitions”, on the other hand, could be invented *ad nauseam* either by predicating of individuals something which is not unique to them (e.g., the definition ‘a father is a male animal with offspring’ makes it such that the category ‘father’ cross-cuts all of the different species of animal, rather than identifying each species uniquely), or by predicating of individuals something which *is* unique to them but which fails to pick out their principal attribute or essence (e.g., Plato’s famous conjecture that a human being is ‘a featherless biped with flat nails’).\(^6\)

A category, then, can be said to be “real” or “natural”, in the Aristotelean sense, just in case it collects together individuals who do indeed share an essence, such that this essence is jointly necessary and sufficient for determining membership in a particular *kind* (or “species”, to use the parlance of the time). Moreover, a properly natural

---

\(^5\) This is Aristotle’s *scala naturae* (“ladder of being”). For Aristotle, all of life could be organized in an absolute hierarchy such that *living things* and *non-living things* are species of *substance*; *animals* and *vegetables* are species of *living things*; and *human beings* are species of animal, and so on. See Ayer (1981). In the medieval and early modern period the ladder of being, or “Great Chain of Being”, was seen as a God-given ordering, with God positioned at the apex followed by angels, demons (i.e. fallen angels), human beings, animals, plants, and minerals. Most importantly, as we will see, objects in the hierarchy were conceptualized as having fixed essences (and consequently, a fixed moral status), such that an individual thing could not, as it were, move up or down the ladder. See Lovejoy (1936/2010).

\(^6\) Aristotle also termed the latter kind of definition a “nominal definition”. See Ayers (1981) for an overview of Aristotle’s approach to scientific classification and explanation.
classification system—and indeed, a complete science—both “cuts nature at its joints” (to use an old Platonic phrase) by providing a comprehensive set of real definitions and reflects the true order of nature by organizing each and every species of thing into an absolute hierarchy according to both their species essence and their genus essence (see note 5 for further discussion).

With the new mechanistic philosophy of the seventeenth century—the view, roughly, that all material objects are explicable mechanically in terms of their fundamental material (as opposed to formal or functional) structure and the common laws of mechanics and motion that govern them—Aristotle’s doctrine of universal natures or substantial forms was in the process of being overturned. What was not lost, however, was the view that the true objects of science are essences. For instance, Descartes and Boyle maintained that they knew the essence of matter, whether it be extension in the case of Descartes, or corpuscularianism (solid particles clashing in a void) in the case of Boyle. Locke objected, arguing in Book III of his Essay that, even if things did have a material essence, this essence could not be known, and so essences are of no help in classifying things into kinds. As we will see in due course, it is the idea of a material essence—and the presumption that we can know a posteriori what these essences are—that now forms the backbone of “New Essentialist” theories of natural kinds.

1.2.2 The Problem of Natural Groups

Essentialism was the predominant assumption in eighteenth century biological classification, even if there was no clear consensus on what the essence of a plant or animal consisted of. Linnaeus, for instance, held that the point of natural classification
was to abstract away from the messiness of nature so as to reveal the perfect orderliness of God’s creation—to recover, as it were, the perfect Adamic language given to Adam by God to enable Adam to correctly name the animals. Linnaeus’ prescription for a proper natural classification was premised on two things. First, metaphysically speaking, species and genera were considered by Linnaeus to be eternally fixed, and any observed intra-specific differences or variations were seen as mere imperfections of God’s orderly creation, rather than observations to be accounted for. Second, epistemologically speaking, for Linnaeus nature itself reveals what the true genera and species are, not the naturalist. Darwin himself would later summarize Linnaeus’ view succinctly: “the characters do not make the genus, but…the genus gives the characters” (Darwin, 1859/2003, 432). In other words, according to Linnaeus each and every plant and animal reveals its own essential properties and thereby “speaks” its own name.

In addition to essence, eighteenth century taxonomists carried over from Aristotle (or at least the Christianized version of Aristotle from medieval scholasticism) the assumption that living things ought to be classified in an absolute hierarchy. For instance, Linnaeus classified living things according to the method of genus and difference. Following this method, *animal* would be a genus of *mammal*, but a species for *living

---

7 As he puts it: “all genera are natural, and have been such since the beginning of time” (Linnaeus, 1751, 100).

8 There is some debate about whether or not Aristotle was actually interested in developing an absolute hierarchy of species, but I gloss over these interpretive difficulties here. See Lloyd (1961) and Pellegrin (1986) for further discussion. What is clear is that the “great chain of being” conception of nature dominated thought well into the nineteenth century in spite of emerging evidence from palaeontology and geology pointing away from the fixity of species and toward organic evolution (or “the transmutation of species”). See Larson (2006).
thing; feline would be a species of mammal but a genus of tiger, and so on. Thus a whole nested hierarchy becomes possible. Species are grouped into genera, genera into families, families into classes, classes into orders, and so forth.

This is when a series of problems emerged. Suppose a taxonomist is attempting to arrange individual things into groups. First, two descriptive questions arise: (1) in which ways do the individuals in question resemble each other? and (2) which method of sorting best represents the ways in which these individuals resemble each other? Linnaeus (1735, 1753), for instance, introduced the “sexual system” of classification for the purpose of sorting plants. Following this system, plants were classified according to the number of stamens and pistils (the male and female sexual organs of a plant, respectively) in the flowers of plant. The number and arrangement of stamens gave rise to 24 classes; within each class a plant was then assigned an order based on the number of its pistils.

Linnaeus’ system was generally recognized as ingenious and expedient. Yet, a further, ontological, question emerged in response to Linnaeus’ classification of both plants and animals: (3) are Linnaean genera and species “natural” or “artificial”? Or, to put it in contemporary terms, are Linnaean kinds real or are they convenient fictions? Why, for instance, suppose that the sexual organs of a plant are the resemblances or similarities that, as nature would have it, determine whether an individual belongs to one species and not another?

To rephrase the latter question more generally: (4) which resemblances or

---

9 For instance, Linnaeus (1758) gave tiger the name felis tigris, felis being the generic name followed by the specific epithet tigris to distinguish tigers from other felines or felidae. This came to be known as binomial nomenclature. Prior to Linnaeus plants and animals were typically classified using unwieldy phrases such as ‘annual, much-branched Physalis, with strongly-angled, glabrous branches and leaves with sawtoothed edges’ The Linnaean binomial equivalent for the same plant is Physalis angulata.
similarities are the ones that, as nature would have it, determine whether an individual belongs to one species or kind and not another? That is, which resemblances matter and which don’t? In Linnaeus’ case the inability to answer this latter question led many of his contemporaries to dub his system of classification artificial as opposed to natural. With respect to plants, they thought the singling out of one feature, sexual organs, as an organizing feature of classification an arbitrary choice. Michel Adanson (1763), for instance, identified that there are upwards of 65 different ways that plants resemble each other, and a proper natural classification scheme would take as many of these resemblances as possible into account. By “artificial” these authors did not mean that species and genera were arbitrary in the way that the motley collection of things on my desk might be. Sameness of sexual organs is a natural similarity. The point, however, was that plants resemble each other in many ways, and the privileging of one similarity to the exclusion of others is an arbitrary choice. Artificial classes, then, were seen as useful classes invented by botanists, as opposed to classes which accurately represented the natural order of things.

1.2.3 Beyond Resemblances and Essences: Postulating Historical Groups

In spite of his critics, Linnaeus was resolute in maintaining that beyond our imperfect classificatory schemes are real genera and species. This view was shared by Darwin. Nevertheless, while Darwin retains the hierarchical ordering of living things, he revolutionized biological taxonomy by proposing a way in which biological classification can move beyond mere resemblances in the absence of stipulative essences. As is well known, this solution came almost a century after Linnaeus when Darwin postulated:
According to Darwin, objects can, in principle, be classed in innumerable ways. Previous naturalists, such as Linnaeus and Adanson, respectively, tended to classify objects “artificially by single characters, or more naturally by a number of characters” (ibid., 431); sometimes things were classed together for the purpose of making generalizations about the members so classed (ibid, 432, 437). Yet, each and every one of these strategies failed to identify “the hidden bond” (ibid., 437) that would serve to explain why similarities between members of a species obtain. For Darwin, individuals resemble each other in the ways that they do because they are part of a common lineage. Genealogical succession or descent, then, is the true natural organizing principle for classifying living things.

1.2.4. Summary: Natural Kinds of Species

The eighteenth and nineteenth century debate about natural classification that culminated in Darwin’s The Origin of Species is noteworthy in a number of respects. First, the debate itself presupposes both that it is a fact about nature that there are kinds of species, and that the similarities and differences among individuals in virtue of which they are divided into kinds can serve as a basis for classification. In short, it presupposes that there are natural kinds of species and that they are knowable a posteriori. Second,
concerning the classification of living things, it would seem that Darwin’s solution to (4) required an answer to a further question: (5) how are individuals in nature related—i.e. what kinds of things are they? Darwin’s postulation that species are historical entities provides an answer, and, as far as classification is concerned, a natural organizing principle whereby some headway can be made on the question of which resemblances matter and which don’t. Rather than sorting living things by means of arbitrarily choosing one resemblance to the exclusion of others—or attempting the Sisyphean task of taking into account all of the ways in which individuals resemble each other—we consider those “characters…which have been inherited from a common parent”, since the relationship to a common ancestor is the reason why individual members of a species resemble each other in the ways that they do. What makes a category of species “real” or “natural”, then, is not the sharing of some eternal essence, but those properties which correctly reflect its place in the tree of life.

Be that as it may, Darwin’s solution for the classification of living things does not solve the problem of classification writ large. While life is genealogical and thereby can be classified in an hierarchical fashion, this still leaves wide open the question of how other aspects of nature are ordered. For instance, Mendeleev’s periodic table of elements broke with the long-standing tradition in which attempts were made to organize all kinds of inanimate stuff in a tree-like fashion. Minerals, too, are no longer organized hierarchically.\(^\text{10}\) Furthermore, although the principles used to classify species of plant and

\(^{10}\text{Mineralogical specimens are now classified according to both their chemical make-up and their crystalloid shape. In the eighteenth century and prior it was common to extend the hierarchical approach used for classifying plants and animals to the classification of minerals and other inanimate substances.}\)
animal were at one time extended to the classification of mental disorders\textsuperscript{11}, there have been many candidates; and, indeed, as we will see, psychiatric classification is still awaiting its own natural organizing principle.

1.3 Natural Kinds

Mill was the first to translate eighteenth and nineteenth-century debates about scientific classification into natural kind talk. Like Darwin and Linnaeus—and unlike his contemporary William Whewell—Mill thought it plausible that some of our categories reflect real kinds of animals, plants, chemical elements, and so on.\textsuperscript{12} Mill also recognized, however, that not all of our scientific categories reflect real kinds. Many of our scientific categories are the subjects of open-ended investigation, and will surely be displaced in favour of new ones. And, indeed, in ‘Of Classification as Subsidiary to Induction’ in his \textit{Logic}, Mill even seems to suggest that the goal of scientific classification is not necessarily to discover and then classify real kinds. Real kinds make up only a portion of our scientific classification schemes, and there are other legitimate uses to which classification can be put. In light of this, the important question regarding natural kinds, for Mill, is not simply, “What is a natural kind”\textsuperscript{?}\textsuperscript{13}—but, given that there are natural kinds, and that they are knowable only \textit{a posteriori}, “How do we know which of our scientific categories are real kinds and which are not?” I describe Mill’s answer to this question in what follows.

1.3.1 Mill on Artificial Classes, Practical Classes, and Natural Groups

\textsuperscript{11} See Foucault (1972/2006), Part II, Chapter 1.

\textsuperscript{12} Mill uses the phrase ‘real Kind’ rather than ‘natural kind.’ I use these terms interchangeably throughout this dissertation.
Mill’s first gesture at a description of real kinds can be found in the section ‘Kinds have a real existence in nature’ of *Logic*. Here Mill points out that there is “a very remarkable diversity in this respect between some classes and others” (*Logic* I vii § 4). The class of white things, for instance, “are not distinguished by any common properties except whiteness; or if they are, it is only by such as are in some way dependent on, or connected with, whiteness” (ibid.). According to Mill, members of the class of white things do not share natural properties other than those which were used to mark the kind (namely, the property white). It is an artificial class. In contrast, members of classes such as *animal*, *plant*, *sulphur*, and *phosphorus* have in common properties in addition to those which are used to mark the kind. As he states: “a hundred generations have not exhausted the common properties of animals or of plants, of sulphur or of phosphorus; nor do we suppose them to be exhaustible, but proceed to new observations and experiments, in the full confidence of discovering new properties which were by no means implied in those we previously knew” (ibid.). The latter, for Mill, are real kinds.

But there is more to the story. Mill does not presuppose that all of our current categories of plants and animals and elements are real kinds. Indeed, Mill acknowledges that there are many uses to which classification can be put, and not all of these uses will be real kinds.

---

13 Indeed, Mill even seems to suggest that we will never discover all of the properties of a real kind, since the common properties of real kinds are inexhaustible. In a recent treatment of natural kinds, Khalidi (2013, 49-52) makes much of Mill’s notion of ‘inexhaustibility’, suggesting that it is an unnecessarily strict condition that Mill places on kindhood. After all, for any kind $K$ it impossible to say definitively whether or not the properties of some kind will be inexhaustible or not, and there is in principle no reason to suppose that there will not be instances in which we discover all of the properties of a kind at the outset of our investigation. While there are certainly instances in which Mill seems to suggest that inexhaustibility (he also uses the terms ‘indefiniteness’ and ‘indeterminacy’) is a condition of kindhood (see, especially, *Logic* I vii § 4), he also says that the properties of a kind are inexhaustible in that “Our knowledge of the properties of a Kind is never complete” (*Logic* IV vii § 4). In other words, Mill seems to be making the banal observation that it will always be possible that we will discover new properties of a kind that we did not at first consider. For this reason, I do not treat the condition of inexhaustibility seriously in what follows.
furnish us with real kinds. For instance, in ‘Of Classification as Subsidiary to Induction’ Mill observes that some classes are formed merely with a “particular end in view” (Logic IV vii § 2). Call these “practical classes”. With respect to these classes, the properties or characters that we attend to for the purpose of classification are the ones that merely serve certain parochial practical ends. A farmer who grows both flowers and garlic, and is interested in profiting from the sale of these two crops, does not consider garlic a lily, even if garlic is a lily (i.e. subsumed under the category Liliaceae) according to the botanist. According to the farmer and her interests, garlic and lilies are two very different kinds of thing. Garlic is a root crop and requires different care than a lily. Garlic is planted and harvested differently than a lily; it is dried, cleaned, and eventually used to flavour a dish. The implication of this, with respect to classification, is that the putative kind garlic may be classified in at least two different ways, both as a lily and not as a lily. As Mill makes clear, the varied uses to which we put classification makes it such that “the same objects…may admit with propriety of several different classifications” (Logic IV vii § 2).

Be that as it may, it would be a mistake to think that a botanist’s classification scheme provides us with real kinds, whereas a farmer’s does not. As you will recall, Linnaeus’ critics dubbed his classification of plants “artificial” on the basis that the choice of sexual organs as an organizing feature of classification was an arbitrary one. Mill agrees, but for slightly different reasons. For Mill, the choice of sexual organs is arbitrary not merely because it is one property among many; rather, it is arbitrary (and, he adds, “of

---

14 Mill provides a similar example: “A farmer does not divide plants, like a botanist, into dicotyledonous and monocotyledonous, but into useful plants and weeds” (Logic IV vii § 2).
little use”) because classifying in this way does not allow us to say anything more about the thing that we are classifying.\(^\text{15}\)

For Mill, then, there are many ways to divide the world, some approaches being useful and some less so. Most importantly, however, not all uses are on a par. When our goal is to learn about the world, and not merely to construct categories that serve our parochial practical ends, we search for natural groups. As Mill writes:

But when we are studying objects not for any special practical end, but for the sake of extending our knowledge of the whole of our properties and relations, we must consider as the most important attributes, those which contribute most, either by themselves or by their effects, to render the things like one another, and unlike other things; which give to the class composed of them the most marked individuality; which fill, as it were, the largest space in their existence, and would impress the attention of a spectator who knew all their properties but was not specially interested in any. Classes formed on this principal may be called, in a more emphatic manner than any others, natural groups (Logic IV vii § 2; emphasis mine).

In order to classify the world as it really is we must attend to “the most important properties, those which contribute most, either by themselves or by their effects, to render the things like one another.”\(^\text{16}\) It would seem that, for Mill, we know that a kind is “natural” or “real” when we recognize or discover those properties which make a thing the kind of thing that it is—the similarities \textit{in nature} in virtue of which things are, as it were, divided into kinds.

But how do we know which properties “contribute most” to rendering individuals like one another? At this point Mill seems to suggest that the importance of a property

\(^{15}\) As Mill puts it: “to think of [plants] in that manner is of little use, since we seldom have anything to affirm in common of the plants which have a given number of stamens and pistils” (Logic IV vii § 2).

\(^{16}\) Regarding natural classification, Mill says: “the test of its scientific character is the number and importance of the properties which can be asserted in common of all objects included in a group” (Logic IV vii § 2).
consists in its aptness for making inductive generalizations. For instance, Mill writes:

*The ends of scientific classification are best answered, when the objects are formed into groups respecting which a greater number of general propositions can be made...the properties, therefore, according to which objects are classified, should, if possible, be those which are causes of many other properties: or at any rate, which are sure marks of them. Causes are preferable, both as being the surest and most direct of marks, and as being themselves the properties on which it is of most use that our attention should be strongly fixed (Logic IV vii § 2; emphasis mine).*

According to Mill, the important properties of a kind ensure the manifestation of other properties of the kind, either by causing them or being predictive (i.e. “sure marks”) of them. Given this, it would seem that, for Mill, “important properties” are projectable properties\(^\text{17}\)—i.e. properties which point reliably to other properties.

But beyond the fact that natural groups allow us to make useful generalizations, why does Mill insist that classifying natural groups in the sense outlined above comes closer to tracking real divisions in nature? There is a clear rationale for this. If we are able to make reliable inductive inferences about individual members of a kind, then there must be something in nature making this possible.\(^\text{18}\) Hence, the projectability of a category is an epistemic mark of its naturalness (viz. realness).

1.3.2 From Natural Groups to Natural Kinds

But projectability is not the whole story about real kinds. For Mill, natural groups come closer to tracking real kinds in nature than artificial classes and practical classes, but not all natural groups are real kinds (see, especially, *Logic* IV vii § 4). You will recall

---

\(^{17}\) Typically, projectability is used to characterize terms or predicates which figure in reliable inductive inferences (Cf. Goodman, 1983). In this dissertation I will often say that kinds, properties, categories, or concepts are projectable, which is just to say that the terms or predicates which refer to them are.

\(^{18}\) As Mill puts it: “If the class did not possess any characters in common, what general assertions would be possible respecting it?” (Logic IV vii § 4).
that, in the previous section, I said that for Darwin the question of (4) which resemblances or similarities are the ones that, as nature would have it, determine whether an individual belongs to one species or kind and not another turned on the question: (5) how are individuals in nature related—i.e. what kinds of things are they? Mill appears to share the same intuition. Although Mill was writing in pre-Darwinian days and thus was not privy to Darwin’s solution to (4) in the context of biological species, he takes his cue from chemistry. Of elementary chemistry Mill writes:

The substances, whether simple or compound, with which chemistry is conversant, are Kinds, and, as such, the properties which distinguish each of them from the rest are innumerable; but in the case of compound substances (the simple ones are not numerous enough to require a systematic nomenclature), there is one property, the chemical composition, which is of itself sufficient to distinguish the Kind; and is (with certain reservations not yet thoroughly understood) a sure mark of all the other properties of the compound (Logic IV vii § 5; emphasis mine).

In Mill’s view, there are many properties associated with a particular chemical compound. Nevertheless, with respect to classifying a real kind of chemical compound, one property matters more than any other: chemical composition.

But why suppose that chemical composition is the property that matters? It seems Mill is appealing here to Antoine Laurent-Lavoisier’s system of classifying chemical compounds, which first appears in Method de nomenclature chimique\(^\text{19}\) in 1787. Implied in this method of classification was the view that chemical compounds are composed of simple chemical substances linked by certain chemical affinities—e.g., the chemical

\(^{19}\) Co-authored with Guyton de Morveau, Berthollet, and Foucroy.
compound H₂O being composed of hydrogen and oxygen molecules in a ratio of 2:1.²⁰ To be sure, Mill is somewhat hesitant in his claim about chemical composition, acknowledging its theory-dependence with the qualification “with certain reservations not yet thoroughly understood”. Such reservations notwithstanding, Mill’s central point is that, given that the arrangement of simple chemical substances does make a chemical compound the kind of thing that it is—i.e. given that Lavosier is right that the structural properties of chemical compounds “render the things like one another”—chemical composition is the property that matters.

According to Mill, then, natural kinds are at bottom collections of individuals in nature who have in common important properties, and we can be reasonably certain that a kind is “natural” or “real” when we discover rather than stipulate what these properties are. Moreover, ‘important’ here should be taken to mean “theoretically important” in the sense that the properties indicative of a real kind stand in an explanatory relationship to one another. For example, while it was once commonplace for chemists to use color as a basis for distinguishing between chemical kinds (the motivation being that substances which were otherwise alike in their observable properties often differed in color), color turned out to be an unreliable indicator of real chemical kinds. Why? Because, as it happens, there is no fundamental relation between color and chemical constitution. And, if chemical constitution does indeed explain why it is that chemical kinds are similar in the ways that they are, then color is not a reliable indicator of a real kind of chemical

²⁰ This example is slightly anachronistic, as Lavoisier thought (at least for a short time) that the gases hydrogen and oxygen were themselves compounds of heat and another substance. See Lefèvre (2012, 2018) for an account of Lavoisier’s method of classification according to composition.
element or compound. In short, color is not an important property.

The appeal to important properties is crucial for Mill. Even though projectable categories are “more natural” (inasmuch as individual members of these kinds have many properties in common) than arbitrary categories and mere practical categories, this does not entail that they are real kinds. After all, presumably many of our incomplete scientific categories and extra-scientific (or folk) categories are projectable. A farmer, for instance, can make generalizations about garlic, even if garlic is not a real kind according to our current best theories of the world. What is needed then—in order to individuate a real kind—is the discovery of those properties in nature that unify individual members of a kind.

1.3.3 Summary: Millian Kinds

The eighteenth and nineteenth century debate about classification was premised on a distinction between natural and artificial classes. Although Darwin proposed a way of moving from artificial to natural classes in the context of biology, he did not explicitly say what it is that makes a kind “natural” or “real” more generally. This was Mill’s contribution. Since, in my view, Hacking’s “many radically incompatible theories of natural kinds now in circulation” can be read as different ways of responding to Mill’s account, it will be useful to first summarize the theses to which Mill appears committed:

I. Mind-independence: For Mill, like Darwin, it is a fact about nature (independent of how we happen to classify the world) that there are kinds of things, substances, organisms and so forth, and the similarities among individuals in virtue of which they are divided into kinds (viz. important properties) “are made by nature” (Logic IV vii § 4).
II. Discoverability: It is possible to discover or recognize a posteriori what the important properties of a kind are. Hence, it is possible to discover and classify natural kinds.

III. Mind-dependence: The recognition of the important properties of a kind “as grounds for classification and of naming, is… the act of man” (Logic IV vii § 4). Natural kind categories are “mind-dependent” in the sense that the only way to know whether a property is important or not is to appeal to our best theories about the world. Insofar as our theories about a kind are correct, that kind is “natural” or “real”.

IV. Definability: A rough-and-ready characterization of “natural kind” is that it is a collection of individuals in nature who have in common important properties. As we have seen, Mill tells us what he thinks the important properties of chemical compounds are, but he does not say what kinds of properties are important more generally. As we will see in the next section, the attempt to define ‘natural kind’ more precisely has been the central focus of natural kind theorists in the twentieth and twenty-first century.

V. Utility: Since a category or kind-concept might be put to many different uses, the utility of a kind inevitably varies across time and across space. This, however, does not entail that all uses are on a par. When our goal is to learn about the world, and not merely to construct categories that serve our parochial practical ends, our categories will tend toward real kinds in nature.

VI. Classification: All classification schemes are “conventional” in the trivial sense in which each and every category or kind-concept is formulated relative to certain human interests. Be that as it may, relative to certain interests (e.g., identifying a unique species or a chemically pure substance), we are able to recognize what the important properties of
a natural kind are. Hence, a natural classification scheme is possible. Indeed, a natural classification scheme is desirable inasmuch as our goal as scientists is to develop categories which reflect the natural structure of the world.

1.4 Theories of Natural Kinds

Most believers in natural kinds agree with Mill’s conceptualization of natural kinds as collections of individuals who share important properties (Cf. Boyd, 1999a, 143-144; Kornblith, 1993, 36-40; Khalidi, 2013, 44, 54; Putnam, 1975a, 239)\(^\text{21}\). As far as the distinction between kinds that exist in nature and kinds constructed for our own convenience goes, this seems to me a reasonable requirement. Consider, for instance, the kind gold, an oft-used exemplar of a natural kind. As Kornblith (1993, 37) has noted, two samples of gold need not share all of their properties (contrary to what Locke has contended\(^\text{22}\)). Two samples of gold may differ in size, shape, weight, temperature, and so on. If we were to require that members of the same natural kind have all of their properties in common, then this would result in each sample of gold being a kind unto itself. If, on the other hand, we were to simply loosen this requirement so that members of the same natural kind need only have some of their properties in common (irrespective of whether or not they are important), then this would fail to exclude artificial kinds such as the class of white things. Indeed, this condition would fail to exclude any class of individuals that can be cobbled together on the basis that individuals have one or more of

\(^{21}\)Not all philosophers use the phrase ‘important property’. Other oft-used terms are ‘salient property’, ‘real property’, ‘genuine property’, ‘real similarity’, ‘homeostatic property’, and so forth. In each and every case, however, these phrases are intended to convey the idea that the properties associated with natural kinds ought to be the ones that make it the kind of thing that it is. More on this below.

\(^{22}\)For instance, Locke says: “[I]t is...impossible, that two things, partaking exactly of the same real Essence, should have different Properties...” (III, iii, 17).
their properties in common.

Clearly, in order to make the case that gold is a real kind, and that white things are not, we need to presuppose that natural kind categories are defined by more than mere similarity. Not all philosophers conceive of natural kinds in exactly this way, however. In the past century or so theories of natural kinds have deviated from Mill in one of two ways. Some philosophers think of natural kinds as vague categories which collect together individuals who are loosely similar, effectively excluding the condition that members of the same natural kind have in common important properties. Others add further conditions on natural kinds. Before proceeding with our discussion of the natural kind status of mental disorders in the next chapter, it will be useful to distinguish both of these positions from Mill’s.

1.4.1 Cheap Similarity: Natural Kinds as Folk Categories

Far from Mill’s view that natural kinds are realities in nature that science is in the business of discovering, both Russell (1948) and Quine (1969) contend that natural kinds have their origin in pre-scientific or “folk” categories. Russell (ibid, 335), for instance, concludes that we posit kinds such as dogs and cats in order to establish such common-sense inductions as “dogs bark” and “cats meow.” The assumption that dogs and cats are natural kinds, however, is merely a stopgap on the road to the more fundamental categories that science aims to uncover.

Quine’s (1969) analysis of natural kinds is similar. According to Quine, at the beginning stages of inquiry we place things that we find similar in categories that we deem to be natural kinds. Initially, such categories are grounded on intuitive or innate
judgements of similarity, or what Quine calls an “innate quality space” (ibid., 129). For instance, we might first organize things into kinds based on similarities in color—Mill’s class of white things, if you will. Nevertheless, as science matures, intuitive judgments of similarity are replaced with scientifically sophisticated ones, such that natural kinds based on “superficial” or “spurious” similarities such as color are replaced with categories “that have been found by scientific experience to facilitate induction” (ibid). Natural kind categories, on this account, do not reflect realities in nature; rather, they are merely useful crutches that enable us to accomplish fairly limited goals.

Dupré’s (1981, 1993, 2001, 2002) pluralism or promiscuous realism deviates slightly from the positions which I have just outlined. Dupré, like Quine, sees the concept of a natural kind as being grounded in judgements of similarity. Unlike Quine, however, Dupré does not think that folk categories will, as a matter of course, be superseded by scientific categories. His reasoning for this is fairly straightforward. Since there are many ways in which individuals in nature are similar, there are many ways in which one might legitimately divide the world. It all depends on what one is trying to do.

Take, for instance, the classification of whales. At one time biologists classified whales as kinds of fish based on shared habit and similarities in other gross phenotypical features. At the present time, however, whales are no longer classified as fish on the basis that they are descendants of a distinct evolutionary line. In fact, the lack of gills, the ability to give live birth, in addition to other typical mammalian properties, also distinguishes them phenotypically from other organisms that we tend to label ‘fish’. For Dupré, however, the purposes for which biologists classify organisms does not
necessarily coincide with that of the population at large. Even if biologists no longer consider whales as kinds of fish, the classification of whales as fish might be warranted in certain folk biological contexts, such as commerce. After all, given that whales are similar to fish in some respects, and given that the similarities that we attend to for the purpose of constructing a kind are relative to the uses to which they are put, whales qua fish also merit the label “natural kind”.

In the philosophical literature on natural kinds, Russell, Quine, and Dupré are most often interpreted as proposing distinct theories of natural kinds (see, for example, Cooper, 2004, 74-75, 2007, chapter 4; Hacking, 1991, 112-113, 2007, 226-227; Khalidi, 2013, 56-65). Their respective attitudes toward natural kinds, however, are completely at odds with the approach adopted by Mill; and, relative to Mill, it is somewhat misleading to label these as theories of natural kinds. After all, according to the foregoing authors “natural kinds” are merely rough-and-ready categories which collect together individuals who are loosely similar. Conversely, for Mill, natural kinds are not categories at all, but kinds in nature—i.e. individuals in nature who are unified by real similarities.

This is a crucial difference. According to the former reading of natural kinds a category can be said to be natural if it collects together individuals who are similar in some way. Hence, chemical kinds which are distinguished on the basis of differences in color are “natural” in this sense. On Mill’s view, however, color is not sufficient to distinguish real chemical kinds from bogus ones. The reason for this, as we have seen, is that there is no fundamental relation between color and chemical constitution, and chemical constitution (so far as we know) explains the other properties of a chemical kind.
—e.g., flammability, toxicity, acidity, reactivity, heat of combustion, oxidation states, etc. In short, color is not an important property, and categories which are not constructed on the basis of important properties are not natural kinds.

In light of the fact that natural kinds as Russell, Quine, and Dupré understand them are characterized by mere similarity, and mere similarity is not sufficient for establishing that something is a real kind, it would be more fruitful to think of them as either practical or artificial kinds in the sense outlined in section 1.3. I will make this case more forcefully in chapter 3. For now, it is important to recognize that the phrase ‘natural kind’ means something very different for Russell, Quine, and Dupré than it does for Mill.

1.4.2 Essentialism and Non-essentialism about Natural Kinds

If our purpose in positing natural kinds is to distinguish between kinds in nature and kinds constructed for our own convenience, then it is clear that members of natural kinds must be alike in important respects. The problem, however, is to say in just which respects members of the same natural kind must be alike. Even if we grant Mill’s point that the important properties of a natural kind are the ones that make a thing the kind of thing that it is, this does not appear to be all that informative. After all, what \textit{kinds} of properties make a thing the kind of thing that it is? And must each and every one of these properties be possessed by each and every member of a natural kind? That is to say, are individual members of a kind unified in virtue of sharing an essence, or is it sufficient that members of the same kind share enough of their important properties in common? Let’s address the question about essences first before moving on to the question about what kinds of properties ought to be associated with a natural kind.
Some philosophers favour a position, which goes under the name of essentialism, whereby members of the same natural kind must have all of their important properties in common, such that these properties are necessary and sufficient for determining kind membership. Familiar examples of putative essential kinds include kinds of chemical elements or compounds such as *gold* or *water*, respectively. For instance, with respect to chemical elements, some essentialists have contended that the number of protons in the nucleus of an atom is jointly necessary and sufficient for defining a particular chemical element (Ellis, 2001). That is to say, a particular arrangement of subatomic particulars is its real essence. And, given that each natural kind has a real essence, it will always be clear whether or not an individual is a member of a particular kind.

Most philosophers have argued against essentialism on the basis that many of our actual scientific categories do not in fact meet these conditions, but that it is also implausible to suppose that such categories are merely invented and have no claim to accurately describing a portion of nature. For instance, in a series of articles philosopher Richard Boyd (Boyd, 1989, 1991, 1999a, 1999b) has contended that, once we leave the world of chemistry for, say, biology, many of our kind-concepts are vague or fuzzy in the sense that members of the same kind share some but not all of their properties in common, and it will not always be clear when an individual is a member of a kind. For instance, there are no properties essential to the kind *tiger* (*Panthera tigris*). Intraspecific differences such as body size, pelage coloration, striping patterns, skull dimensions, craniological details, genetic, and molecular structures make it such that it is unclear whether the kind *tiger* is one kind of thing or many (Mazak, 2008). Yet, insofar as
members of the kind tiger are morphologically or genetically similar, they are plausibly members of the same evolutionary group, and so are distinguishable from other kinds of animals—as well as other species in the same genus *Panthera*, such as lions, leopards, jaguars, and snow leopards.

The essentialist position and the non-essentialist position are often pitted against each other in a never ending dialogue of the deaf. It is not difficult to see why. The non-essentialist response, while plausible, is not a dialectically effective response to essentialism. The fact that many of our *concepts* are vague, fuzzy, and *in some way* connected to nature does not entail that *kinds* in nature (supposing that such kinds exist) are in fact structured in this way. It is perfectly possible that many of our current categories are convenient fictions which will be displaced in favour of new categories which, in the fullness of time, will meet some or all of the conditions of essentialism.\(^\text{23}\)

Consider a hypothetical kind \(K\), which is associated with properties \(P_1, \ldots, P_6\). Suppose, further, that the condition we place on natural kinds is that the individual members of \(K\) need only have some of their important properties in common. In this scenario it will always be possible that there is some subset of individuals of \(K\) which are more fundamentally similar to each other than they are to other individuals of \(K\) (say, by consistently sharing properties \(P_1, \ldots, P_3\) and not properties \(P_4, \ldots, P_6\)). If this is the case, then we have reason to subdivide \(K\) into \(K_1\) and \(K_2\). If individual members of \(K_1\) and \(K_2\) still do not share all of their important properties in common, then there is nothing

\(^{23}\) It also possible, as some essentialists contend (Cf. Ellis, 2001), that only a small set of privileged categories may rightly be considered natural kinds.
from stopping us from revising our kinds further. And so on.

This difficulty, I believe, is one reason why essentialists insist that natural kinds must be defined by essences. Many of our scientific concepts are fuzzy in light of our incomplete knowledge of the world. Yet, if real kinds themselves are vague or fuzzy—i.e. if nature itself does not have clear boundaries—then how could we ever expect to differentiate between inchoate scientific categories and real kinds in nature? In other words, how could we ever know which categories are vague in virtue of our incomplete knowledge of the world and which categories are vague in virtue of the way the world is in fact structured?

Essentialism, of course, provides a ready solution to this problem. A kind is a natural kind just in case each and every one of the properties associated with the kind in question are shared by each and every individual member of the kind. We know that a category is a real kind when it conforms to this definition. Be that as it may, this solution appears *ad hoc*. For one thing, the fact that essentialists place conditions on natural kinds which are convenient for clearly distinguishing between natural kinds and other kinds of categories does not entail that natural kinds are in fact structured in the way that essentialists suppose. It is perfectly possible that many of our scientific categories are indeed natural kinds, and that we simply cannot know with certainty whether they are

---

24 The reverse is true as well. We may find that there is significant overlap between individual members of $K_1$ and $K_2$ such that we are justified in combining two kinds into one.

25 But there are other reasons for positing essences. As has been pointed out elsewhere (Cf. Barker & Kitcher, 2014, 40; Hacking, 2007a, 237), the idea of an essence is also invoked to explicate the idea of a law of nature or natural necessity. To take the simplest case, the fact that we can make law-like generalizations about *hydrogen* is explained by the fact that many of the properties of hydrogen flow from its essence—i.e. the distinctive features of the atoms which make it up, in this case each atom consisting of a single proton and a single electron.
natural kinds, or whether they might be further revised so as to conform more closely to nature.

Furthermore, even if we grant that some natural kinds have essences—e.g., kinds of chemical elements such as gold, or kinds of chemical compounds such as water—this does not entail that all kinds in nature are structured in this way. After all, many complex biological kinds, such as kinds of tiger, do not (as far as we know) have essences. Should we simply conclude that tigers are not natural kinds, allowing the proverbial philosophical tail to wag the scientific dog? This seems unreasonable. Short of appealing to some metaphysical intuition about the nature of reality, we simply cannot know a priori whether or not the kind tiger is a natural kind in virtue of failing to have an essence.

1.4.3 Kinds of (Natural Kind) Properties

From the standpoint of the life sciences the thesis that all natural kinds come equipped with essences would seem to be highly dubious. Nevertheless, even if we cannot conclude that individual members of a real kind are unified by an essence, can we say something more about what kinds of properties make a thing the kind of thing that it is?

As I outlined in the previous section, there are times when Mill seems to suggest that, with respect to natural kinds, certain kinds of properties are more important than

---

26 Indeed, even in the chemical case there are worrying examples. For instance, as Khalidi (2013) argues, isotopes of the same chemical element share the same atomic properties, but may differ in other properties (e.g., melting point) as a function of differences in mass. Isomers and polymers also present difficulties for essentialism. For a recent criticism of essentialism, see Khalidi (2013, chapter 1 and chapter 5). I will not dwell on essentialism at length in this dissertation, as it is generally agreed by philosophers of psychiatry that psychiatric kinds do not have essences.
others (*Logic IV* vii § 2). For instance, when discussing natural groups Mill appeals to projectable properties (including causal properties); and, when discussing real kinds, Mill seems to appeal to the structural properties of a chemical compound.

Be that as it may, Mill never straightforwardly addresses the question of which kinds of properties are associated with a natural kind, leaving us to wonder how we might generalize the example of chemical compounds to other putative natural kinds. Fortunately, other philosophers have since attempted to fill in that which Mill left vague. For instance, in discussing the kind water, Putnam contends that to be considered a member of this kind a liquid must “agree in important physical properties” (1975a, 239) with other liquids of the same kind:

Normally the ‘important’ properties . . . are the ones that are structurally important: the ones that specify what the liquid . . . is ultimately made out of . . . and how they are arranged or combined to produce the superficial characteristics. From this point of view the characteristic of a typical bit of water is consisting of H₂O (ibid., emphasis mine).

Similar to Mill, Putnam believes that, with respect to the kind water, chemical composition is the property that matters. The only difference between Putnam and Mill is that Putnam gives us a clearer explanation as to why it matters. Rather than simply saying, as Mill does, that chemical composition is important because it is “a sure mark of all the other properties of the compound” or that “it renders the things like one another,

---

27 A likely explanation for this is that Mill does not think that natural kinds are held together by properties at all, be they causal or essential. Rather, he thinks that natural kinds are regularities in nature that obtain between different things at the same time. Such regularities are held together by the laws of causation, specifically laws of coexistence which have the same reality as diachronic causal laws. See, especially, Magnus (2015) for an illustration of differences between Mill and other philosophers on this score. I will ignore this complication in what follows, since it is generally agreed that many kinds—particularly kinds in the social sciences—are not held together by laws, and there is no *a priori* reason to suppose that causal properties cannot play the same role that laws play in Mill’s account.
and unlike other things” Putnam makes clear to us that the particular arrangement of hydrogen and oxygen molecules is “important” because it is structured in such a way as to produce the other “superficial” properties of water, such as being colorless, odourless, tasteless, and transparent.

It is perhaps tempting to conclude from this that the particular arrangement or combination of hydrogen and oxygen molecules is the essence from which the other properties of water flow. To be sure, there are instances in which Putnam does speak in this manner. Nevertheless, Putnam’s more general point is that, with respect to a natural kind, there is a “hidden structure” (Putnam, 1975b, 235, 241) which produces the other superficial or stereotypical properties of a kind, and it is this hidden structure which generally “determines what it is to be a member of that natural kind” (ibid., 241).

The metaphor of a hidden structure is instructive, though stretched quite thin if taken literally. After all, what is the “structure” of a kind of species or a kind of disease? Consider the latter. One might plausibly point to etiology, since the etiology of a particular disease produces its characteristic or stereotypical symptoms. But most diseases arguably involve the interplay of many etiological and constitutive factors, and are perhaps best explained not by appeal to a defined structure but rather complex networks or patterns of causal mechanisms, some of which conspire to produce a disease in one instance but not in another (see, especially, Thagaard, 1998). For example, a duodenal ulcer might be produced by a Helicobacter pylori infection, but many people with this

---

28 For instance, when putting forth the idea that members of the same natural kind have “the same general hidden structure” he adds (in both parentheses and scare quotes) “the same ‘essence’, so to speak”. See ibid., 235. Putnam’s account of natural kinds is often lumped together with Kripke’s (1980), but, as has been argued elsewhere (Hacking, 2007a, 227-228; 2007b), Putnam, unlike Kripke, does not conceive of natural kinds as generally being defined by essences.
infection do not get ulcers. Moreover, some people without a *H. pylori* infection do get duodenal ulcers, a hypothesized effect of either genetic predisposition, smoking and stress, the heavy use of aspirin, or some combination of the aforementioned.

Of course, one might argue that diseases are not natural kinds on the basis that they don’t have the same sort of microphysical structural make-up as chemical compounds. But, as with the essentialist thesis, this conclusion begs the question in favour of microphysical properties. The fact that the kinds that we happen to think are real are defined by microphysical properties does not entail that all natural kinds must be defined in this way.

At this point we would seem to be right back where we started. If the important properties of a natural kind are not necessarily microstructural properties, then, again, what *kinds* of properties are associated with a natural kind? It is commonplace for both essentialists and non-essentialists alike to put forward the general desideratum that the properties associated with natural kinds ought not be superficial properties, but rather those properties which are responsible for the superficial properties of a kind. But what kinds of properties are these? As many philosophers have argued (Boyd, 1989, 1991, 1999a, 1999b; Craver, 2009; Khalidi, 2013; Kornblith, 1993) the most plausible candidate would seem to be *causal properties*. After all, if members of some kind consistently share

---

29 Indeed, some essentialists do hold such a position. Cf. Wilkerson, 1988, 41.

30 Another reason for contending that the properties associated with a natural kind ought to be microphysical properties is that microphysical properties are “more fundamental” than macrophysical properties inasmuch as they constitute them. Nevertheless, as has been noted elsewhere (Hacking, 2007a, 205; Khalidi, 2013, 38-40), the properties that philosophers typically associate with natural kinds of chemical elements and compounds—molecules, atoms, protons, neutrons, electrons—are themselves (according to current physical theory) composed of more fundamental particles such as quarks, bosons, and fermions. If some kinds are not natural or real in virtue of being constituted by more fundamental “microphysical” properties, then surely the same logic extends to kinds of chemical elements and compounds.
a cluster of immediately observable or superficial properties, then presumably they do so because a causal relationship obtains between a set of properties whose co-instantiation causes the instantiation of other, more superficial, properties of a kind.

The appeal to causal properties would also seem to make intelligible Mill’s claim that the important properties of a natural kind are those which make a thing the kind of thing that it is, since, presumably, a thing possesses a characteristic set of properties in virtue of the properties which hold them together—namely, the properties which causally sustain or maintain them. The general picture of a natural kind on this account is not a kind of thing with a microphysical structure or essence, but rather a property cluster whereby the presence of some properties tends to favour the presence of others (Cf. Boyd, 1989, 16; Khalidi, 2013, 78).

Craver (2009, 579) and Khalidi (2013, 80) have labelled this latter conceptualization of natural kinds the “simple causal theory” of natural kinds. There are a number of reasons why this conception of natural kinds might be attractive. First, the simple causal theory supplements Mill’s theory of natural kinds by specifying for us what the important properties of a kind generally are. Moreover, while the simple causal theory presupposes that members of the same natural kind do have a shared “causal structure” (inasmuch as different causal properties conspire to produce a shared cluster of superficial or observable properties), it does so in such a way as to not stipulate what the precise structure of a natural kind is, thereby accounting for a wider range of natural kinds that might be found within different scientific disciplines.

---

31 Boyd calls the cluster of properties causally responsible for the other properties of a kind “homeostatic causal mechanism(s)".
Most importantly, however, the simple causal theory can be seen as a middle-way between pluralism and essentialism about natural kinds. Natural kinds, on this account, are not inchoate folk categories defined by spurious or superficial similarities. Nor are they collections of individuals who are defined by some essence. Rather, they are kinds inasmuch as they are collections of individuals who are similar in virtue of the causal properties that make them so. As Khalidi (2013, 74) puts it: “once the existence of the properties within the cluster is understood to spring from certain causal mechanisms, this assures us that the properties have not been associated with each other on artificial grounds, merely as a result of our predilections to group certain properties together”.

1.4.4 Summary: Causal Relations as Hidden Bonds

On the simple causal theory causation plays a major role in the individuation of natural kinds. Nonetheless, one might legitimately ask: why is it that causal properties or relations especially qualify as being indicative of which kinds are or are not real? A fully satisfactory answer to this question is beyond the scope of this dissertation, but two considerations are worth noting. First, if natural kinds are indeed collections of individuals who are fundamentally the same, then there must be some “hidden bond”, as Darwin put it, that would serve to explain why individuals are similar in the ways that they are. The appeal to causality would seem appropriate here, since causality is commonly acknowledged to be, in Hume’s words, the “cement of the universe” (THN Abstract ¶35).

Furthermore, in many instances science itself treats causality as a criterion for the reality of a kind. Consider, for instance, the kind dementia. Prior to the twentieth century
dementia was a broad clinical syndrome associated with global declines in intellectual functioning (e.g., difficulties with memory, abstract reasoning, and comprehension) (Cf. Berrios, 1987). As it happens, however, dementia is not a disease in and of itself. Rather, the term ‘dementia’ is a catch-all for different kinds of diseases which affect cognition, such as frontotemporal dementia, Alzheimer’s disease, Huntington’s disease, and perhaps even schizophrenia (formerly, “dementia praecox”). Theoretically, each of these diseases is distinguished from each other on the basis of causality—most often brain pathology, but also etiology in cases where etiology is known. To cite one example, in 1907 Aloys Alzheimer discovered a patient who did not have obvious cerebral atrophy but nevertheless had microscopic changes in their brain (the patient’s brain was cluttered with abnormal debris known as amyloid plaques and neurofibrillary tangles). A century later these accumulations of protein—more concentrated in particular areas of the brain such as the medial frontal lobe, the inferior and medial parietal lobes, the medial temporal lobe, and the posterior cingulate cortex—are still considered the hallmarks of Alzheimer’s disease.

Supposing that science is in the business of explaining things, and to explain something (e.g., Alzheimer’s disease) is to show how it fits into the causal structure of the world (Craver, 2007; Salmon, 1984), causal relations would seem to play a crucial role in determining which of our categories or kind-concepts ought to be considered real kinds.
But even if causal relations are not the privileged features of reality, they are certainly viable candidates for the role of “important properties” in Mill’s sense of the phrase. Since the kinds I will be talking about in this dissertation are typically explained in terms of causal relations, I will proceed under the assumption that causation plays an important role in the individuation of natural kinds.

1.5 Realism About Natural Kinds

Thus far I have argued that the central idea running through the history of natural kind theorizing is that natural kinds are, at bottom, collections of individuals who have in common important properties. Moreover, “important properties” should be taken to mean properties which cluster or “hang together” in virtue of being produced by causal properties.

The appeal to causal properties is crucial for natural kind realism. In the absence of knowledge about the causal structure of the world, it would be impossible to know whether the properties definitive of a kind do indeed cluster together, or whether these groupings are merely a reflection of our predilections to group certain properties together.

One skeptical worry raised by this account is how can we know, or reasonably believe, that the causal properties we’ve identified are real rather than nominal? The appeal to unobservable causal properties would seem to be a sham, unless we have reason to believe that unobservable causal properties are themselves real and not merely

---

32 Some philosophers, for instance, have contended that fundamental physics is not in the business of describing causal relations at all, but, rather, the structural properties of the whole universe expressed through laws. As I mentioned earlier (see note 27), Mill himself preferred to characterize natural kinds as being held together by laws of coexistence rather than causes or causal relations. Be that as it may, there is no logical inconsistency in maintaining that some natural kinds are held together by laws and some by causal relations.
Realists about natural kinds typically respond to this line of skeptical reasoning by invoking an amended version of the “no-miracles argument”. The no-miracles argument was put forth by Hilary Putnam, who contended that scientific realism “is the only philosophy that doesn't make the success of science a miracle” (Putnam, 1975c, 73). According to Putnam, the best explanation for the success of science is the truth (or at least approximate truth) of science. If our best scientific theories were far from the truth, then the presumed success of these theories would simply be a miracle. Hence, the preferred explanation for the success of our best scientific theories is that they are approximately true, and closer to the truth than past scientific theory.

In *Inductive Inference and Its Natural Ground: An Essay in Naturalistic Epistemology* Hilary Kornblith says more-or-less the same thing about natural kinds:

If the scientific categories of mature sciences did not correspond, at least approximately, to real kinds in nature, but instead merely grouped objects together on the basis of salient observable properties which somehow answer to our interests, it would be utterly miraculous that inductions using these scientific categories tend to issue in accurate predictions. Inductive inferences can only work, short of divine intervention, if there is something in nature binding together the properties which we use to identify kinds. Our inductive inferences in science have worked remarkably well, and, moreover, we have succeeded in identifying ways in which the observable properties which draw kinds to our attention are bound together in nature. In light of these successes,

---

33 A similar point is made by Locke. In his *Essay* Locke argues that the classes or kinds that we recognize are an intellectual invention, their essences nominal rather than real. This is not to say that there is nothing in nature sustaining the observed similarities among things. The point is that, since we know nothing of the internal essence of things, the sorting of things into kinds is entirely “the workmanship of the understanding.” As Locke (III, iii, 13) notes in the context of species of animal: “I would not here be thought to forget, much less to deny, that nature in the production of things, makes several of them alike: there is nothing more obvious, especially in the races of animals, and all things propagated by seed. But yet, I think, we may say, the sorting of them under names, is the workmanship of the understanding, taking occasion from the similitude it observes amongst them, to make general ideas, and set them in the mind, with names annexed to them...to which, as particular things existing are found to agree, so they come to be of that species, have that denomination, or are put into that classis [class]”.
we can hardly go on to doubt the existence of the very kinds which serve to explain how such successes were even possible (Kornblith, 1993, 42).

Kornblith goes on to argue that, in science, we postulate unobservable causal properties for the purpose of explaining why observable properties are found to co-occur. When we postulate the existence of unobservable causal properties—or “hidden bonds”, as Darwin put it—and this postulation leads to explanatory and predictive success, then we have “extraordinarily good evidence that the structure postulated does genuinely exist” (ibid., 43).

Similarly, Khalidi argues that the aim of science is to “identify projectible properties, particularly clusters of properties that point reliably to yet other property clusters” (Khalidi, 2013, 80) and “the fact that these properties are projectibly clustered indicates that there are casual links between them” (ibid). In other words, successful inductive inference would be utterly miraculous if our cluster concepts were not accommodated to the causal structure of the world.

One obvious difficulty with this response to skepticism about natural kinds is that ‘success’ is both an ambiguous and vague notion. It is ambiguous in the sense that, even if we limit “success” to the epistemic purposes to which scientific investigation is put34, a category might be successful in some respects and not others. Consider, again, the concept of ‘dementia’. Although no serious scientist still considers dementia a kind unto itself, the concept of ‘dementia’ was successful in the sense that it allowed one to predict the general course and prognosis of the putative disease—namely, progressive cognitive decline. As it happens, however, the concept of ‘dementia’ turned out not to be

---

34 Cf. Khalidi (2013, chapter 2).
explanatorily adequate, since individual members of the kind did not share a causal structure that sufficiently explained their shared symptomatology.

The term ‘success’ is also vague inasmuch as it admits of degrees. That is to say, a science and the kinds that it postulates can be more-or-less successful. For example, there was a time in the early classification of chemical kinds when color was moderately successful in distinguishing between different kinds of things. As Kornblith (1993, 39) notes, “differences in color among substances are indeed accompanied by differences in their possibilities of chemical combination”; yet, color turned out to be a “very imperfect indicator of chemical kinds” (ibid., 40).

It would seem, then, that limited predictive success is not sufficient for considering a kind natural—explanatory success is also necessary. In order to be reasonably sure that a kind is natural our postulation of unobservable causal properties ought to successfully explain why it is that the observable properties of a kind are found to co-occur.

This point will become especially acute when we discuss psychiatric kinds. As I will argue in due course, many psychiatric kinds have enjoyed some predictive success, but limited explanatory success.

1.6 Conclusion

The theories of natural kinds outlined in this chapter do not exhaust everything that has been said about natural kinds.\(^{35}\) Be that as it may, the idea that natural kinds are

\(^{35}\) In chapter 3 I will address a slight variant of the simple causal theory when discussing mental disorders. For further discussion of natural kinds, two of the best book-length surveys, with thorough references to the past forty years of debate, are Khalidi (2013) and Laporte (2004).
collections of individuals who are alike in important respects is sufficient to cast doubt on Hacking’s suggestion that skepticism or pluralism about natural kinds are the only possible ways of responding to the murkiness of the natural kind concept and its idiosyncratic usage amongst philosophers. In defining natural kinds more broadly as collections of individuals unified by important properties or real similarities, we can concede to the skeptic that there is no well-defined class of natural kinds, but also deny that natural kinds are merely collections of individuals similar in some respects.

More importantly, however, this conception of natural kinds will serve as an ideal starting point for assessing psychiatric kinds. As we will see in due course, debates about natural and artificial classes have not been lost on medicine.
CHAPTER 2

Psychiatric Kinds and Natural Kinds

[In the human mind there is a] a strong propensity not only to make divisions in knowledge where there are none in nature, and then to impose the divisions upon nature, making the reality thus comfortable to the idea, but to go further, and to convert the generalizations made from observation into positive entities, permitting for the future these artificial creations to tyrannize over the understanding.

— Henry Maudsley, in *The Physiology and Pathology of the Mind*, 1867, 323-324.

2.1 Introduction

Debates about natural kinds might seem remote from concerns about psychiatric classification, but the controversy over natural and artificial classification introduced in the previous chapter has never been lost on psychiatry. A recurring theme in nineteenth century psychiatry was whether or not the then nascent science of psychiatry ought to tow the line of biology and medicine more generally by attempting to build a so-called natural classification of mental disorders. The 1860-1861 debate on psychiatric classification organized by the *Société Médico Psychologique* serves as an apt illustration of this.\(^1\) In the session of 26 November 1860 Jules Falret accused current species (viz. kinds) of mental disorder as being artificial, in much the same way as Adanson and Darwin accused Linnaeus’ species and genera as being artificial. He argued that the categories of mental disorder recognized at the time\(^2\) tended to be arbitrarily constructed around one feature (e.g., the involvement of the intellectual faculty, a predominant idea or emotion or act, or

---

\(^1\) See Berrios (1999) for an extended summary of this debate.

\(^2\) In the nineteenth century leading European classifications (or “nosologies”) of mental disorders were typically made up of a short-list of conditions including *melancholia* (*lypemania*), *mania*, *monomania*, *dementia*, *paralytic insanity*, *idiocy*, and *stupidity*. As Berrios (1999) has noted, this list appears to have been a minor nosological expansion or modification of Cullen’s 1872 list (which included mania, melancholia and dementia) to which Pinel added idiocy, Esquirol added monomania, Georget added stupidity, and Parchappe added paralytic insanity.
the presence of this or that delusion) rather than a set of important features which, in the
different individuals, evolved similarly across time.

While the way in which a disorder evolves over time (i.e. the order in which
symptoms appear, etc.) was an important criterion for psychiatric classification according
to Falret, it was not the only candidate. Candidates for a natural organizing principle of
psychiatric classification included: etiology (the cause of a disorder), anatomy (lesions or
atrophy in the brain observed post-mortem), clinical outcome (whether a disorder was
curable or incurable), phenomenology (whether or not a particular symptom was present
or absent, e.g., delusions), psychology (whether or not a particular mental faculty was
implicated, e.g., the intellectual, emotional, or volitional faculties), and disease
progression (i.e. whether or not a mental disorder worsened over time).

Unlike biology, however, there was no classificatory revolution in psychiatry.
Some commentators, such as Philippe Buchez, suggested that, since the objective of
psychiatric classification is diagnosis and treatment, mental disorders ought to be
classified according to their causes, not their signs and symptoms. Yet others, such as
Delasiauwe and Garnier, rebutted that there are multiple features, factors, and influences
that contribute to mental illness, and so the search for the cause of a mental disorder is at
best utopian. With this in mind, Garnier himself suggested that the classification of
mental disorders ought to be based on mental faculties, presumably under the assumption
that a particular mental faculty was impaired in each and every instance of mental
disorder, and that the various potential causes leading to such an impairment could then
be investigated.³

Today, more than 150 years after the foregoing debate took place, confusion concerning the ontological status of mental disorders, as well as how mental disorders ought to be classified, has not dissipated.⁴ For the purposes of this chapter I will set aside the question of the appropriateness or usefulness of natural classification to psychiatry and instead concentrate on whether or not mental disorders are natural kinds. As I briefly mentioned in chapter 1, part of the confusion over whether or not kinds of mental disorder are “real” or “natural” can be attributed to disagreement about what exactly real or natural kinds are. Another source of confusion I suspect stems from a subtle ambiguity in the question itself, an ambiguity which is most often glossed over in the philosophical literature on natural kinds and mental disorders. When one asks “Are mental disorders natural kinds?” one might mean one of three things: (i) Might there plausibly be natural kinds of mental disorder, or is there something in principle that prevents mental disorders from being natural kinds?; (ii) Supposing that nothing in principle prevents mental disorders from being natural kinds, are mental disorder categories listed in our current classification schemes natural kinds?; and, finally, (iii) Are some mental disorder categories not listed in our current classification schemes natural kinds?

In what follows I contend that, despite a few notable exceptions, there is no reason

³ Although, he would later admit that classifying according to impairments in, for example, judgement and reasoning was more difficult than it at first appeared, since these mental capacities are also affected by other mental capacities such as memory and attention.

⁴ As I outlined in the introduction, some philosophers and scientists contend that there is a sense in which some kinds of mental disorder are natural or real and that psychiatric classification ought generally to be developed in terms of a natural kind approach, since it would be an improvement on current psychiatric classification (Cooper, 2005, 2007; Kender et al., 2011; Tsou, 2007, 2008, 2012, 2013, 2016). Others contend that current mental disorders are not natural kinds at all, and that the natural kind approach to psychiatric classification may not even be possible or desirable (Hacking, 1992, 1995a, 1995b, 1999, 2007c; Hyman, 2010; McGuinn, 1991; Zachar, 2002, 2014a, 2014b).
to think that psychiatric categories or kind-concepts are natural kinds. In particular, I argue that, while nothing in principle prevents mental disorders from being natural kinds, paradigm examples of kinds of mental disorder, such as schizophrenia and depression, are individuated by superficial rather than important similarities, and so cannot be said to be natural kinds in the sense outlined in chapter 1. First, I critique and ultimately dismiss three general arguments for the conclusion that mental disorders cannot in principle be natural kinds. Second, I utilize my discussion of natural kinds in the previous chapter to make the case that one paradigmatic psychiatric kind, depression, is not a natural kind. I then generalize my treatment of depression to other categories of mental disorder—adding the qualification that some non-paradigmatic mental disorders might plausibly be natural kinds.

2.2 Can Mental Disorders Be Natural Kinds?

Oftentimes it is supposed that there is something fundamentally different about mental disorders which prevents them from being natural kinds. In this section I consider three arguments in favour of the conclusion that mental disorders cannot be natural kinds:

(2.2.1) **The Argument from History.** Mental disorders cannot be natural kinds because they have histories quite unlike paradigmatic natural kinds; (2.2.2) **The Argument from Looping.** Mental disorders cannot be natural kinds because they are subject to looping effects, whereas natural kinds are not; (2.2.3) **The Argument from Mind-Dependence.** Mental disorders cannot be natural kinds because they are mind-dependent, whereas natural kinds are not.

2.2.1 The Argument from History
In comparison to some paradigmatic natural kinds, many kinds of mental disorder have atypical histories. Whereas hydrogen and helium are as old as the Big Bang, most mental disorders seem to have emerged only recently. The putative kind schizophrenia, for instance, was not introduced until 1908 by Eugen Bleuler\(^5\), and descriptions of a coherent schizophrenia-like presentation are virtually absent in the medical literature prior to the nineteenth century\(^6\). What is more, some mental disorders seem to be what Hacking (1998, 100) calls “transient”. They appear at specific times and places, eventually fade away, in some cases reappearing at other times and places. Hacking points to dissociative fugue, hysteria, and multiple personality disorder, but one could no doubt add others: neurasthenia, chronic fatigue syndrome, anorexia nervosa, attention-deficit/hyperactivity disorder (ADHD), substance addiction, monomania.\(^7\)

What are we to make of the apparent waxing and waning of psychiatric kinds over the course of the last 200 years? Some philosophers simply dismiss out of hand such kinds as hysteria and neurasthenia under the assumption that, since the categories are no longer in use, science has found them not to be real kinds (Cf. Khalidi, 2013, 64; Kornblith, 1993, 48-49). Other philosophers make stronger claims. They contend that such kinds cannot be natural kinds either because natural kinds are more-or-less ahistorical (Zachar, 2002; 2014, 149), whereas most psychiatric kinds are not, or because

---

\(^5\) Cf. Bleuler (1908)

\(^6\) See Shorter (1997, 61-62). Delusions, hallucinations, diminished emotional expression, disorganized speech, and disorganized or catatonic behaviour are the most recently listed diagnostic criteria for schizophrenia. See APA (2013, 99).

\(^7\) For instance, in the first half of the nineteenth century monomania was the disease entity du jour within French psychiatry; in the latter half of the nineteenth century it was hysteria. See Goldstein (2001) [1987] for a historical account of the rise and fall of the ‘monomania’ and ‘hysteria’ concepts in nineteenth century French psychiatry.
natural kinds are more-or-less stable objects of inquiry, whereas psychiatric kinds are not (Tsou, 2008, 66-67; Hacking, 1992, 1995a, 1995b, 1999, 2007c). Let’s consider the latter two claims concerning the possibility of mental disorders being natural kinds.

If it is possible for species to be natural kinds, then the assertion that mental disorders cannot be natural kinds since they are not ahistorical is clearly implausible. After all, as far as evolution by means of natural selection is concerned, species are historical kinds. They evolve across time and space. New species emerge and most become extinct. The historical nature of species is also reflected in biological taxonomy, where individuals are most often grouped together in virtue of their shared history. Although some biological taxonomists will consider synchronic features of organisms when grouping members of the same kind (namely, the phenetic school of taxonomy), the dominant trend in biological taxonomy is to group together individuals based on diachronic features (e.g., causal history, phylogeny, and descent).

But can species be natural kinds? As I outlined in chapter 1, the principle distinction between a natural kind and an artificial kind is that, in the former case, a kind is individuated by important properties, whereas, in the latter case, a kind is individuated by means of arbitrarily selecting some properties or similarities to the exclusion of others (see, especially, sections 1.2 and 1.3). Moreover, I said that ‘important properties’ is most often taken to mean causal properties, since, insofar as causal properties are responsible for the other properties of a kind, they make a thing the kind of thing that it is (see, especially, sections 1.4.3 and 1.4.4). Species are good candidates for real kinds in the sense described. Generally speaking, with respect to species a shared history is a good
indication of other shared properties, both superficial and causal. (Cf. Khalidi, 2013, 131; Laporte, 2004, 21). This licenses both historical (or, one might say, “etiological”), physiological, behavioural, and ecological explanations with respect to why a particular species is the kind of thing that it is. Polar bears, for instance, can swim long distances in icy waters because their ancestors are amongst a group of bears which survived by adapting to such conditions, later passing on traits that make for efficient swimming in icy waters. That’s the historical explanation. But we may also appeal to the anatomical/physiological features that make it such that polar bears are capable of swimming in icy waters, or we may point out the polar bears enjoy these adaptive features because they successfully carved out an ecological niche by moving from one habitat to another in search of more abundant resources (e.g., seals).

The fact that some natural kinds might exist for a period of time and then subsequently disappear also should not be surprising in light of what we know about infectious disease. By the sixteenth century smallpox had become a leading cause of mortality in much of the world. But diseases disappear. The last documented case was in 1977, and the World Health Organization (WHO) announced its eradication in 1980.8

---

8 “Smallpox”, WHO Factsheet. Of course, the smallpox virus (there are two variants, Variola major and Variola minor) itself might still exist (in biowarfare laboratories, for example), leaving open the possibility that smallpox might arise at some future date.
With 22 reported cases in 2017, polio might soon follow. Clearly, if there is something that prevents kinds of mental disorder from being natural kinds, it’s not the fact that they do not exist at all times and in all places.

While many philosophers of psychiatry do accept that historical objects such as species and diseases might be natural kinds, some still maintain that psychiatric kinds have histories totally unlike the histories of natural kinds, and this prevents them from being natural kinds. Most notably, whereas kinds of species and disease are relatively stable over time, and are thus stable objects of inquiry, (at least some) kinds of mental disorder appear to be “moving targets”. ‘Moving target’ is a phrase employed by Hacking (2006) to describe kinds which are always “on the move” (ibid., 293) in the sense that the properties associated with such kinds rapidly change over time. For instance, Hacking (1995a) asserts that multiple personality disorder, which was considered a special case of hysteria prior to the publication of DSM-III, became a kind-concept or category unto

---

9 “Poliomyelitis”. WHO Factsheet.

10 We can also account for the rise and fall of many infectious diseases by appeal to causation. In the transition from nomadic to Neolithic society plants and animals were increasingly domesticated, forcing human beings to establish more permanent settlements. More-or-less daily contact with animals put in place the appropriate conditions for viruses, once exclusive to animals, to mutate and transfer to humans. The subsequent spread of infectious disease can be explained by a variety of factors: faecally-polluted water, population growth, population congestion in villages, towns, and cities, etc. The “fall” of (some) infectious diseases is of course more familiar. In the case of smallpox, the development of a smallpox vaccine (preceded by the method of inoculation or “variolation”)—coupled with more sanitary living environments, farming practices, and a concerted effort by the WHO and other health organizations to administer the vaccine worldwide—led to its demise. See Porter (2002).

11 Unless, of course, we suppose that a kind is “natural” or “real” only if it has existed since the beginning of time. Indeed, Zachar makes this assumption in several places (see, e.g., Zachar, 2002, 167; 2014, 149), though he does not defend it.

12 Tsou (2008, chapter 2), for instance, makes this assumption in distinguishing between so-called “stable” psychiatric kinds—namely, schizophrenia—and “unstable” psychiatric kinds such as hysteria. The former, for Tsou, is a natural kind, whereas the latter is not.

13 Since the publication of DSM-IV “Multiple Personality Disorder” has since been relabelled “Dissociative Identity Disorder”. The defining features of both the former and the latter are virtually identical from DSM-III to DSM-5.
itself following a series of reported cases of “multiples” in the 1970s. According to Hacking, the classification “multiple personality” was transformed because the experiences and behaviours of people who might have previously been diagnosed with *hysteria* rapidly changed to such an extent that having two or more distinct personalities became a predominant feature rather than just one of many hysterical signs. Let’s address the claim that “unstable” or “mutable” historical kinds cannot be natural kinds.

Although there might be reason to think that kinds of mental disorder change much more quickly than kinds of species, bacteria, viruses, and so forth, the assertion that kinds of mental disorder cannot be natural kinds in virtue of this fact is clearly a *non sequitur*. Even if it were true that the properties of mental disorders are continually changing, this would only imply that such kinds are difficult to study and classify, not that they cannot possibly be natural kinds. For instance, if Hacking’s story about *multiple personality disorder* is correct, it is possible that multiple personality arose as a kind unto itself not because we “made it up”, or because we finally identified a disorder which had always existed, but, rather, because a constellation of causal factors which were previously absent conspired to produce the cluster of signs and symptoms now associated with the disorder. If this is a possibility—and I think that it is—it is likewise possible that *multiple personality disorder* (and, indeed, any other “transient” or “mutable” psychiatric kind) is a natural kind in accordance with the account of natural kinds outlined in the previous chapter. In any case, in claiming that unstable kinds cannot be natural kinds, it would seem that some philosophers have confused an epistemological problem for an
ontological one.  

2.2.2 The Argument from Looping

While the Protean nature of some psychiatric kinds does not in principle prevent them from being natural kinds, it has been suggested that the way in which kinds of mental disorder change makes them fundamentally different from natural kinds. For instance, in a series of publications Hacking distinguishes between “human kinds” and “natural kinds”, claiming that the former are subject to “looping effects” whereas the latter are not. The phrase ‘looping effect’ refers to the way in which individuals’ experiences and behaviours change in response to being classified or defined in a certain way, forcing us to constantly revise our categories in light of these changes. Unlike kinds of physical particles, chemical elements, and (most) species, human beings are both aware of, and concerned with, how they are classified. This results in a kind of feedback mechanism that appears to be absent in cases of paradigmatic natural kinds.

A looping effect, for Hacking, operates in the following way: Once we give a label to some human phenomenon—e.g., ‘hyperactive child’—we “get the notion that there is a definite kind of person” (Hacking, 1999, 27), the hyperactive child. This “kind” then becomes reified to an extent that parents, teachers, therapists, etc., begin to regard, and

---

14 Although I do not think that instability in principle prevents mental disorders from being natural kinds, in practice instability might make it highly unlikely that our psychiatric kind-concepts or categories will reflect the natural structure of the world. I return to this issue in chapter 4.

15 Cf. Hacking (1992, 1995a, 1995b, 2007c). In more recent work (Hacking, 1999, 103-106) Hacking uses the phrases ‘interactive kind’ and ‘indifferent kind’. If I understand Hacking correctly, these latter phrases should not be taken to be equivalent to the terms ‘human kind’ and ‘natural kind’, respectively, but as pointing to one crucial difference between the kinds of things which philosophers have traditionally called “natural kinds” and the kinds of things (e.g., child, adolescent, teenager, woman refugee, the unemployed, etc.) of which Hacking speaks.

16 This is my example. In the quotation just cited Hacking discusses the kind child television viewer.
interact with, some children as if they were a distinct species of child. Children, being self-aware, respond in kind. They may begin to think of themselves as hyperactive children and alter their behaviour to further conform to this category. They may also react against, or reject, the label and/or the theories associated with it, changing their behaviour to further distance themselves from the putative kind. In either case, however, alterations in the experiences and behaviours of children who are categorized eventually results in individuals subsumed under the category ‘hyperactive child’ having different properties than they did previously. This change in properties then “loops back” (in the sense that a change in the properties previously associated with the kind is noticed by others), forcing us to revise our beliefs about the kind in question. This process can repeat indefinitely, making kinds of this sort “moving targets”.

In extreme cases the kind of feedback that Hacking discusses in connection with human kinds may result in the manifestation of an entirely new kind. Consider, again, multiple personality disorder. When patients with distinct personalities started to appear in popular culture in the 1970s (e.g., on American talk shows, in the book Sybil and the subsequent miniseries featuring the same title\(^{17}\)) more and more people started presenting with signs and symptoms now characteristic of multiple personality disorder. Accompanying the rise of cases of multiples was also a rise in the number of alters (i.e. distinct personalities) reported by patients and therapists\(^{18}\), as well as a change in the kinds of alters being reported (e.g., personalities of the opposite sex and animal

\(^{17}\) The book *Sybil* was written by Flora Schreiber and published in 1973. The miniseries *Sybil* was released in 1976.

\(^{18}\) Kihlstrom (2005), for instance, cites a number of studies suggesting a rise from one alter personality being typically associated with “multiple personality” prior to 1970 to an average of 13 by the mid-1980s.
personalities). In Hacking’s (1995a) view, multiple personality is a prime example of how looping effects may result in “making up [kinds of] people” that did not previously exist (ibid., 6). Prior to the publication of DSM-III multiple personality was not considered a kind unto itself. Once the label ‘multiple personality disorder’ entered the public consciousness, however, it served as a “semantic contagion” (ibid., 238, 247, 255-259). Individuals were increasingly regarded as a distinct species, expected to behave in ways consistent with the description of multiple personality. What is more, in the same period multiple personality became increasingly bound up with the idea that the emergence of an alternative personality was a coping mechanism to deal with trauma from child abuse (often sexual in nature). According to Hacking, the explanation of multiple personality in terms of traumatic memories of child abuse provided the “narrative structure” whereby people could subsequently “make up themselves” (ibid., 6) (i.e. quite literally alter their own experiences) by redescribing and reworking their own memories.

Granting, for the sake of argument, that human beings are subject to the kinds of looping effects that Hacking describes, in what way is this kind of feedback pernicious to the claim that mental disorders can be natural kinds? As some commentators have pointed out (Cf. Bogen, 1988; Cooper, 2004, 78-80; 2005, 59-61), it cannot merely be the fact that the classification of human kinds results in feedback. After all, some natural kinds are also altered by feedback as a consequence of our classificatory practices. Consider, for

---

19 See, also, Hacking (1986).

20 The assumption Hacking makes here is that people “act under a description”.

21 As he puts it: “[G]ood stories use explanations. Dissociation is explained as a coping mechanism. The multiple comes to understand that she is as she is now because of the way she deployed coping mechanisms in the past. A narrative structure is available that can then be filled in with the appropriate scenes” (ibid., 256).
example, herbicide-resistant weeds (or “superweeds”). In the 1990s engineers discovered a way to develop glyphosate-resistant crops. This allowed farmers to spray fields with herbicide (e.g., Monsanto’s roundup weedkiller, glyphosate being the primary ingredient), quickly killing weeds with little chance of harming the genetically-modified crops. While this method of using herbicides was much more efficient than methods used in the past, over time some kinds of plants developed a resistance to glyphosate and passed on this genetic material to offspring.\(^\text{22}\) Evidently, classifying certain kinds of plants as “weeds” triggered a chain of events whereby (some of) the very plants classified underwent significant change.\(^\text{23}\)

The fact that the classification of human kinds results in feedback is not enough to prevent human kinds from being natural kinds, but perhaps feedback materializing in the case of human kinds is of a special sort. In several places Hacking appears to make this suggestion, asserting that only with respect to human kinds is it true that the ways in which an individual thinks about themselves thereby changes them. The reason for this is that only human beings are both aware of and care about how they are categorized.\(^\text{24}\)

Consider, for instance, Hacking’s (1999, 105-106) discussion of microbes. Microbes are not passive. When we attempt to kill harmful microbes with penicillin, some may respond to this intervention by mutating in species-beneficial ways, eventually

---

\(^{22}\) A family of weeds that includes waterhemp (Amaranthus rudis), for instance, has developed glyphosate-resistant strains. See Service (2013).

\(^{23}\) Similarly, as Bogen (1988) has argued, classifying marijuana as ‘illegal’ can change the physical appearance of the plant itself, since illegal marijuana plants are often grown in different environments—e.g., basements, attics, closets, etc. The same type of feedback also appears to occur in the selective breeding of animals. See, e.g., Khalidi (2010, 2013, 147-148).

becoming resistant to our antibacterial medication. Microbes are thus “interactive” in this sense that they change in response to our intervening on them. Be that as it may, microbes do not interact with the idea of microbes. That is to say, they are not interactive in the sense that they change in virtue of being aware that they are called microbes. Rather, they are “indifferent”.

In appealing to “awareness” on the part of those classified, Hacking appears to have pointed to an important difference between human kinds and natural kinds. The problem with this suggestion, however, is that Hacking himself does not seem to think that awareness on the part of the individual classified is a necessary condition for looping to take place. And for good reason. Consider once more the case of the hyperactive child. Even if a hyperactive child is unaware of how they are classified, they may nevertheless react to interventions predicated on the idea of hyperactive children. For instance, they may be placed in different classrooms, be given medication or psychological intervention, be treated differently at home or school, and so forth. Over time such interactions may result in hyperactive children having different properties than they did previously, forcing us to revise our beliefs about them.

As Khalidi (2013, 146-150) has argued, if awareness on the part of the individual classified is not a necessary condition for looping to occur, then there does not seem to be any fundamental difference between human kinds and natural kinds. After all, even if we stipulate that at least someone must be aware of a classification for looping to take place, this would fail to exclude so-called natural kind categories. Our awareness of microbes,

---

25 See, for example, Hacking (1999, 32, 103).
for instance, occasions us to intervene on microbes in the ways that we do. This changes the way in which microbes behave, and loops back to force changes in our classifications and knowledge about them.

Notwithstanding the fact that some natural kinds appear to be subject to looping effects, Hacking nevertheless maintains that the way in which looping occurs in the case of human kinds marks a fundamental difference. For example, after posing the question, “Is there not a looping effect between the microbe and our knowledge?” (Hacking, 1999, 106), Hacking replies: “My simple-minded reply is that microbes do not do all these things because, either individually or collectively, they are aware of what we are doing to them…they do interact with us. But not because they know what they are doing” (ibid.). This passage is somewhat cryptic, but here’s what I think Hacking is driving at: only human beings, as a distinct species, interact with ideas about them, either by becoming perspicaciously aware of how they are classified (i.e. hearing a word and understanding what it means), or by becoming more subtly aware of how they are classified in virtue of experiencing their own existence in a “larger matrix of institutions and practices surrounding this classification” (ibid., 103). It would seem that, for Hacking, microbes don’t interact with ideas about them not only because they are not (perspicaciously) aware of how they are classified, but because they don’t even experience themselves as being classified in a certain way.

---

26 As I have already noted in chapter 1, Hacking has since abandoned the notion of a natural kind. It is not clear, however, what implications this has for his later distinction between interactive and indifferent kinds. Even if there is no such thing as a natural kind, one might still maintain a distinction between kinds which are subjects to looping effects and kinds which are not.

27 Although Hacking himself does not put it in this way, he gestures at these different forms of awareness in several places (Cf. Hacking, 1999, 103, 105).
If Hacking is right that only human beings interact with ideas about them\textsuperscript{28}, and that looping occurs exclusively as a function of this interaction, then this certainly marks a difference between human beings and other kinds of species, chemical elements, physical particles, etc. What is not clear, however, is how this prevents human kinds from being natural kinds. As Cooper (2004, 79; 2005, 60) has pointed out, some kinds of things are vulnerable to mechanisms to which other kinds are not, but we would not cite this as evidence that they not natural kinds. For example, the fact that animals can be selectively bred whereas chemical elements cannot does not imply that the former are not natural kinds.

Unfortunately, Hacking himself is not altogether clear on how being subject to looping effects prevents a kind from being natural or real. In some places he suggests that feedback which results from the classification of human kinds occurs at a significantly greater rate than that which manifests in other kinds, ultimately confounding our ability to study and classify such phenomena.\textsuperscript{29} As I have already argued (see section 2.2.1), however, the fact that a kind rapidly changes does not \textit{in principle} prevent it from being natural kind. In order to establish the conclusion that human kinds are not natural kinds Hacking must show that the difference between the former and the latter is metophysically (as opposed to epistemologically) significant.

\textsuperscript{28} One might, for instance, object to this by maintaining that some non-human animals (e.g., dogs) interact with ideas about them in the sense that they experience themselves as being classified in a certain way. I think Hacking’s likely reply to this would be that part of experiencing being classified in a certain way is caring about how you are being classified. Insofar as non-human animals are indifferent to how they are classified, they do not interact with ideas about them in the same way that human beings do. I will not expand on this point here, since, even if Hacking is right about this, I don’t think this prevents human kinds from being natural kinds.

\textsuperscript{29} Hacking (1999, 108) also claims that the rate of feedback marks a fundamental distinction between the natural and human (or “social”) sciences. See also: Hacking (1992, 190).
In other places Hacking appears to be working with an understanding of natural kinds which is at odds with how I’ve been discussing them. For instance, just after distinguishing between interactive and indifferent kinds, Hacking writes: “When philosophers talk about natural kinds, they take the indifference—in my technical sense—of natural kinds for granted” (Hacking, 1999, 106). According to Hacking, when the concept of a natural kind came into currency philosophers thought of Nature as passive, not active. That is to say, in philosophizing about tigers, horses, water, sulphur, and so forth, philosophers did not anticipate that there might be kinds of things—namely, human kinds—which interact with their very own classifications. This might well be true, but it does not imply that human kinds are not natural kinds. Even if paradigmatic natural kinds are more passive than human kinds, human kinds might still be considered natural or real inasmuch as members of those kinds have in common important properties. Nothing about the idea of natural kinds logically implies that they are passive.

In my mind, Hacking has identified something unique about the way in which classification affects human beings. What he has not shown, however, is that this prevents human kinds from being natural kinds. Quite the contrary. If human kinds do indeed change in virtue of individual members of a kind interacting with ideas about them, then this mechanism of change would seem to be an important property of human kinds (and, hence, psychiatric kinds) qua natural kinds, one that ought to be taken into account if our goal is to have our categories reflect the natural structure of the world.

2.2.3 The Argument from Mind-Dependence

When Hacking first introduced his influential looping effect argument he assumed
that the properties associated with human kinds do indeed change as a consequence of
looping. He has since conceded the possibility that in some instances looping might only
alter the superficial or stereotypical properties of a human kind, leaving the important
posits that the putative kinds schizophrenia and childhood autism, though subject to
looping effects, might be natural kinds in the sense that they are ultimately produced by
some underlying “biological pathology $P$” (ibid., 119). According to Hacking, if
childhood autism is (i.e. is identical to) a certain pathology $P$ (e.g., a genetic,
biochemical, or neurological abnormality), then it is “natural” or “real” for the reason that
$P$ “is, by hypothesis, not what it is in virtue of anything conscious, self-aware” (ibid.).
While the idea of childhood autism might be “constructed” or “socially constructed” (this
is Hacking’s terminology; see ibid., 121) in the sense that it interacts with researchers,
therapists, and autistic children themselves, childhood autism itself—qua $P$—is not.
Hacking’s discussion of childhood autism (and schizophrenia) highlights another possible
reason for thinking why at least some human kinds cannot be natural kinds. As you will
recall, in the previous chapter I called attention to a crucial feature of natural kinds
according to Mill and other natural kind theorists: natural kinds are mind-independent in
the sense that they exist independently of how we think of them. But if (some) human
kinds are indeed shaped by mental properties such as our beliefs about ourselves, or
others’ beliefs about us, then they would seem to be mind-dependent in a way in which
paradigmatic natural kinds are not.

Mind-independence is often taken to be an important criterion for realism about
the external world, whether we are considering individual objects or kinds (Cf. Boyd, 1989; Devitt, 2005; Kitcher, 1992). But as both Cooper (2004, 79-80; 2005, 61) and Khalidi (2013, 142-150) have noted, ‘mind-independence’ is an ambiguous notion. By “mind-independent” we might simply mean that some object or kind does not pertain to the mind. For instance, gold and sulphur are mind-independent in the sense that they are not associated with any mental properties, whereas schizophrenia is. “Mind-independent” might also refer to the fact that some object or kind is not a causal product of the mind, but has its properties in virtue of other, non-mental, processes. Finally, “mind-independent” might imply that some object or kind exists independently of our beliefs about it, as opposed to being a name or category that we have devised merely for the sake of convenience or out of our own ignorance.

If Hacking is to be believed, (most) psychiatric kinds are mind-dependent in the second sense. Being subject to looping effects implies that a kind is, at least in part, a causal product of the mind. It is not obvious, however, in what way being a causal product of the mind prevents a kind from being natural or real. Minds, after all, are part of the universe, and we should not prejudge the issue of whether or not some mental states or processes can be natural kinds. For the sake of continuity, consider once more the case of multiple personality disorder (MPD). It’s easy to be skeptical about MPD. If Hacking is right, had we not introduced the notion of MPD then there would have been no MPD to discover. This suggests some sort of contrivance. Be that as it may, if we grant that a group of individuals sharing a cluster of MPD-like properties did indeed emerge as a consequence of how we chose to divide the world, then we are forced to admit the
possibility of MPD emerging as a real kind.

In my mind, only the third sense of mind-independence is relevant to the assessment of whether or not a kind is natural or real, and there is nothing to suggest that mental disorders cannot in principle be mind-independent in this way. To be sure, a psychiatric category might be mind-dependent in the sense that it is invented arbitrarily, out of mere convenience, or tainted by certain biases, presuppositions, vested interests, etc. But this might be true of any category. And, as far as I can see, no fundamental difference exists between psychiatric kinds and other kinds of things that would merit the conclusion that this sort of mind-dependence is inevitable.

2.2.4 Summary

In this section I refuted three arguments that purport to show that mental disorders cannot be natural kinds. Other arguments (some of which were discussed in the context of essentialism in chapter 1) have been given for restricting natural kinds to the basic sciences such as physics and (perhaps) chemistry as opposed to the non-basic or “special” sciences, such as biology, psychology, psychiatry, and so forth. Nevertheless, these arguments are not exclusive to psychiatric kinds and so will not be discussed further in this dissertation. Let us now move on and consider whether or not mental disorders are in fact natural kinds.

2.3 Are Mental Disorders Natural Kinds?

Before considering whether or not current kinds of mental disorder are natural kinds it would be fruitful to remind ourselves of the account of natural kinds developed in the previous chapter. I argued that the central idea of a natural kind is that it is a collection
of individuals who are alike in important properties. These properties are “important” in
the sense that they cluster or “hang together” in virtue of being produced by certain causal
mechanisms. In short, natural kinds are clusters of properties which cluster together
because they have a common causal structure. This is what philosophers typically mean
when they say that the important properties of a kind make it the kind of thing that it is.

As we have seen, natural kinds are most often discussed in connection with kinds
of things or stuff—e.g., physical particles, chemical elements, biological species, etc.
Kinds of mental disorder, however, are generally believed to be kinds of states or
processes. That is to say, psychiatric kinds are not entities in themselves, but states or
processes (most often considered to be pathological in nature) that an entity is subject to.
Examples of natural kinds of states might include atomic states (e.g., oxidation state),
biological states (e.g., homeostasis), neurological states (e.g., vegetative states),
psychological or mental states (e.g., sadness, happiness, disgust, fear), and so on.
Examples of natural kinds of processes might include chemical reactions, pathological
disease processes, and biological processes such as metabolism and reproduction.

Although it might sound odd to say that there are individual members of a natural
kind of state or process, the account of natural kinds discussed in the previous chapter can
be readily adapted to deal with natural kinds of states or processes. Rather than saying
that a natural kind of process is a collection of individuals who are alike in important
properties, we might say that a natural kind of state or process refers to instances or
happenings in nature that involve the same set of important features or properties.
Consider, for example, homeostasis. Homeostasis is a biological process whereby an
organism maintains a stable internal state in spite of environmental variation and disturbances. Homeostatic regulation generally involves the following features: (i) **System Variable** (the biological characteristic to be regulated/maintained—e.g., temperature, glucose levels, weight); (ii) **Set Point/Range** (the optimal value or state of the system variable—e.g., a core body temperature of 37 degrees Celsius or 98.6 degrees Fahrenheit for human beings); (iii) **Detector** (the physiological structure/mechanism that monitors the system variable with respect to deviations around the set point—e.g., the Preoptic Area of the Hypothalamus (POA) and Anterior Hypothalamus (AH) monitor their own structural temperature, as well as receiving input from thermoreceptors in the skin); and **Correctional Mechanism** (a mechanism that restores the system variable to the set point—e.g., homoeothermic or endothermic organisms such as mammals and birds often make use of internal physiological mechanisms—e.g., vasoconstriction or vasodilation—to maintain an almost constant body temperature, whereas poikilothermic or ectothermic organisms such as amphibians, reptiles, and most fish lack an internal, physiological mechanism of temperature regulation, and so regulate temperature by changing their environment).30

To be sure, different organisms regulate different variables, have different set points, as well as different mechanisms for detecting and correcting their own internal states. Nevertheless, within a particular species particular homeostatic processes tend to be uniform. Human beings, for example, use the same set of causal mechanisms to

---

30 Mammals, of course, might also change their environments for the purpose of temperature regulation (e.g., find appropriate shelter, protection, cover and/or relief from cold or heat; put on clothes, take off sweaters) in addition to relying on internal physiological mechanisms.
achieve temperature regulation—and, in light of this, one can speak of *homeostatic temperature regulation* as a natural kind of process.

For a type of mental disorder to be considered a natural kind, then, the causal properties of instances of the disorder must all be similar, much like the causal properties of homeostatic temperature regulation. Unfortunately, as Cooper (2005, 72) has previously noted, in many instances the causes of mental disorder are so poorly understood that it is impossible to tell whether this condition is met. While I cannot, in this dissertation, do justice to each and every category currently listed in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) or the *International Classification of Diseases* (ICD), I will proceed here by arguing that one paradigmatic psychiatric kind, depression, cannot be considered a natural kind. Following this I will suggest that what is true of depression is, *mutatis mutandis*, true of most (though not all) categories of mental disorder currently recognized by psychiatry.

2.3.1 Characteristics of Depression (“Major Depressive Disorder”)

If one were to simply look at the most recent version of the DSM—the DSM-5—in order to determine whether or not the category of depression reflects a real kind in nature, one should quickly conclude that it does not. The category of depression, as with most categories of mental disorder currently listed in the DSM, is highly disjunctive. What I mean here by “disjunctive” is that a category or kind-concept collects together individuals who might have very little or nothing in common. For example, because individuals meet the criteria for depression so long as they present with five or more
symptoms, none of which are necessary\textsuperscript{31}, there are, logically speaking, 227 possible ways to meet the symptom criteria for major depressive disorder (MDD) (Cf. Zimmerman et al., 2015). It is also possible for two individuals diagnosed with depression to have very little in common. One individual, for instance, might be classified as depressed in virtue of meeting criteria (1)-(5) (see note 31), whereas another individual might be classified as depressed in virtue of meeting criteria (2) \textit{and} (6)-(9).

If two or more individuals diagnosed with depression share very few properties in common, then it is highly unlikely that MDD is a natural kind. After all, if MDD were a real kind, then we would expect instances of this kind to have more-or-less the same range of observable properties, since instances in nature that involve the same set of important causal properties typically involve a similar range of effects. As Kendell (1975, 69) puts it: “it’s almost impossible to visualize how a discontinuity could exist at a symptomatic level without being a reflection of a more fundamental discontinuity elsewhere.”\textsuperscript{32}

There are two possible responses to this kind of assessment of the DSM category of depression, as well as other similarly disjunctive categories. First, one might object that, although diverse symptom profiles are theoretically possible, \textit{in practice} individuals present with much more homogenous symptom profiles. Second, one might attempt to

\textsuperscript{31} At least one of these symptoms must be low mood or loss of pleasure, but there is no one symptom which is necessary for being considered a member of the putative kind depression. The nine criteria for depression 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) 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. See APA (2013, 160-161).

\textsuperscript{32} See also: Guze (1992).
weaken the link between causes and effects and maintain that, although individuals
diagnosed with MDD might have heterogeneous or diverse symptom profiles, the signs
and symptoms of mental disorders are superficial properties rather than causal ones, and
so it is possible that individuals classified as depressed are superficially dissimilar yet
importantly the same. Let’s consider each of these points in turn.

Recent research does support the view that MDD is more homogenous in practice
than in theory. For instance, examining more than 1500 patients who met the DSM-IV
criteria for MDD, Zimmerman et al. (2015) found that individuals met MDD symptom
criteria in 170 different ways, with a relatively small number of combinations accounting
for more than 40% of patients. More recently, Park et al. (2017) identified 119 different
symptom profiles amongst 853 individuals who met the DSM-IV criteria for MDD, the
most common of which consisted of meeting all nine criteria.

Although these findings do suggest that MDD is more homogenous than a
superficial analysis of the DSM might lead us to believe, this does not imply that MDD is
a natural kind. For one thing, 119 possible symptom profiles for MDD is still quite
heterogenous in comparison to, say, schizophrenia, in which there are 27 possible ways to
meet DSM criteria, or anorexia nervosa, which only has 1 diagnostic combination. Most
importantly, however, even if we grant that MDD is a unified syndrome, there is no
guarantee that this putative syndrome or cluster of properties is produced or generated by

---

33 The criteria for MDD in DSM-IV and DSM-5 are virtually identical (compare: APA, 1994, 327 and APA,
2013, 160-161). The only important difference between the newest version of the DSM (i.e. DSM-5) and
DSM-IV is that the “Bereavement Exclusion” was removed in DSM-5. In DSM-IV the bereavement
exclusion states: “The symptoms are not better accounted for by Bereavement, i.e., after the loss of a loved
one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment,
morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor
retardation”.
the same set of causal properties. While one might plausibly infer that the same range of effects will result from the same type of disease process across different token instances, the reverse inference often does not hold. For example, two individuals might both present with a cough, chest pain, fatigue, weight loss, and fever—yet, in one case the cluster of symptoms is the effect of *pulmonary tuberculosis*, whereas in the other case it’s the effect of *lung cancer*. To conclude that MDD is a natural kind on the basis that there are a set of individuals who share the same immediately observable properties would be the rough equivalent of concluding that *pulmonary tuberculosis* and *lung cancer* are two of a kind, or that *gold* and *pyrite* (iron sulphide or “ersatz gold”) are the same kind of thing since they have in common the observable properties of being shiny, yellow, and metallic.

But what of the objection that individuals classified as depressed might be superficially dissimilar yet importantly the same? Although we might plausibly maintain that the same set of causal properties involves a similar range of effects across each and every instance of a particular disease, a scenario in which this is not the case is not altogether unusual. *Systemic Lupus Erythematosus* (“Lupus”), for instance, is a chronic autoimmune inflammatory disease that can affect any area of the body. Since lupus can affect different parts of the body, symptoms tend to vary significantly from person to person depending on what part of the body is affected. Symptoms may also come and go, and may be mild, moderate, or severe.

Given that lupus is a natural kind of disease, one that manifests in various ways though has a distinct pathophysiology, we cannot rule out the possibility that MDD
reflects a real kind in nature on the basis that it is a heterogenous concept. In the end, what we need to know is whether or not there are causal properties that would serve to unify diverse manifestations of MDD. It is to this question which I now turn.

2.3.2 Causes of Depression

Genetic. What is the causal structure of depression—i.e. what are the causal properties that make depression the kind of thing that it is? Some genetic research supports the view that depression is heritable, with monozygotic twins having a concordance rate of 40% (Sullivan, Neale, & Kendler, 2000). Nevertheless, this only makes it moderately likely that one twin will develop depression if the other does, and the search for a “depression gene” has thus far been met with frustration. Perhaps the most well-known candidate is the serotonin transporter gene, 5-HTTLPR, which can be found in the promoter region near the DNA sequence for the serotonin transporter protein. The promoter region controls the level of expression of the serotonin transporter protein, which itself is responsible for ensuring that serotonin can move back into the axon terminal from the synapse (“reuptake”) to be reused. In theory, the higher the expression of the protein, the quicker serotonin can be taken out of the synapse. The body of evidence for the association of the 5-HTTLPR gene with major depression remains mixed. Some studies have replicated these and similar findings whereas others have not (Munafo, 2012). One meta-analysis of 39 studies between 1996 and 2009 (Clarke, Flint, Attwood, & Munafo, 2010) did find a statistically significant association between 5-HTTLPR polymorphism (namely, a “short” allele, which causes different levels of

---

34 Heritability is around 30% for men and 40% for women.
transporter expression than a “long” allele) and unipolar depression, although the authors caution that the effect size is quite small. In other words, even if the 5-HTTLPR gene does play a role in depression, then this role is unlikely to be fundamental, since the effect size attributable to the single genetic variant is very small.35

**Neurochemical.** We often hear depression characterized as a “chemical imbalance”. Neurochemical theories of depression first emerged in 1950s following observations of the effects of certain medications, such as resperine (a blood pressure medication), on mood. The most influential of these theories is the “monamine hypothesis” of depression, in which depression is said to be caused by a deficiency of the monoamine neurotransmitters (dopamine, norepinephrine, and serotonin), particularly serotonin. Evidence for the role of serotonin in depression comes from a few different sources. First, early studies found that depressed individuals had lower levels of serotonin metabolites in their cerebralspinal fluid (Cf. Traskman-Bendz, Asberg, Bertilsson, & Thoren, 1984). Second, reducing serotonin levels by eating a diet low in tryptophan appears to produce depression-like symptoms (Booij et al., 2002; Smith, Fairburn, & Cowen, 1997), as well as negative mood bias (van der Veen, Evers, Deutz, & Schmitt, 2007)—i.e. responding more strongly to negative rather than positive emotional stimuli. Third, some more recent studies have found abnormal (lower) levels 5-HT (serotonin) 1A receptors in the brains of individuals diagnosed with depression as compared to controls (Drevets et al., 1999; Sargent et al., 2000). And, finally, antidepressant medications—

---

35 It should also be mentioned that both short and long alleles of 5-HTTLPR gene have roughly the same frequency in the overall population (45% of the alleles in the gene pool are short, whereas 55% are long; see Caspi et al., 2003), suggesting that there must be more to depression than the 5-HTTLPR gene, since it is estimated that, at any given time, only about 5% of people worldwide meet current diagnostic criteria for depression.
namely, monoamine oxidase inhibitors (MAOIs), tricyclic antidepressants (TCAs), and selective serotonin reuptake inhibitors (SSRIs)—act by boosting serotonin levels in the synapse (and, in the former two cases, dopamine and norepinephrine as well) either by blocking the breakdown of serotonin and other monoamines in the presynaptic cell or by blocking the reuptake of serotonin in the presynaptic cell.

The evidence for the role of serotonin in depression would seem to be overwhelming; yet, the story of depression appears to be much more complicated than simply “low levels of serotonin cause low mood”. For one thing, reducing serotonin levels by eating a diet low in tryptophan doesn’t appear to have the same effect on healthy individuals in comparison to those who have previously experienced an episode of depression (Murphy et al., 2002), suggesting that low serotonin levels are not sufficient to produce depression. What is more, the link between lower levels of serotonin receptors and depression is inconsistent. While some studies have found such a link, others have not (Meyer et al., 2004; Parsey et al., 2006), and still others have found raised levels of serotonin receptors (Reivich, Amsterdam, Brunswick, Yann Shiue, 2004).

The evidence from treatment is likewise unconvincing. Some medications with little to no effect on serotonin (e.g., desipramine, bupropion) are just as effective in treating depression as SSRIs (Clayton et al., 2006). Furthermore, many treatments which appear to alleviate symptoms of depression—e.g., sleep deprivation, medically-induced seizures, in addition to antidepressants—also happen to stimulate the expression of a protein known as brain-derived neurotrophic factor (BDNF), which is responsible for

---

36 For instance, in some experimental studies medications which affect glutamate signalling appear to have antidepressant-like effects (Mathews & Zarate, 2013).
regulating the growth, survival, and plasticity of neurons. Studies of depressed suicide victims have found lower levels of BDNF principally in the plastic regions of the brain, such as the hippocampus (Dwivedi et al., 2003), suggesting that perhaps antidepressants (as well as other effective treatments of depression) work by acting on BDNF rather than on serotonin.\(^{37}\)

In short, we still cannot say whether or not the psychopharmacological agents which we use to treat depression are specific to, or necessary for, their antidepressant-like effects. And, in any case, just because antidepressants (particularly SSRIs) increase serotonin levels, this does not imply that low serotonin is the causal basis of depression. As David Healy puts it: arguing “backward from the efficacy of [a] treatment to what might be the cause of the disorder [makes] little more sense than arguing that because aspirin was useful in treating rheumatoid arthritis there must be some kind of aspirin deficiency implicated in rheumatoid arthritis” (2004, 191).\(^{38}\) To be sure, the neurotransmitter serotonin might be involved in depression \textit{in some way}—perhaps by affecting (or being affected by) brain activity in particular regions of the brain (more on this below)—but this is a far cry from the claim that low serotonin causes or gives rise to the symptoms associated with depression.

\textit{Neural Circuitry}. Not only is the neurochemical theory of depression underdetermined by evidence, but considered as a standalone theory of depression it is

\(^{37}\) Although, as Moncrieff (2007) has noted, research into depression and other brain abnormalities—such as cortisol levels, brain volume abnormalities, and BDNF—have thus far been inconsistent, failing to show that these factors are specific to depression, let alone causal.

\(^{38}\) Interestingly, some studies have found placebos to be nearly as effective as SSRIs in relieving the symptoms of depression (Kirsch et al., 2008; see also Kirsch, 2010). If the efficacy of an SSRI in treating depression is evidence that depression is caused by a serotonin deficiency, then, by parity of reasoning, the efficacy of a placebo in treating depression is evidence that depression is caused by a placebo deficiency.
implausible. Neurotransmitters are a necessary component of neurotransmission (namely, by transmitting excitatory or inhibitory signals across a chemical synapse), but as far as we know all of our thoughts, desires, emotions, intentions, and so forth, are a consequence of more diffuse brain activity. Even if it were true that neurotransmitters such as serotonin in some sense “cause” the symptoms typically associated with depression, presumably they only do so in virtue of affecting brain activity in specified ways. In light of this, chemical imbalances in the brain are at best part of the causal story of depression.

But what kind of brain activity is associated with depression? In the past two decades neuroimaging studies of depression have shown patterns of underactivity and overactivity in a widespread network of brain regions, including the anterior thalamus, hippocampus, anterior cingulate, subgenual cingulate, orbitofrontal cortex, lateral prefrontal cortex, and medial frontal cortex (Seminowicz et al., 2004). The subgenal cingulate cortex appears to be a key node in this network, as it is consistently overactive in individuals diagnosed with depression and returns to normal levels of activity in those who have recovered from depression, regardless of treatment method (Drevets, Savitz, & Trimble, 2008).³⁹ The subgenal cingulate plays a central role in emotion regulation via its connection to the amygdala and other core limbic structures. Although the amygdala itself is able to assign emotional valance to external stimuli by drawing on different sensory modalities (e.g., visual, auditory, somatosensory, gustatory, and olfactory inputs), it could easily overreact to external stimuli (e.g., a rubber snake) or fail to react to external stimuli.

³⁹ And in some instances using deep brain stimulation in order to reduce activity in this area results in remission. See Mayberg et al. (2005).
(e.g., the death of a loved one) if not modulated by the more sophisticated cortical networks found in the subgenal cingulate.

Communication between the amygdala and the subgenal cingulate is not all there is to emotion regulation, however. Research suggests that many other cortical areas modulate the activity of the subgenal cingulate itself. For example, the orbitofrontal and ventrolateral prefrontal cortices have strong inputs into the subgenal cingulate, and increased activation or deactivation in these areas appears to be associated with the use of context and cognitive strategies to appraise or reappraise a distressing stimulus (Johnstone, van Reekum, Urry, Kalin, & Davidson, 2007). Interestingly, in individuals diagnosed with depression increased activation of the ventrolateral cortex is correlated with increased activation of the amygdala, whereas in healthy control subjects increased activation of the ventrolateral cortex is correlated with decreased activation of the amygdala. This suggests that perhaps in the former case increased activation is a sign of negative reappraisal or “catastrophizing” (e.g., failing a course is interpreted as career-ending), whereas in the latter case its a sign of positive reappraisal or an individual making light of a situation (e.g., failing a course is interpreted as not a big deal in the grand scheme of things).  

All of this is to say that the relationship between brain activation and emotion regulation is not altogether straightforward and predictable. Depressed patients and non-

---

40 It should be noted that the precise role of the amygdala in human emotion is still contentious. Individuals with bilateral lesions to their amygdala typically have normal emotional functioning with the exception of the ability to learn and express fear (Hayman et al., 1998), suggesting that the amygdala is not necessary for processing other emotions such as happiness, sadness, disgust, anger, and surprise. For instance, S.M.—oftentimes referred to as the “woman without fear”—lacked an amygdala on both sides of the brain (and no other brain tissue) as a consequence of Urbach-Wiethe disease. Although the loss of her amygdala impaired S.M.’s capacity for experiencing and recognizing fear, it did not remove her capacity to experience and recognize other human emotions (Cf. Young, 2010).
depressed patients alike may have similar activity in the same brain region, yet this activity has differential effects elsewhere in the brain. What is more, even amongst depressed patients themselves the same brain regions—e.g., the dorsolateral prefrontal cortex and the dorsomedial prefrontal cortex—can be either underactive or overactive, suggesting that in some instances of depression there is a lack of cognitive modulation, allowing subcortical regions such as the amygdala to operate unchecked, whereas in other instances of depression there is too much cognitive modulation in the form of negative reappraisal (Oschner et al., 2004).

Network models of depression which focus on “abnormal” interactions between limbic areas responsible for emotional states and cortical areas which modulate these states are not wholly inconsistent with neurochemical and genetic theories of depression. For example, a 5-HTTLPR polymorphism may cause an imbalance in serotonin which may, in turn, cause an increase or decrease in brain activity in particular neural networks. The problem, however, is that at present this is pure speculation. While a bottom-up causal process of this sort is possible, it is equally possible that neurochemical and genetic changes are being driven by top-down processes, such as an increase or decrease in brain activity in response to environmental stimuli (more on this below). Most importantly, for our present purposes, is the possibility that what we now call “depression” is not one kind of natural process, but many. For instance, bottom-up processes may be more prevalent in some instances of “depression” whereas top-down processes may be more prevalent in others. And, as I have already said, even so-called

---

41 See, for instance, Mayberg’s (1997) limbic-cortical dysregulation model.
top-down processes, such as the modulation of limbic areas by cortical areas, might differ between depressed patients.42

Psychological and Environmental. Thus far I have highlighted some of the putative causal factors involved in depression at the level of genes, chemistry, and neurocircuitry. In addition to this there is a much older tradition which has emphasized the important causal role of both psychological and environmental factors in depression. For example, in the preface to the *Anatomy of Melancholy* Robert Burton writes: “There is no greater cause of melancholy than idleness, no better cure than business” (1621/1651, 20). As Radden (2000, 4-19) argues, in the long tradition of writing about melancholy43 there is an evident association between melancholy and idleness (and boredom), as well as the palliative effects of labor (particularly, meaningful labor). Idleness, however, was not seen to be the only possible cause of melancholy. In addition to idleness, Burton himself recognizes many “inward” and “outward” causes of melancholy (see Burton, 1621/1651, 127-128), including excessive labor/study, brain injury, excessive sexual indulgence, excessive alcohol use, bad diet, sorrow or loss, fear, worry, envy, shame, hatred, anger, ambitiousness, covetousness, and pride.

42 Indeed, even bottom-up processes might differ between depressed patients. Consider the fact that damage to different subcortical limbic regions of the brain might similarly affect different individuals. For instance, some patients with a tumour in the ventromedial hypothalamus have exhibited depressive-like symptoms (Veendrick-Meekes et al., 2007). As well, reduction in hippocampal volume has been linked to depression and bipolar disorder (Frey et al., 2007; Roddy et al., 2019). Furthermore, a bilateral lesion to the ventral globus pallidus results in symptomatology consistent with severe depression with anhedonia (Miller et al., 2006). As it happens, the ventral globus pallidus is adjacent to the ventral striatum, which has been deemed the “pleasure centre of the brain”—and stimulating the nucleus accumbens of the ventral striatum does in fact decrease symptoms of severe depression in some individuals (Benezzouz et al., 2000).

43 The term ‘melancholy’ does not have the exact same connotation as ‘depression’ or ‘major depressive disorder’. As Radden (2000) notes, from Galen onwards the central states associated with melancholy have been fear and sadness. While sadness (i.e. ‘depressed mood’) remains a part of the DSM diagnosis of depression, the subjective state of fear has dropped out entirely. *NB*. ‘Melancholy’ should not be confused with ‘melancholia’. The latter term often referred to a range of conditions, sometimes conditions similar to depression or melancholy, but oftentimes conditions which more closely resembled delusional disorders.
Medical theorizing in Burton’s day was dominated by the humoral theory of medicine, whereby excess or deficiency of thought, desire, or action were considered bad for both body and mind (the proximal cause being excess or deficiency in one or more of the four bodily humors: blood, yellow bile, black bile, and phlegm). What is striking, however, is that four hundred years on many of Burton’s claims have been corroborated. Depression is associated with traumatic brain injury (Jorge, Robinson, & Moyer, 2004), particularly damage to the left hemisphere. Depression is comorbid with Alcohol Use Disorder—though, it is not entirely clear whether alcohol causes depression or whether depression causes alcohol use (Boden & Fergusson, 2011). Lifestyle factors such as sleep, diet, and exercise are also implicated in depression, most likely via their affects on various neurobiological pathways such as neurotransmitter processes, immuno-inflammatory pathways, the hypothalamic–pituitary–adrenal (HPA) axis, antioxidant defence systems, neuroprogression, and mitochondria (Lopresti, Hood, & Drummond, 2013). In addition, individuals diagnosed with depression are three times more likely than those not diagnosed with depression to have experienced a stressful life event prior to the onset of depression (Harkness, 2008), particularly childhood trauma or loss.

Nevertheless, as with the genetics, chemistry, and neural circuitry, the environmental level of analysis is not fully explanatory. For example, although

44 Of course, it is also possible that some third factor explains the cooccurrence of both depression and alcohol use.

45 In fact, children who are victims of physical, sexual, and/or emotional abuse are two to five times more likely than those who have not been abused to develop depression in adulthood (Harkness & Lumely, 2007). As well, according to one estimate, nearly seventy-five percent of depressed individuals have suffered at least one severe loss event in the six months prior to the onset of depression (Brown & Harris, 1989), a finding which hearkens back to Freud’s (1917/1961) claim in *Mourning and Melancholia* that at least some instances of depression are a consequence of an inability to cope with loss. “Severe loss event” might refer straightforwardly to the death of a loved one. It might also refer to other significant life events such as the loss of one’s job, the loss of one’s home, etc.
experiencing a stressful life event is a strong predictor of depression, not everyone who experiences a stressful life event ends up depressed. This suggests that there must be some other factor accounting for why stressful life events trigger depression in some individuals and not others. To this end, personality and cognitive psychologists tend to invoke psychological or cognitive mechanisms, often insisting that “stressful life event” should be taken to refer to events which are interpreted as stressful by an individual, rather than a class of events deemed to be stressful from some objective standpoint. Freud, for instance, pointed out that even a relatively “insignificant” life event (e.g., failure on an exam, making a mistake at work, etc.) might be interpreted as stressful by some individuals. The American psychiatrist Aaron Beck later developed this insight, insisting that one’s emotional response to a situation is determined by the way in which a situation is appraised or evaluated, and people prone to depression tend to appraise even the most innocuous events (e.g., not being acknowledged by an acquaintance in the grocery store) negatively (Beck, 1972).

Beck’s cognitive model of depression is certainly consistent with our current understanding of emotion, whereby emotions are controlled/modified by cognitive interpretation occurring after physiological reaction to some stimulus. Be that as it may,

46 Freud referred to these cases as instances of “imagined loss” whereby an individual unconsciously interprets comparatively banal life events as severe loss events. See Freud (1917/1961).

47 This is an example of jumping to conclusions, a cognitive distortion in which one interprets things negatively in the absence of facts to support one’s conclusion. Other examples of cognitive distortions include: all-or-nothing thinking (“I’m a failure if I don’t get an ‘A’ in calculus”), overgeneralization (“I never do anything right”), and magnification or catastrophizing (“Failing this course will ruin my life”).

48 Since Schacter & Singer (1962) showed that emotions require both a physical (visceral) and cognitive (interpretation of the context) component, most theories of emotion have incorporated both bottom-up visceral factors and top-down contextual factors, differing only in the degree of emphasis placed on either bottom-up or top-down processing.
it is still not clear from this why some individuals are more likely than others to appraise situations negatively. Beck himself suggested that some individuals’ negative cognitive style is a consequence of a “depressive schema” (i.e. rigidly held negative core beliefs about the self, world, and the future) which has developed over time in response to early experiences with the world (Beck, 1979). But this is consistent with competing etiological accounts of depression. For instance, a negative cognitive style might arise in response to insecure attachment to a parent (Bowlby, 1980), early maltreatment internalized as a child (Young, 1994; Lumely and Harkness, 2007), or even a 5-HTTLPR polymorphism (Hayden et al., 2008).

Perhaps most importantly, it is by no means clear that each and every case of depression involves cognition. As I said earlier, it is possible that in some instances of depression there is a lack of cognitive modulation, allowing subcortical limbic regions responsible for emotion to operate unchecked. It is also possible that one or more of these regions is damaged (or at least not functioning optimally), causing depressive-like symptoms and behaviour (see note 42).

2.3.3 The Natural Kind Status of Depression

The fact that depression as we currently conceive it is associated with a variety of complex causal factors has led some to describe it as a “multifactorial” illness, much like diabetes or heart disease. The problem with describing depression as a multifactorial

49 For example: “I’m unlovable”, “People are out to hurt me”, “The world is a dangerous place”, etc.

50 In particular, Hayden and her colleagues found that seven-year-old children who are homozygous for the short allele of the 5-HTTLPR gene displayed a bias toward processing information more negatively than those with other genotypes.

51 This is the tack most often taken in textbook descriptions of depression (see, for example, Dozois & Firestone, 2010, 173; Eagleman & Downar, 2015, 546).
illness, however, is that it begs the question in favour of the conclusion that depression is indeed *one* kind of thing—a real kind in nature. As I have tried to make clear in the foregoing, there are good reasons to doubt this. Our current category of depression is heterogenous with respect to both superficial and causal properties. Far from reflecting a set of properties which cluster or “hang together” in virtue of sharing a similar causal structure, ‘depression’ would seem to be a ragbag term which collects together instances which are quite dissimilar both in terms of their observable and unobservable properties.

It must be emphasized that my conclusion that depression is not a natural kind should not be read as implying that, in order to be considered a natural kind, each and every instance of depression ought to share a common essence, or that individual members of the kind “depression” share all of their properties in common. It is general consensus in the philosophy of psychiatry that kinds of mental disorders, much like species, do not have essences (Cf. Kincaid & Sullivan, 2014). My point, rather, is that there is no evidence to suggest that the signs and symptoms which are characteristic of depression cluster together because of a common causal structure. In the absence of knowledge of the causal structure of “depression”—i.e. how unobservable causal properties conspire to produce the observable signs and symptoms of depression—it is impossible to know whether or not the properties of depression do indeed cluster together, or whether these groupings are merely a reflection of our predilections to group certain properties together.

To put this in terms of the notion of ‘scientific success’ discussed in chapter 1 (see section 1.5), the concept of ‘depression’ has enjoyed moderate predictive success (e.g., we
can predict that some individuals diagnosed with depression will respond to antidepressant medication) but very little explanatory success. As I suggested in the previous chapter, a kind is plausibly natural when our postulation of unobservable causal properties successfully explains why it is that the observable properties of a kind are found to co-occur. At present, this cannot be said to be true of depression. Indeed, some have argued that, while the most severe form of clinical depression—which has variously been called melancholia, endogenous depression, or nuclear depression—is a kind unto itself, most instances of clinical depression fall on a continuum with normal unhappiness (Cf. Haslam, 2002, 211; Horowitz & Wakefield, 2007).

Nor should my argument be taken to imply that, in order to be considered a natural kind, kinds of mental disorder ought to have a common unobservable biological causal structure. I agree with Khalidi (2013) that there is simply no a priori or a posteriori reason to restrict natural kinds to biological kinds, or even to physical kinds. As I argued above (see section 2.2.3), the fact that at least some kinds of mental disorder are “mind-dependent” in the sense that they are (at least in part) a causal product of the mind does not entail that they cannot be real. Be that as it may, there is also no evidence to suggest that instances of depression involve a common unobservable psychological or cognitive causal structure. In fact, as Horowitz & Wakefield (2007) have argued, there are good reasons to think that severe depression primarily involves a common biological causal structure, whereas mild to moderate depression primarily involves psychological
2.3.4 The Natural Kind Status of Other Mental Disorders

When one considers other DSM mental disorders, the heterogenous nature of mental disorder concepts is even more pronounced. According to Olbert, Gala, and Tupler (2014), there are currently 636,120 possible ways to meet the criteria for posttraumatic stress disorder (PTSD), 116,200 for ADHD, 32,647 for conduct disorder, 23,422 for panic disorder, and so on. Indeed, in most cases two individuals diagnosed with the same mental disorder will share on average less than half of their symptoms—and in some cases they will share no symptoms in common at all. As I said earlier in the context of depression, while heterogeneity itself does not imply that a kind is not natural, it seems highly unlikely that such extreme heterogeneity in symptomatology would not implicate the involvement of different causal processes.

Most importantly, however, even in cases in which there is much less heterogeneity in symptomatology relative to other mental disorders, such as schizophrenia and autism spectrum disorder, there is no consensus amongst experts with respect to which causal properties explain the superficial properties of these putative disorders. Schizophrenia, for instance, has been described as a genetic disorder, a neurochemical disorder, a neurodegenerative disorder, a neurodevelopmental disorder, and even a psychogenic disorder. In chapter 3 and chapter 4 I will consider schizophrenia in more detail. For now it is enough to say that schizophrenia is so poorly understood that

---

52 In fact, as Radden (2000) has argued, it has long been thought that there is a form of what we now call “depression” that is at least qualitatively distinct from other forms of depression. Historically, this form has been referred to as “melancholy”. See note 43.

53 There are 27 diagnostic combinations for schizophrenia and 15 for autism spectrum disorder. See ibid.
at minimum we must remain agnostic about whether or not it is a natural kind.

Furthermore, from a historical standpoint, given that categories of mental disorder have been, and still are, constructed on the basis of signs, symptoms, course, and outcome—as opposed to knowledge of the causal structure of the world—it would be serendipitous indeed if the causal properties which are responsible for psychopathological phenomena stood in simple causal relationships to DSM constructs (Cf. Tabb, 2015). Just as the combined presence of coughing, chest pain, fatigue, weight loss, and fever turned out to be an unreliable indicator of a natural kind of disease, so the combined signs and symptoms of current psychiatric concepts are likely to be at best imperfect indicators of natural kinds of mental disorder.

There are exceptions. Down Syndrome is a kind of genetic disorder in which the causal properties are well-known. It is caused by an anomaly on the twenty-first chromosome—specifically, a trisomy on chromosome 21 (i.e. an extra twenty-first chromosome) which causes biochemical abnormalities which, in turn, impair brain development. While it would be incorrect to assert that each and every case of Down Syndrome is identical (e.g., some individuals with Down Syndrome are higher functioning than others), it is nevertheless clear that sharing a genetic anomaly of this sort explains why it is that individuals with Down Syndrome share other similar properties. The same can be said for Fetal Alcohol Spectrum Disorder, Fragile X syndrome, and neurocognitive disorders such as Huntington’s Disease.

2.4 Conclusion

In this chapter I argued that, while nothing in principle prevents mental disorders
from being natural kinds, there is no reason to think that the mental disorder categories listed in our current classification schemes are natural kinds. In fact, when one considers research into the causes of paradigmatic mental disorders such as depression, the more likely conclusion to draw is that most mental disorder concepts are not natural kind concepts, but rather catch-all terms which potentially conceal important distinctions that exist between individuals diagnosed with these disorders.

As I said in the introduction to this dissertation, the most recent version of the DSM, DSM-5, includes a total of 541 diagnostic categories, 151 of which are defined by predetermined inclusion-exclusion criteria. In an essay of this scope I simply cannot address each and every one of these categories, nor can I accord proper justice to even several of them. While I do consider depression and schizophrenia to be paradigmatic kinds of mental disorder—and, in light of this, my conclusions are likely to generalize to most kinds of mental disorder—ultimately the question concerning whether or not some kind of mental disorder is natural or real needs to be determined on a case-by-case basis. All of this is to say that what I have said about depression here, and what I will have to say about schizophrenia in the next two chapters, should not be taken to apply necessarily to each and every category listed in the DSM.
CHAPTER 3

Psychiatric Kinds and Homeostatic Property Cluster Kinds

[I]t is often imagined that progress has been made simply because fancy names have been given to old things.


3.1 Introduction

Thus far I have argued that natural kinds are collections of individuals who have important properties in common—namely, a shared causal structure—and that most psychiatric kinds are not natural kinds in this sense. In fact, when one considers research on paradigmatic mental disorders such as depression, the more plausible conclusion to draw is that many, if not most, mental disorder concepts are catch-all terms which potentially conceal important distinctions that exist between individuals diagnosed with these disorders.

Even though psychiatric kinds are not natural kinds in the sense in which I have been discussing natural kinds, in recent years some philosophers have maintained that there is a sense in which at least some kinds of mental disorder can be considered “natural” or “real”. In particular, philosophers of psychiatry have defended the Homeostatic Property Cluster (HPC) view of natural kinds (Cf. Kender, Zachar & Craver, 2011; Tsou, 2007, 2008, 2012, 2013, 2016) as a way of conceptualizing mental disorders. In addition to maintaining that some psychiatric kinds are natural kinds in accordance with the HPC theory, some philosophers have advanced the further claim that HPC kinds are “useful models for psychiatric classification” (Kendler et al., 2011, 1143). That is to
say, assuming that at least some of our current psychiatric categories are not HPC kinds, revising our categories in such a way as to reflect HPC kinds would be an improvement on current psychiatric classification.

In this chapter I argue that philosophers of psychiatry who claim that mental disorders are HPC kinds, and that the HPC theory is a useful model for psychiatric classification, often equivocate between two very different versions of the HPC theory of natural kinds. In particular, I contend that there is both a strong and weak reading of the HPC theory, and that mental disorders are only HPC kinds in the weak sense. The upshot of this, I argue, is that philosophers cannot consistently hold both that psychiatric kinds are HPC kinds and that HPC kinds are useful models for psychiatric classification. I conclude this chapter by clarifying the sense in which I think kinds of mental disorder can be said to be “natural” or “real”.

3.2 Homeostatic Property Cluster Kinds

As I tried to make clear in chapter 1, the phrase ‘natural kind’ can be construed in a variety of ways, and philosophers still disagree about which theory of natural kinds ought to be adopted. Given such uncertainty, the descriptive claim that some mental disorders are natural kinds in the HPC sense of ‘natural kind’ requires further commentary. In this section and the next I will clarify what it means to say that a category is an HPC kind, as well as attempt to situate HPC kinds amongst the theories of natural kinds discussed in chapter 1. In section 3.4 I will address the argument that (some) mental disorders are HPC kinds.

3.2.1 Homeostatic Property Cluster Kinds
In chapter 1 I briefly introduced Boyd’s (1989, 1991, 1999a, 1999b) account of natural kinds as an alternative to essentialism about natural kinds. For Boyd, while there may be properties essential to kinds of chemical elements, there are not (as far as we know) properties essential to kinds of species. Yet, according to Boyd it would be hasty to conclude from this that the kinds of species that we currently recognize are merely invented and have no claim to accurately describing a portion of nature. To be sure, members of a particular species do not share some simple, deterministic essence; nevertheless, unlike the motley collection of things on my desk or Mill’s category of white things, they do have much in common. On Boyd’s view species are “homeostatic property cluster” (HPC) kinds in the sense that members of a species imperfectly share homeostatically related morphological, behavioral, and physiological features. On this view, natural kinds are not groups of entities that have every property in common; rather, natural kinds are groups of entities that share stable similarities. What is more, the HPC approach supposes that there are causal mechanisms (or “homeostatic causal mechanisms”) that are responsible for the similarities apparent among members of a particular kind. That is, the underlying mechanisms of a species cause the members of that species to have similar features.\(^1\)

Boyd summarizes his view by listing ten features of HPC kinds (ibid., 16-17). I’ve further condensed these into seven characteristics below:

F1) **Property Cluster:** a cluster of properties that regularly co-occur;

F2) **Homeostatic Mechanism:** a mechanism, set of mechanisms, or processes that

\(^1\) See, especially, Boyd (1999b), 81.
explains why a cluster of properties regularly occur together;

F3) **Causal Importance:** the property cluster together with the homeostatic mechanism(s) feature in some theoretically or practically important inductive generalization;

F4) **Extensional Vagueness**\(^2\): an HPC kind term \( t \) refers to individuals that display the homeostatic clustering of *some but not all* properties in a property cluster (and *some but not all* of the homeostatic mechanisms associated with the property cluster in question) such that it is sometimes unclear whether or not an individual is a member of some kind, \( k \);

F5) **Naturalism**\(^3\): HPC kind term definitions are *a posteriori*, theoretical definitions (as opposed to conceptual, analytic definitions). And, given extensional vagueness, it is often difficult to determine theoretically (if at all) which properties belong to the definition of a kind term, \( t \);

F6) **Revisability**\(^4\): revising our taxonomy of kinds *in order to remedy extensional vagueness* would be at the expense of preserving the naturalness (viz. realness) of the kind in question; and

F7) **Individuation:** HPC kinds are individuated historically, not extensionally.

With respect to F1, a property cluster refers to some more or less stable phenomenon of interest. In the context of psychiatry, a property cluster may refer to the

---

\(^2\) This feature is intended to summarize Boyd’s (1989, 16-17; 1999, 143-144) features vi and viii, “imperfect homeostasis” and “extensional indeterminacy”, respectively.

\(^3\) This feature is intended to summarize Boyd’s (ibid.) features v and vii, “natural definitions” and “*a posteriori* knowledge”, respectively.

\(^4\) This feature corresponds with Boyd’s (ibid.) feature x, which outlines the refinement of natural kind terms. I discuss Boyd’s (ibid.) iv and ix below, as I take these to be summarizations of what HPC kind terms are.
“signs and symptoms” of mental disorders. For instance, the phenomenon of major depressive disorder is defined as a cluster of symptoms which often co-occur, including depressive mood, insomnia or hypersomnia, anhedonia, fatigue, and feelings of worthlessness or guilt. Likewise, hallucinations and delusions—common symptoms of schizophrenia—regularly occur together (ibid.).

F2 is invoked by Boyd in defence of realism about natural kinds. Boyd’s argument for realism is abductive. According to Boyd, there exist property clusters that are useful for explanation, prediction, and control. The best explanation for these property clusters is that there exists something in the natural structure of the world which sustains them—namely, homeostatic mechanisms. The phrase ‘homeostatic mechanism’ appears to have two slightly different implications for Boyd. First, a mechanism is homeostatic if it is able to explain the co-occurrence of phenomenal properties in the property cluster. Second, a mechanism is homeostatic if it maintains a stable cluster of properties by ensuring that deviations from the cluster have a low chance of persisting.\(^5\)

F3 is more difficult to interpret. Boyd says that the homeostatic clustering of properties in a property cluster is “causally important.” That is, “there are (theoretically or practically) important effects which are produced by a conjoint occurrence of (many of) the properties in F together with (some or all of) the underlying mechanisms in question” (1989, 16; 1999, 143). It is not entirely clear what Boyd intends by this. As

---

\(^5\) It should be noted that, while Boyd often employs the notion “underlying mechanism”, he also allows that a “homeostatic mechanism” might be defined by causal relations among the phenomenal properties themselves. Thus a homeostatic mechanism need not be “underlying” or “lower-level.” It might involve etiological causes of a property cluster; it might also involve causes at different levels of explanation (for instance, evolutionary, genetic, psychological, physiological, behavioral, social, etc). For example, in the case of species, the homeostatic mechanisms may be intrinsic (e.g. gene exchange within a population or developmental factors) or extrinsic (e.g. a species’s evolutionary niche provides common selective factors).
Craver (2009, 578) points out, “causally important” may mean either that (i) HPC kinds figure in generalizations that are important for our own aims or objectives, or that (ii) the property cluster itself is “causally relevant” to something (i.e. it makes some difference).

The first notion of “causal importance” clearly introduces conventional elements, inasmuch as HPC kinds are recognized partly on the basis of whether they are deemed by some scientist as “important” phenomena. The second notion of “causal importance” rests on the putatively more objective notion of “causal relevance.” An inclusive definition of “causal relevance”, however, will certainly recognize more kinds than science will ever find useful.

I am inclined to side with the first interpretation, for two reasons. First, Boyd himself neither introduces nor defends the notion of “causal relevance” here. Second, Boyd argues elsewhere (Boyd, 1999a, 148, 157-162) that the recognition of a natural kind is “discipline relative.” That is to say, whether or not some kind is deemed “real” consists in part on the contribution that that kind-concept makes in accommodating our inductive practices to the relevant causal factors that sustain them. The implication of this—with respect to classification—is that there is no one unique way to classify some phenomenon of interest.\(^6\) What is more, natural kinds, on this account, can be found both within and outside of scientific practice. Thus our ordinary language category “lily” (a bulbous, flowering plant with large trumpet-shaped flowers) is a natural kind category (in the HPC sense of ‘natural kind’) even if it does not map onto the botanists’ category “Liliaceae”, which also includes onions and garlic. According to Boyd, the term ‘lily’ is useful for

---

\(^6\) Indeed, Boyd agrees with this aspect of Dupré’s (1981) “promiscuous realism”. See ibid., 160-161.
gardeners and horticulturists in achieving particular horticultural and aesthetic aims precisely because it is accommodated to causal factors that are relevant to the growing and cultivation of such a plant (i.e. the causal factors that serve to sustain the cluster of properties—e.g., aesthetic properties—that one is interested in). In any case, I return to the notion of ‘accommodation’ in the next section.

F4 is introduced to distinguish HPC kinds from essentialist kinds. With respect to F4, Boyd (1989) notes that “property cluster definitions” do not provide necessary and sufficient conditions; rather, they provide only sufficient conditions (viz. the possession of an adequate number of properties) for something to fall within the extension of a term. As such, HPC kind terms are characteristically vague and imprecise. An HPC kind term \( t \), however, does refer to things that display the homeostatic clustering of most (but not all) properties in a property cluster. Thus, for Boyd, HPC kind terms refer to kinds of things that are more-or-less stable inasmuch as they share an imperfect homeostasis.

F5 is introduced by Boyd to highlight the \textit{a posteriori}, theoretical nature of HPC kind terms. According to Boyd, HPC kinds are indeterminate as a consequence of extensional vagueness. Since members of an HPC kinds do not have all of their properties in common, it is a difficult theoretical question which properties or homeostatic mechanisms belong in the definition of an HPC kind, \( k \). Nevertheless, determining which phenomenal and causal properties provide the natural definition for a kind term, \( t \), (if this can be determined at all) is an \textit{a posteriori}, theoretical question, not a conceptual one. Contrast this with essentialist definitions, which—at least on one account—are supposed
to be analytic even if discovered *a posteriori* (Cf. Kripke, 1980; Putnam, 1975b).

Craver (2009) argues that F6 is susceptible to both a weak and strong reading. The strong reading implies that “the correct taxonomy of natural kinds must be immune to revision in light of future discoveries about the mechanistic structure of the world” (ibid., 579). That is to say, a correct taxonomy of natural kinds should reflect the natural structure of the world by ignoring *causally irrelevant* differences between members of a kind and recognizing *causally relevant* differences between members of a kind. The weak reading implies that a taxonomy of natural kinds simply recognize “sets of things that are similar enough for our explanatory or instrumental projects” (ibid.). For instance, for botanists, tomatoes are fruits; for culinarians, tomatoes are vegetables. In short, any taxonomy of kinds will depend on which properties one considers to be part of the definition of a kind. This, in turn, will depend on the purpose for which the definition itself is being used.

In my view, both of these interpretations are incorrect. While it is clear that, for other reasons, Boyd is committed to the weak reading (a point which I will argue below), in this particular passage Boyd is concerned here with naturalism, not regimenting classification. These are Boyd’s own words:

> No refinement of usage *which replaces t by a significantly less extensionally vague term* will preserve the naturalness of the kind referred to. Any such refinement would either require that we treat as important distinctions which

---

7 This does not dismiss the likelihood that it is difficult, epistemically speaking, to determine which properties belong in the definition of essentialist kinds such as “water.” The point seems to be—or at least this is how I read Boyd—that the ontological nature of HPC kinds makes this determination more difficult. After all, a member of an HPC kind differs from each and every other member in some way or another. Thus, with regard to HPC kind membership, one must determine which of these differences do or do not make a difference. As Boyd puts it: “In such cases, the relative importance of the various properties in F and of the various mechanisms in determining whether the thing falls under t—if it can be determined at all—is a theoretical rather than a conceptual issue” (Boyd, 1989, 16).
are irrelevant to causal explanation or to induction, or that we ignore similarities which are important in just these ways (Boyd, 1989, 17; emphasis mine).

Nothing in this passage suggests a normative constraint on taxonomy. The point here seems to be that revising our taxonomy of kinds in order to remedy extensional vagueness or indeterminacy would be at the expense of preserving the naturalness (viz. realness) of the kind in question.

This interpretation is made more clear, I think, from Boyd’s discussion of property cluster definitions just prior to discussing the characteristics of HPC kinds (see especially the final paragraph of ibid., 15). Here Boyd notes that property cluster definitions have been defended by the ordinary language philosophical tradition as being perfectly appropriate and necessary for ordinary linguistic usage. Boyd, however, thinks that such definitions are also appropriate and necessary, mutatis mutandis, for linguistic usage in the sciences. The reason for this is that, as he states, “there are a number of scientifically important kinds, properties, relations, etc. whose natural definitions are very much like the property-cluster definitions postulated by ordinary-language philosophers” (ibid., 15). This is particularly evident in the special sciences which study “complex structurally or functionally characterized phenomena” (ibid., 16). Boyd uses biological species as the paradigm case of HPC kinds throughout his writings (see, for example, ibid., 17-18; Boyd, 1991, 142). As a consequence of evolution (specifically, heritable variation), he says, members of a species imperfectly share homeostatically-related morphological, behavioral, and physiological features. That is, the extension of species terms is necessarily indeterminate as there exist populations which are intermediate between two
Thus, Boyd concludes, to employ only essentialist definitions in order to remedy such indeterminacy would be “scientifically inappropriate and misleading” (Boyd, 1989, 18; see also: Boyd, 1991, 142).

Finally, Boyd introduces F7 in an attempt to account for the variability of a kind itself. After all, if, over time and space, the property cluster and homeostatic mechanisms of a kind change (as F4 and F5 entail), then in what respect is it the same kind? Boyd’s answer to this is that HPC kinds are not individuated extensionally but historically. He explains this in the context of species:

[T]he property cluster and homeostatic mechanisms which define a species must be individuated non-extensionally as a process-like historical entity. This is so because the mechanisms of reproductive isolation which are fundamentally definitional for many sexually reproducing species may vary significantly over the life of a species. Indeed, it is universally recognized that selection for characters which enhance reproductive isolation from related species is a significant factor in phyletic evolution, and it is one which necessarily alters over time the species’ defining property cluster and homeostatic mechanisms (ibid., 18; emphasis mine).

The point here seems to be that, while a species may vary across time and space, both in terms of its “defining property cluster and homeostatic mechanisms”, it is still the same species. According to Boyd, this is a necessary feature of HPC kinds in accounting for species, since species evolve across time and space.

3.2.2 HPC Kinds in the Taxonomy of Natural Kinds

In the taxonomy of natural kinds Boyd’s HPC view is most often presented as a third way between skepticism and essentialism about natural kinds (Boyd, 1989, 1991, 1999a; Craver, 2009; Griffiths, 1999; Kornblith, 1993). The HPC framework is non-essentialist inasmuch as it neither restricts natural kinds to kinds defined by internal,
microstructural essences (e.g., the atomic structure of gold or the DNA of a particular species), nor does it require that each and every member of an HPC kind have all of their properties in common. Boyd is also a realist about natural kinds, and distances himself from skepticism by maintaining that the property clusters indicative of natural kinds are the consequence of similarity-generating (or “homeostatic”) mechanisms in nature, not merely the way human beings think about the world. As others have pointed out (Cf. Craver, 2009, 578; Griffiths, 1999, 218) homeostatic mechanisms on Boyd’s account are supposed to serve the same explanatory purpose as essences do on the traditional account (Cf. Boyd, 1999b, 218).

In light of what has been said so far, it would seem that a term, for Boyd, denotes a natural kind when it picks out a cluster of properties which are held together by one or more known casual mechanisms.8 Things are not quite so simple, however. As I mentioned earlier (note 6), Boyd also aligns himself with Dupré’s doctrine of promiscuous realism, and this presents some difficulty for interpreting the HPC theory of natural kinds. As you will recall from chapter 1, the only condition that the promiscuous realist places on natural kinds is that individual members of a kind are similar enough for our explanatory or instrumental projects. As I argued in chapter 1, however, mere similarity is not sufficient for establishing that something is a real as opposed to an artificial kind. Individual members of the putative kind depression, for instance, might be similar with respect to some of their signs and symptoms, nevertheless differing in some more fundamental way (e.g., one individual is reacting to loss whereas another is

---

8 This rendering of Boyd’s account would also place the HPC theory of natural kinds alongside the simple causal theory described in chapter 1 (see, especially, section 1.4.3).
suffering from some brain abnormality). While depression would count as a natural kind on Dupré’s account, the assertion that depression is a natural kind would seem to belie Boyd’s point that natural kinds are the consequence of similarity-generating (or “homeostatic”) mechanisms in nature, as there is no reason at present to think that individual members of the putative kind depression are anything more than superficially similar.

Boyd’s commitment to both realism and pluralism about natural kinds leaves us with at least two possible versions of the HPC theory of natural kinds, both of which have implications for the ontological status of psychiatric kinds qua HPC kinds and the status of the HPC framework as a usefulness model for psychiatric classification. The former, stronger reading implies that a correct taxonomy of natural kinds will identify clusters of properties which are homeostatically-related. For example, if schizophrenia is an HPC kind, then we would expect there to be a causal relationship between the observable properties of schizophrenia (e.g., delusions, hallucinations, etc.) and the properties (e.g., excess dopamine) which produce them. If, on the other hand, the observable properties of schizophrenia turn out to be produced by distinct causal mechanisms in different individuals, then a correct taxonomy should recognize that difference. The latter, weaker reading implies that a taxonomy of natural kinds which picks out sets of things that are similar enough for our explanatory or instrumental projects is sufficient to merit the label ‘natural’. If schizophrenia is a useful category for clinicians, then there is no need to nitpick over the details. It all depends on what one is trying to do.

In the next section I argue that Boyd is in fact committed to the weak version of
HPC kinds. On Boyd’s view, a category is a natural kind just in case it is accommodated to the causal structure of the world *in virtue of* being accommodated to the inferential practices of the researcher. The implication of this for classification is that categories on the HPC account are revised in accordance with the instrumental objectives of the researcher, first and foremost, and only secondarily to the causal structure of the world. This presents a dilemma for philosophers of psychiatry who would commit themselves to Boyd’s account of natural kinds. As I intend to argue in section 3.4, the argument that paradigmatic mental disorders such as schizophrenia and depression are HPC kinds only goes through if we presuppose Boyd’s version of HPC kinds (particularly Boyd’s accommodation thesis, which I discuss below). Boyd’s accommodation thesis, however, is too weak to serve as an objective arbiter of classification. This conclusion should be unpalatable to philosophers of psychiatry who hold *both* that psychiatric kinds are HPC kinds, *and* that HPC kinds are useful models for psychiatric classification. Far from being “useful models for psychiatric classification”, HPC kinds place no constraints whatsoever on classification.

This consequence of the HPC theory should not be surprising given a close reading of Boyd’s account of HPC kinds, however. As we will see in the next section, the HPC theory of natural kinds was not developed for the purpose of providing a definition of natural kinds to which our classifications ought to aspire—rather, Boyd’s theory is a post-hoc explanation of the success of our inductive practices. In particular, Boyd wanted to offer an explanation of how some of our concepts can be deemed “natural” despite being heterogenous and vague.
3.3 Boyd’s Accommodation Thesis

Boyd’s accommodation thesis is an aspect of the HPC account that is rarely touched upon in the philosophy of psychiatry literature. This is unfortunate, since, as I will argue in the following, it is clear that this thesis commits Boyd to the weak reading of HPC kinds, and this significantly loosens the prescriptive force of the HPC theory for classification. What is more, philosophers of psychiatry who assert that some mental disorders are HPC kinds, or that the HPC framework is a useful model for psychiatric classification, appear to be implicitly committed to this thesis.

3.3.1 HPC Kinds and Accommodation

A superficial reading of Boyd’s HPC theory of natural kinds would seem to suggest that HPC kinds are individuated by the recognition of both a property cluster and the homeostatic mechanism(s) which explain the clustering of said properties. This, however, is not Boyd’s view. In response to the question, “how are we to tell which of our categories are natural kinds?”, Boyd contends that questions concerning the natural kind status of our categories reduces to questions about how well our categories are accommodated to inductive practices. In defending his “accommodation thesis”, Boyd says:

It is a truism that the philosophical theory of natural kinds is about how classificatory schemes come to contribute to the epistemic reliability of inductive and explanatory practices. Quine was right in “Natural Kinds” (1969) that the theory of natural kinds is about how schemes of classification contribute to the formulation and identification of projectible hypotheses (in the sense of Goodman 1973). The naturalness of natural kinds consists in their aptness for induction and explanation... (Boyd, 1999a, 147; emphasis mine).
According to Boyd, explanatory and inductive success is the best guide to the naturalness of our categories. The best explanation of this success, moreover, is that there is an “accommodation between conceptual and classificatory practices and causal structures” (Boyd, 1999a, 142). Thus, our categories are “natural” inasmuch as they are accommodated to the inferential practices of the researcher, objectives which are assumed by Boyd to be accommodated to the causal structure of the world.

But in what sense do classificatory schemes “contribute to” the reliability of our inductive practices? Boyd has a ready answer for this:

[W]e are able to identify true generalizations in science and in every-day life because we are able to accommodate our inductive practices to the causal factors that sustain them. In order to do this—to frame such projectable generalizations at all—we require a vocabulary, with terms such as sodium salt and flame, which is itself accommodated to relevant causal structures. This is the essence of the accommodation thesis regarding theoretical natural kinds (ibid., 148; emphasis mine).

According to Boyd, classificatory schemes “contribute to” the reliability of our inductive practices inasmuch as the success of our inductive practices is contingent upon the utilization of natural categories—i.e. categories that are accommodated to the causal structure of the world. The picture is something like this: in scientific practice we postulate a category. Many categories will not be projectable. Indeed, we can postulate categories that will not be projectable despite fitting all past data—e.g., grue and bleen (Cf. Goodman, 1983). Presumably we revise, or eliminate, such categories. Eventually, however, some provisional categories will prove to be successful inasmuch as they allow us to make successful predictions. Suppose, for instance, that we begin conducting laboratory experiments on samples of a substance that we have provisionally termed
‘phosphorus’. In so doing we find that each and every sample of phosphorus reacts similarly under similar laboratory conditions. That is, we are able to predict how a sample of phosphorus will react when placed under various laboratory conditions. Why are we able to make such predictions? According to Boyd, such predictions are possible precisely because the term ‘phosphorus’ is accommodated to the causal structure of the world.

One must recognize, however, that for Boyd natural kind concepts do not themselves provide conditions for individuating a kind. That is to say, we do not know that a concept is a natural kind concept because we have some prior idealized definition of natural kind that our current concept conforms to. This interpretation would be completely at odds with Boyd’s naturalism discussed in the preceding section (see, especially, my rendering of Boyd’s F5 and F6). Rather, we know that a concept is a natural kind concept when we are able to wield the concept in the service of making projectable generalizations. The fact that we are able to make projectable generalizations at all is explained by our concepts being accommodated to the causal structure of the world.

Hence, with respect to our classificatory schemes, we know that we have a natural classification when the categories that make up a classification can be put in the service of making projectable generalizations. Again, the fact that we are able to make projectable generalizations at all is explained by our categories being accommodated to the causal structure of the world. In short, Boyd’s HPC kinds are explanatory; they are able to explain why a classification is successful, but do not themselves provide the resources for individuating kinds.
It follows from Boyd’s accommodation thesis that there are many different ways to classify phenomena while still maintaining the “naturalness” of one’s classification. Indeed, as I mentioned earlier, Boyd agrees with this aspect of Dupré’s “promiscuous realism” (see Boyd, 1999a, 160-161). After all, if one can make projectable generalizations using one’s categories, then they are able to do so only insofar as their categories are accommodated to causal features of the world. If their categories are accommodated to causal features of the world, then they are “natural”. Hence, if one can make projectable generalizations using one’s categories, then they are in this sense “natural kinds”.

Boyd’s accommodation thesis also entails that natural kind classifications are not limited to scientific practice, but are central to many ordinary practices. The example Boyd uses is our ordinary language category ‘lily’. The category ‘lily’ is a natural kind category, even if it does not include onions and garlic (as the the botanists’ category ‘Liliaceae’ does). According to Boyd, the term ‘lily’ is useful for gardeners and horticulturists in achieving particular horticultural and aesthetic aims precisely because it is accommodated to causal factors that are relevant to the growing and cultivation of such a plant. Likewise, for botanists, tomatoes are fruits; for culinarians, tomatoes are vegetables.

The accommodation thesis, as far as I can see, is employed by Boyd for the purpose of preserving a naturalistic interpretation of some of our scientific and extra-scientific categories while at the same time avoiding thorny metaphysical questions about kindhood. Boyd, it would seem, is sceptical that we can know which similarities do or do
not—metaphysically speaking—underpin kinds. This is clear from the explanatory nature of HPC kinds and the fact that Boyd maintains that it is often difficult to determine theoretically (if at all) which properties belong to the definition of a kind. More importantly, however, Boyd agrees with Dupré’s view that our natural kind categories are conventional inasmuch as they are relative to the use to which they are put (see, especially, Boyd, 1999a, 158-162). As Dupré (1981, 82) argues, there are many “sameness relations that serve to distinguish classes”, none of which are privileged. The “sameness relations” or “similarities” that serve to underpin a “kind” are ultimately contingent on the relevance that that kind-concept has to some inductive or explanatory goal.

3.3.2 Weak and Strong Accommodation

The upshot of Boyd’s accommodation thesis regarding natural kinds is that a category need not “cut nature at the joints” or reflect the causal structure of the world precisely in order to be considered a natural kind. In fact, if we take Boyd’s accommodation thesis seriously, the only condition that Boyd places on natural kind-concepts is that they be projectable. The projectability of a concept, for Boyd, ensures that it is connected to to the causal structure of the world in some way. Call this “weak accommodation”.

Natural kinds as Boyd understands them are a far cry from the definition of natural kinds that I’ve been working with thus far. Take, for instance, the kind dementia discussed briefly in chapter 1. On Boyd’s account, dementia would be considered a natural kind. After all, if an individual is diagnosed with “dementia”, then we can fairly
confidently predict their general course and prognosis—namely, progressive cognitive decline. The category of dementia, in other words, is projectable. Be that as it may, no serious scientist would consider dementia a kind unto itself. Individuals members of the putative kind *dementia* are at best superficially similar—they possess no shared causal structure that sufficiently explains their shared symptomatology. Indeed, it is now well-recognized that the term ‘dementia’ is a catch-all for different kinds of diseases which affect cognition, such as frontotemporal dementia, Alzheimer’s disease, Huntington’s disease, etc.

By all accounts the newer concepts of frontotemporal dementia, Alzheimer’s disease, and Huntington’s disease more accurately reflect the way the world is than the nineteenth century concept of dementia. To put this in the context of Boyd’s accommodation thesis, the former concepts are strongly accommodated to the causal structure of the world (call this “strong accommodation”), whereas the latter concept of dementia is at best weakly accommodated. Boyd’s HPC theory of natural kinds, however, makes no distinction between weak and strong accommodation, and thus places the concept of dementia ontologically on a par with the concepts of frontotemporal dementia, Alzheimer’s disease, and Huntington’s disease.

This is an unintuitive consequence, one that is particularly problematic for philosophers of psychiatry who claim that revising psychiatric kinds in line with the HPC framework would be an improvement on current classification. In medicine, at least, heterogeneous concepts like dementia are the kinds of concepts that we are trying to avoid. To conclude that dementia is a natural kind, and that we should revise our
categories in such a way as to reflect natural kinds in this sense, would be would be the rough equivalent of concluding that *cough disorder* is a natural kind, and that our classifications of other kinds of disease should resemble that of “cough disorder”.

In the next section I contend that, in their attempts to establish that some mental disorders are natural kinds in the HPC sense of ‘natural kind’, this is precisely the kind of absurd consequence that some philosophers of psychiatry have unwittingly committed themselves to.

### 3.4 Psychiatric Kinds and Homeostatic Property Cluster Kinds

Philosophers of psychiatry generally agree that, as with species, mental disorders are not good candidates for the essentialist model of natural kinds (see, especially, Kincaid & Sullivan, 2014). According to Kendler et al. (2011), essentialism cannot capture the very real, yet complex and “fuzzy” nature of psychiatric disorders. First, essentialism entails that a common essence is directly and causally responsible for the defining features of a particular psychiatric disorder. While the essentialist model of natural kinds may be able to account for static elements such as gold, it is unlikely that essentialism can capture the nature of complex biological phenomena. For instance, Kendler et al. (2011) argue that “evolving organisms vary too much from one to the next...for the concept of an ‘essence’ to be very useful” (ibid., 1144). More specifically, many illnesses (including psychiatric illnesses) arise “from a wide range of genetic, metabolic, behavioral and environmental risk factors” (ibid), rather than from the single, simple causal agent assumed by essentialism. In other words, psychiatric disorders most likely have a variety of different causes related to their manifest signs and symptoms.
In light of these considerations, it has become increasingly common for philosophers of psychiatry to defend the HPC view of natural kinds as a way of conceptualizing and classifying mental disorders. Kendler et al. (2011), for instance, contend that psychiatric disorders should be seen as “mechanistic property cluster (MPC) kinds”, which are similar to Boyd’s HPC kinds. On this view, psychiatric disorders are seen as sharing “stable similarities” both among their clinical features and among their underlying mechanisms and etiology. The MPC framework, it is argued, is the preferable guide for psychiatric classification, as it provides a way for tying psychiatric classification to causes, which would be in line with much of the rest of medicine.

Similarly, in his dissertation and a series of articles (Tsou, 2008, 2012, 2013, 2016) Jonathan Tsou has contended that at least some mental disorders are HPC kinds, and that psychiatric classification should classify them as such in order to provide more useful diagnostic categories.

In the following I show how the argument that paradigmatic mental disorders such as schizophrenia and depression are HPC kinds only works if one presupposes Boyd’s version of HPC kinds—namely, that the concepts of ‘depression’ and ‘schizophrenia’ are weakly rather than strongly accommodated to the causal structure of the world. If this is right, then philosophers of psychiatry cannot consistently hold both that psychiatric kinds are HPC kinds, and that HPC kinds are useful models for psychiatric classification. First, I consider Tsou’s argument for the conclusion that depression and schizophrenia are HPC kinds. I then address Kendler et al.’s suggestion that mental disorders ought to be conceptualized as MPC kinds.
3.4.1 Are Mental Disorders HPC Kinds?

Tsou has been the most prominent defender of the view that some mental disorders are HPC kinds. In particular, Tsou contends that the neurobiological mechanisms that underlie the characteristic signs and symptoms of schizophrenia and depression are best able to account for the fact that we can make successful inferences about members of these kinds. That is to say, the reason that we can predict the prognosis, progression, and treatment response of an individual diagnosed with a particular mental disorder is precisely because the signs and symptoms of the disorder in question are constituted by stable neurobiological mechanisms.

Throughout his writings Tsou is not always clear about whether he has in mind the strong or weak version of HPC kinds, which makes interpreting his claim that depression and schizophrenia are natural kinds particularly difficult. The argument that depression and schizophrenia are HPC kinds in the sense that members of these kinds have a similar biological causal structure would seem to presuppose at the very least that members of the same HPC kind must have in common important causal properties, which is consistent with the strong version of HPC kinds and the general conception of natural kinds presented in chapter 1.9 In fact, however, in several places Tsou appears to make an even stronger claim. Unlike myself, Tsou considers Hacking’s argument from looping (see section 2.2.2) a considerable challenge to the view that mental disorders are natural kinds. While Tsou does not doubt that kinds such as depression and schizophrenia are subject to

---

9 Indeed, Tsou often contrasts natural kinds of mental disorder such as depression and schizophrenia with what he considers to be artificial kinds of mental disorder such as narcissistic personality disorder, hysteria, and multiple personality disorder on the basis that research on the latter “consistently fail to yield any information on the etiology or causal structure of such classifications” (Tsou, 2008, 63).
looping effects, he contends that depression and schizophrenia remain stable in spite of looping effects—and hence, are natural kinds—because “they are constituted by a set of stable biological mechanisms (e.g. neurobiological mechanisms)” (Tsou, 2016, 412). This would seem to imply that—at least as far as mental disorders are concerned—members of the same natural kind must have a shared biological structure.

Nevertheless, Tsou consistently qualifies his claim that some mental disorders are natural kinds by invoking Dupré’s doctrine of promiscuous realism or pluralism (Cf. Tsou, 2008, 68; 2013, 463; 2016, 410). Specifically, Tsou says that, in line with Dupré’s pluralistic approach, he is only articulating “one sense in which some mental disorders can be understood as natural kinds” (Tsou, 2008, 68; emphasis mine)—i.e. one sense in which members of, say, schizophrenia are “relevantly similar to one another with respect to some natural (or ‘real’) properties” (Tsou, 2016, 463). This would seem to imply that members of a kind need not have a shared biological structure—or, indeed, a shared causal structure at all—to be considered a natural kind. After all, as I have already discussed at length, if we adopt Dupré’s pluralistic approach to natural kinds, then it is sufficient that members of natural kind have some properties in common, irrespective of whether or not they are causal.

In light of Tsou’s seeming commitment to different, incompatible versions of natural kinds we are left with the following question: are depression and schizophrenia.

---

10 Tsou uses this same argument throughout his writings (Cf. Tsou, 2007; 2008, 74-79; 2013, 462-463). As I pointed out in chapter 2 (section 2.3), Hacking had already anticipated this objection in his The Social Construction of What (1999). Here Hacking concedes to the possibility that in some instances looping might only alter the superficial or stereotypical properties of a human kind, leaving the important properties unchanged. For instance, Hacking posits that the putative kinds schizophrenia and childhood autism, though subject to looping effects, might be natural kinds in the sense that they are ultimately produced by some underlying “biological pathology P” (ibid., 119), which itself is supposed to be not subject to looping effects.
HPC kinds—and hence, natural kinds—in virtue of (i) a shared biological structure, (ii) a shared causal structure, or (iii) a shared set of natural properties which are projectable? In what follows I argue that Tsou’s claim that depression and schizophrenia are natural kinds is only plausible if we assume (iii).

With respect to (i), one should first of all be skeptical that having a shared biological structure is a necessary condition for a mental disorder to be considered a natural kind. As I have already argued in chapter 2, Hacking’s looping effect argument does not entail that mental disorders cannot in principle be natural kinds. If this is right, then Tsou’s appeal to “stable biological mechanisms”, mechanisms which are assumed not to be subject to looping effects, is superfluous, since even properties which are subject to looping effects can, in principle, be constitutive of a natural kind. Unless we are prepared to argue that biological properties are somehow “more real” or “more natural” than, say, psychological, social, or environmental properties\textsuperscript{11}, then we cannot \textit{a priori} discount the possibility that a mental disorder is a natural kind in virtue of its members having, for example, a shared psychological causal structure—or, indeed, a shared causal structure which includes some combination of chemical, neurological, psychological, social, or environment properties.

Even if it is not \textit{necessary} that members of the same natural kind of mental disorder share the same biological causal structure, having a shared biological causal structure is certainly \textit{sufficient} for being considered a natural kind in the sense of (ii)—and, indeed, in the sense described in chapter 1. But what is the shared causal structure of

\textsuperscript{11} I already tried to dispel this line of argumentation in chapter 1 when discussing microphysical properties (see, especially, section 1.4.3 and note 30 in chapter 1).
depression and schizophrenia?

According to Tsou, “multiple lines of research”\textsuperscript{12} support the “well-confirmed” hypothesis that the positive symptoms of schizophrenia (viz. delusions and hallucinations) are caused by excessive dopamine in the mesolimbic pathway in the brain. As well, for Tsou, pharmacological research shows that depression is caused by deficient activity of serotonin and norepinephrine (monoamine neurotransmitters). Hence, “mental disorders that are MPC kinds, such as schizophrenia and depression, possess a distinctive and stable biological causal structure” (Tsou, 2016, 412). This biological structure (or, in Boyd’s terminology, ‘homeostatic mechanism’) is able to explain the co-occurring properties of schizophrenia, as well as why we are able to make successful predictions about individuals who are diagnosed with schizophrenia.

As I argued in chapter 2, the ‘monamine hypothesis’ of depression is severely underdetermined by evidence. At best, serotonin (and other monamine neurotransmitters) are involved in depression in some way. This, however, is a far cry from Tsou’s claim that deficiencies in monamine neurotransmitters explain how it is that depression comes about. In fact, most researchers have been forced to admit that older theories of depression that merely focus on abnormalities in neurotransmitter systems are far too simplistic.

Furthermore, when we take into account the broader range of evidence concerning the potential causes of depression, it would appear that there are many different ways that one might get depressed—i.e. many different causal processes that produce the

\textsuperscript{12} But mostly pharmacological research on antipsychotics and antidepressants. See, e.g., Tsou, 2008, 84-89; Tsou, 2012; Tsou, 2016, 411-412.
phenomenon of “depression” as we currently conceive it. Even if we restrict our analysis to the brain—the organ which presumably generates all of our perceptions, actions, thoughts, feelings, and emotions—there is at present no evidence that individual instances of the putative kind depression share a causal structure. In fact, as I have already argued in the previous chapter, there is some evidence to the contrary.

Consider also schizophrenia. Evidence for the dopamine hypothesis of schizophrenia has come from two separate observations. First, the use of drugs that increase dopamine levels (e.g., amphetamines, cocaine) can produce some of the positive symptoms of schizophrenia, such as complex hallucinations and paranoid delusions. This is most often referred to in the scientific literature as “amphetamine psychosis”. Second, typical antipsychotic medications work by not allowing dopamine to bind to the D₂ dopamine receptor. In particular, medications which successfully block 65-80% of D₂ receptors in the ventral striatum appear to significantly reduce the positive symptoms of schizophrenia (Howes et al., 2009).

Yet, only one third of schizophrenic patients respond favourably to antipsychotic medication (Cf. Harvey et al., 2007). Recent evidence also suggests that other neurotransmitter systems are involved in schizophrenia. Cannabis use, for instance, significantly increases the risk of schizophrenia, especially if used in adolescence.

---

13 The five (DSM-5) criteria for schizophrenia are as follows: (1) delusions, (2) hallucinations, (3) disorganized speech (e.g., incoherent speech or “word salad”), (4) grossly disorganized or catatonic behavior, (5) negative symptoms (e.g., apathy, avolition, social withdrawal, poverty of speech or thought, flat affect, anhedonia). For a diagnosis of schizophrenia, at least two of (1)-(5) need to be present, and at least one of these symptoms must be either delusions, hallucination, or disorganized speech. See APA (2013, 99).

14 Below this level, positive symptoms of schizophrenia tend to persist. Above this level, significant side effects of low dopamine transmission begin to occur—e.g., tremors, dystopias, akathisia, and Parkinsonian rigidity.
Cannabis acts on a distinct neurotransmitter system involving cannabinoid receptors, which in turn interact with many other neurotransmitters in the brain, not only dopamine. Likewise, some researchers have contended that too little glutamate, rather than too much dopamine, can account for both the positive and negative symptoms of schizophrenia (Olney & Farber, 1995).

Furthermore, as with depression, when we take into account the broader range of evidence concerning the potential causes of schizophrenia, it would appear that much more is involved than neurotransmitters. Biologically-inclined scientists, for instance, have posited more complex neurodevelopmental theories in which schizophrenia is conceptualized as a neurodegenerative disorder characterized by abnormal brain development (Rapoport, Giedd & Gogtay, 2012). These theories have been proposed not only in response to the empirical inadequacy of antiquated neurotransmitter theories, but also in response to evidence that suggests that schizophrenia is associated with a rapid pruning of cortical connections in a widespread set of brain regions such as the anterior and dorsal medial temporal lobes (van Haren et al., 2007), reduced hippocampal and frontal lobe volumes (Vita et al., 2006), as well as cytoarchitectural abnormalities in schizophrenic patients including smaller somas and decreased dendritic spines in pyramidal neurons in the prefrontal cortex (Glantz & Lewis, 2000).

Most importantly, perhaps, while Tsou assumes schizophrenia to have a stable biological basis, it’s not at all clear in what way internal neurobiological mechanisms “cause” schizophrenia. For one thing, there is also evidence that environmental factors are involved in the development of schizophrenia. For example, the risk of developing
schizophrenia is increased by way of maternal infections such as polio, influenza, rubella (Brown & Patterson, 2011); maternal stressors such as depression, grief, flood and famine (Khashan et al., 2008); and even social factors such as childhood trauma, social isolation, living in a city, and belonging to a minority ethnic group (Van Os & Kapur, 2009).

Given this, it is unclear why Tsou thinks that the positive symptoms of schizophrenia are “caused by” excessive dopamine. A cursory glance at the scientific literature suggests that the potential causes of schizophrenia are much more varied than Tsou admits. At most, one can grant that the neurotransmitter dopamine is involved in schizophrenia in some way. This is hardly a ringing endorsement for the view that excessive dopamine in the mesolimbic pathway is a “homeostatic” mechanism which sustains the clustering of the positive symptoms of schizophrenia.

Does this entail that schizophrenia is not an HPC kind? Not necessarily. As I have already said, if by HPC kind we mean that members of a kind (iii) share a set of natural properties which are projectable, then schizophrenia is most certainly a natural kind in Boyd’s weak HPC sense of ‘natural kind’. Consider the following argument:

P1. The DSM category schizophrenia allows us to make projectable generalizations.
P2. The best explanation for this is that the category ‘schizophrenia’ is accommodated to the causal structure of the world.
P3/C1. Therefore (by abduction), the concept ‘schizophrenia’ is accommodated to the causal structure of the world [by P1 & P2].
P4. HPC kinds, according to Boyd, are natural kinds inasmuch as they are accommodated to the causal structure of the world.
P5/C2. Therefore, the DSM kind schizophrenia is a natural kind in the HPC sense of ‘natural kind’.

As you will recall from chapter 1, a category is “projectable” just in case it allows us to

---

15 See also van Os, Kenis & Rutten (2010) for a review of evidence concerning environmental factors and schizophrenia.
make reliable inductive inferences. Hence, insofar as the category schizophrenia allows us to make reliable inductive inferences, it is a natural kind. Although the term ‘reliable’ is left undefined by Boyd (as well as other philosophers who take projectability to be an epistemic marker of a natural kind), I take “reliable inductive inference” here to mean that we can predict with greater than chance probability some fact about members of the category in question. For example, the fact that we can predict with greater than chance probability that any given schizophrenic will experience auditory hallucinations (in fact, the number is somewhere around 70% of individuals diagnosed with schizophrenia; see Andreason & Flaum, 1991) entails that schizophrenia is a natural kind.

Be that as it may, it is crucial to recognize how easy it is for a category to fulfill this condition. As we have just seen, schizophrenia can be considered a natural kind in the foregoing sense even in the absence of knowledge of its causes. Substitute any other DSM category for schizophrenia and what you have is the basis for an argument that could be used to defend the view that any mental disorder is a natural kind. This is an unintuitive consequence. According to critics of the DSM—including Tsou himself—most psychiatric categories are too heterogeneous and thus unlikely to inform precise diagnoses and treatments of mental disorders. What is more, it is assumed that current psychiatric categories are too heterogenous precisely because they ignore causal properties in favour of superficial properties, such as signs and symptoms. Given this, it would seem that the HPC theory of natural kinds is not a useful model for classification after all:

P6. Members of an HPC kind share a set of natural properties which are projectable, but an HPC kind is heterogeneous in the sense that members may differ significantly both in
their superficial and causal properties.
P7. According to critics of the DSM, DSM kinds (including schizophrenia) are too heterogeneous in the sense that members differ significantly both in their superficial and causal properties.
C3. Therefore, conceptualizing mental disorders as natural kinds in the HPC sense of ‘natural kind’ will be of no help in revising current mental disorder categories, if such revisions are meant to pick out more homogenous kinds [by C2 & P6, P7].

This leaves us in a bit of a quandary. If we loosen our definition of natural kinds to include categories which are merely projectable, then the assertion that mental disorder \( \mathcal{x} \) is a natural kind is rather empty, amounting to little more than a rhetorical strategy—giving “fancy names to old things”. If, on the other hand, we insist that members or instances of a natural kind share a common causal structure, then the assertion that mental disorder \( \mathcal{x} \) is a natural kind is descriptively false, since there is not enough evidence at present to know whether this condition is met (as I have argued both here and in the previous chapter).

As well, in order for the prescriptive claim that psychiatric kinds ought to reflect natural kinds to be meaningful at all, it would seem that we must presuppose at the very least that members or instances of a natural kind share a causal structure, requiring us to jettison the conclusion that paradigmatic mental disorders such as depression and schizophrenia are natural kinds.

3.4.2 Are Mental Disorders MPC Kinds?

Although Tsou often equivocates between different, incompatible definitions of “natural kind”, the pervading assumption in his writings is that mental disorders are HPC kinds—and hence, natural kinds—in virtue of having a shared biological structure. In contrast to this view, Kendler, Zachar, and Craver (2011) have argued that mental
disorders are “mechanistic property cluster (MPC) kinds”—and hence, natural kinds—in
virtue of being produced by “stable patterns of complex interaction between behavior,
environment and physiology that have arisen through development, evolution and
interaction with the environment” (Kendler et al., 1147). Psychiatric disorders, on this
view, are defined by stable similarities among properties of mind/brain states, including
interactions between “genes, cell receptors, neural systems, psychological states,
environmental inputs and social–cultural variables” (ibid).

MPC kinds are similar to HPC kinds inasmuch as they posit mechanisms that
ensure that certain properties are instantiated together. Kender et al. (2011) are not
explicit about which of Boyd’s F1-F7 they are committed to. Nevertheless, the following
three passages are informative:

Property clusters do not have simple, deterministic essences. Not all members
need overlap in some single set of traits; rather, members are clustered near
one another in a feature space because of developmental, evolutionary and
physiological causal mechanisms and constraints (ibid., 1146; emphasis
mine).

The ‘kind-ness’ of species is not, from an MPC perspective, produced by a
defining essence but rather from more or less stable patterns of complex
interaction between behavior, environment and physiology that have arisen
through development, evolution and interaction with an environment. It is this
often complex and intertwined mechanism that produces the imperfectly
shared characteristics of the members of a species. Such kinds are more
heterogeneous than elements in a periodic table. Unlike all atoms of gold,
individual members of a species need not share all their properties. Across
the range of a species, some systematic differences may arise in
subpopulations in coloration, body weight or food preference. Hybrids can
also occur. However, the fuzziness of these boundaries does not detract from
their stability (ibid., 1146-7; emphasis mine).

Members of MPC kinds are not similar merely in their superficial properties
(like all the things in refrigerators), but because the co-occurrence of these
properties from individual to individual is explained by causal mechanisms that regularly ensure these properties are instantiated together. Indeed, MPC kinds are useful for prediction, explanation and control precisely because the kinds are sustained by causal mechanisms. Such clusters allow us to make projective inferences about the past, present and future on the basis of an item’s membership in a kind (ibid., 1147; emphasis mine).

The first and third passages commit Kendler et al. (2011) to Boyd’s F1 and F2. While Kendler et al. employ the phrase ‘mechanistic property clusters’ rather than ‘homeostatic property clusters’, the implication is the same: property clusters are explained by the mechanisms which produce them.

The second passage entails Boyd’s F4. As Kendler et al. note, the boundaries of an MPC kind are “fuzzy”; members of an MPC kind “imperfectly share” characteristics; and MPC kinds are “heterogeneous.” In other words, members of an MPC kinds have some but not all of their properties in common (extensional vagueness). As such, it will not always be clear whether or not some thing is member of some kind, $k$ (indeterminacy). Kendler at al. later describe the fuzzy or heterogeneous nature of MPC kinds by making reference to their prototypical structure (see ibid., 1148). According to Kendler et al., one can expect a wide range of members of a kind such that there are members of a “central”, “large” (the terms ‘central’ and ‘large’ are left vague and indeterminate on their account) group who share most of the characteristics of a kind (such as the putative kind “schizophrenia”), and members of outlying groups who share less of the features of the kind in question.

Kender et al. do little to address the other features of Boyd’s account or his accommodation thesis. This is unfortunate, since it is not clear from their description of
MPC kinds whether they think that psychiatric categories are, or ought to be, weakly or strongly accommodated to the causal structure of the world. For instance, the passages just quoted would seem to imply the strong view—namely, that MPC kinds are natural kinds in virtue of members sharing a causal structure, albeit complex. Nevertheless, Kendler et al. also allow that “the same cluster of symptoms might arise from different etiological, underlying or sustaining mechanisms in different cases” (ibid., 1148). This would seem to imply that members of the same MPC kind need not share a causal structure after all.

Yet, if MPC kinds are just as heterogeneous as DSM kinds, then it is unclear in what way MPC concepts are more useful for classification than their DSM counterparts. Indeed, the inference from ‘mental disorders are MPC kinds’ to ‘classifying mental disorders as MPC kinds will provide more useful diagnostic categories’ is a non sequitur; it is based on the false assumption that MPC kinds are less heterogenous than DSM categories. What is more, the paradigm examples of MPC kinds given by Kendler et al. (and, indeed, Tsou) are all examples of heterogeneous DSM kinds, which makes this argument all the more perplexing.

3.4.3 Summary

In the foregoing I have argued that paradigmatic mental disorders such as schizophrenia and depression are HPC kinds only works if one presupposes Boyd’s version of HPC kinds—namely, that the concepts of ‘depression’ and ‘schizophrenia’ are weakly rather than strongly accommodated to the causal structure of the world. Boyd’s weak accommodation thesis, however, is too weak to serve as an objective arbiter of
psychiatric classification, since it permits significant heterogeneity both at the level
phenomenal and causal properties. It would seem, then, that if philosophers would like to
maintain that HPC kinds are “useful models for psychiatric classification”, they must
commit themselves to a stronger view of natural kinds.

3.5 The Reality of Mental Disorders

In chapter 1 (section 1.4.1) I drew a distinction between Mill’s understanding of
natural kinds and that of Russell, Quine, and Dupré. The latter authors, I suggested,
understand natural kinds to be merely rough-and-ready categories which collect together
individuals who are loosely similar. Quine, for instance, thinks that natural kind
categories are grounded on intuitive or innate judgements of similarity, and will be
replaced in the course of scientific inquiry. Conversely, for most natural kinds realists,
natural kinds are clusters of properties which cluster together because they have a
common causal structure. As Khalidi (2013, 46) has argued, natural kinds are the kinds of
categories that come at the end of scientific inquiry, not the beginning.

In my mind, current kinds of mental disorder are at best the kinds of categories
that come in the middle of scientific inquiry. As I have suggested both here and in the
previous chapter, individuals who are diagnosed with a DSM mental disorder are similar
in many respects. For instance, individuals diagnosed with depression often have most of
their observable properties in common (see section 2.3.1). What is more, many DSM
categories are projectable in the sense that they enjoy moderate predictive success. In light of this, psychiatric kinds are undoubtedly more “real” or “natural” than Mill’s class of white things, postal codes, telephone area codes, and the like. Be that as it may, they are not on a par with chemical kinds, biological species, and many kinds of physical and neurological disease. The reason for this is not, as is commonly misunderstood, that they have no essence (Cf. Zachar, 2002; 2014) or underlying biological cause (see section 3.4.1 above). The reason for this is simply that they do not (so far as we know) have unobservable causal structure which explains why the observable properties appear to co-occur.

In light of what has been said, it is best to think of kinds of mental disorder as both “hypothetical constructs” and “useful heuristics”. By “useful heuristic” I mean that psychiatric categories provide us stable “objects” to observe and study, and in some instances might even be both construct valid or predictively valid (see note 16). Be that as it may, they remain “hypothetical constructs” rather than “real kinds” because there is at present no convincing evidence to suggest that the observable properties of kinds of mental disorder are held together by an unobservable causal base.

16 Scientists often distinguish between construct validity and predictive validity to describe psychiatric categories. Traditionally, the requirement for a psychiatric concept to be considered “construct valid” is that either the observable features of the kind-concept are empirically inter-correlated (i.e., they constitute a “syndrome”) or that there is a well-established theoretical framework within which the concept is embedded. The requirement for a concept to be considered “predictively valid” is that it should enjoy predictive success (e.g., predicting that schizophrenia runs in families or that schizophrenia is responsive to certain sorts of treatments). In my mind, even if a psychiatric concept meets the requirements for both construct and predictive validity, this does not imply that it is a natural kind. First, the fact that the observable features of a kind are empirically inter-correlated does not necessarily imply that they cluster or “hang together” because they have a similar causal structure. Second, as I outlined in section 1.5, a kind-concept can enjoy predictive success without reflecting a real kind in nature. If this is right, then concepts such as ‘schizophrenia’ might be both construct valid and predictively valid and still not “natural” or “real” in the sense described in chapter 1—though, others have argued that they don’t not even meet these requirements. Cf. Poland (2004).

17 See also section 2.3.3.
As we will see in the next chapter, from a historical standpoint it is clear that psychiatric categories were constructed not on the basis of established theoretical knowledge of their causal nature, but, rather, in accordance with the hope that, in dividing the psychiatric domain in this way, the true causal nature of mental disorders would be discovered.

3.6. Conclusion

In this chapter I have conceded that mental disorders are natural kinds in Boyd’s weak sense of ‘natural kind’. Be that as it may, it is difficult to imagine that this is what philosophers of psychiatry have in mind when they assert that psychiatric classifications ought to be revised so as to reflect natural kinds. For instance, both Kendler et al. and Tsou assume that the DSM has failed as a medical classification system precisely because it is not a causally-based classification (see, for example, Kendler et al., 2011, 1143 and Tsou, 2016, 406). The HPC view of natural kinds, however, does not require that categories be individuated according to their causes; at most it requires that a category be accommodated to the causal structure of the world in virtue of being accommodated to the inferential practices of the researcher.

It would seem, then, that when philosophers of psychiatry say that psychiatric classifications ought to classify natural kinds, what they really mean to say is that psychiatric classifications ought to classify HPC kinds in the strong sense of ‘HPC kind’, which is consistent with the view of natural kinds discussed in chapter 1. As I have argued throughout this dissertation, there is no reason to think that current psychiatric categories are natural kinds in this sense. This, however, is not to say that psychiatric classification
cannot in principle be revised so as to reflect the causal structure of the world. The next chapter will be devoted to whether or not this normative ideal of scientific classification is plausible.
CHAPTER 4

Psychiatric Kinds and Natural Classification

[H]owever we devise [a diagnostic schema] we realize that it cannot work; that we can only make temporary and arbitrary classifications; that there are a number of different possibilities which account for the fact that different workers construct entirely different schemata; and that classification is always contradictory in theory and never quite squares with the facts. Why then do we keep on making this vain attempt?


4.1 Introduction

Thus far I have been asking the question “Are mental disorders natural kinds?” I have argued that nothing *in principle* prevents mental disorders from being natural kinds. It’s just that, at the present moment, there is no strong evidence to support the view that current psychiatric kinds are natural kinds. A question that I have not yet addressed, however, is whether or not it is plausible to expect that our psychiatric categories will come to reflect natural kinds. Is this normative ideal an appropriate or useful goal for psychiatry, or should we abandon it altogether?

When philosophers suggest that our psychiatric categories ought to reflect natural kinds they presuppose a picture of the world in which the properties of kinds of mental disorder cluster together in ways that are *useful* for psychiatric classification. They also presuppose that by consulting our best theories about the world we can know which clusters are real and which are bogus.

In this chapter I aim to cast doubt on these assumptions. First, I argue that the idea that our best theories about the world are a reliable guide to which psychiatric kinds exist and which don’t is historically false. In fact, the evolution of the ‘schizophrenia’ concept
suggests that schizophrenia is a kind in search of a theory, not a kind firmly established by theory. Our best theories of schizophrenia are not guides to whether or not schizophrenia is real or bogus. Quite the contrary. In theorizing about schizophrenia one presupposes the existence of the very thing they are attempting to theorize about.

Second, I challenge the assumption that kinds of mental disorder do indeed cluster together in ways that are useful for psychiatric classification. In psychiatry, presumably our ultimate goal is to treat those individuals who suffer from problems which are psychopathological in nature. I contend that the phenomenologically and causally complex nature of psychopathological phenomena makes the traditional approach of dividing up the psychopathological domain by postulating hypothetical constructs impractical for these purposes.

In light of the above considerations I conclude that the normative ideal of natural classification—i.e. classifying real or natural kinds in the sense described in chapter 1—is not appropriate as a general strategy for psychiatry. In the conclusion to this dissertation I propose an alternative way forward. Rather than assuming a priori that there are indeed real kinds of mental disorder that cluster together, and expending our energy on establishing their reality, a more tractable and theoretically promising approach would be to first explain how particular kinds of experiences and behaviours—e.g., hallucinations, delusions, disorganized thinking, mania, low mood, etc.—are produced. In short, classification should be secondary, not primary, to psychiatric research.

4.2 The Origin of Schizophrenia
Psychiatrist and prominent schizophrenia researcher Nancy Andreasen once said that to be ahistorical is “to lack the capacity the see things within the context of the broad sweep of time” (Andreasen, 1994, 1405). In the context of surveying changing definitions of ‘schizophrenia’ she describes the ahistorical fallacy in the following way:

The Ahistorical Fallacy rests on three faulty assumptions. The first assumption is that “proposition X must be true because it is what the ‘experts’ are teaching”. The second assumption is that “proposition X must be true because it is the most recent one to be put forth”. The third assumption, which buttresses the first two, is that “if information increases, knowledge increases as well” (ibid.; emphasis in original).

Realists about natural kinds suggest that the only way to know whether or not a kind is “real” or “natural” is to consult our best theories about the world (Cf. Khalidi, 2013, 2; Kornblith, 1993, chapter 3). This may very well be true. This does not imply, however, that our best theories of the world inevitably, and as a matter of course, furnish us with real kinds. As we have already seen (see chapter 3), philosophers who argue that kinds of mental disorder, such as schizophrenia and depression, are natural kinds are often content to defer to the experts. In doing so they presuppose the existence of the very thing they are attempting to describe, and often assent to theories of mental disorders that are either simplistic or plainly false. In short, they have fallen prey to the ahistorical fallacy, assuming that the kinds that we currently recognize, as well as our theories about them, are reliable guides to the existence of real kinds of mental disorder. What I have to say about schizophrenia in this chapter should disabuse us of this notion.

4.2.1 The Schizophrenia Concept

---

1 As so Khalidi continues: “[I]t is not that our best theories and settled knowledge actually determine which kinds exist, but rather that they serve as our best guide to the existence of the kinds of things in the world” (ibid.; emphasis in original).
If schizophrenia is a natural kind, then we would expect that, as a matter of historical fact, the concept would have been progressively refined over time so as to more closely reflect the real structure of the world. According to one hypothesis—call it the “continuity hypothesis”\textsuperscript{2}—this is indeed the case. Our current concept of ‘schizophrenia’ began with partial descriptions in the early to late nineteenth century. Haslam (1809) observed a schizophrenia-like presentation in young persons, with chronic delusions and hallucinations accompanied by disorganized thinking.\textsuperscript{3} In the middle of the nineteenth century Morel (1860, 516, 566) coined the term ‘démence précoce’, meant to describe cases of cognitive deficit in early adolescence. During the late nineteenth century Kahlbaum (1874) observed patients who had marked psychomotor disturbances, including mutism, negativism (no response to instructions or external stimuli), stereotypies (frequent and repetitive movements), catalepsy (muscular rigidity and fixity of posture), waxy flexibility (resistant to being moved), and verbigeration (very fast and incomprehensible speech). He called this clinical presentation “catatonia”. And, finally, during the same decade Hecker (1871) coined the term ‘hebephrenia’, which meant to capture an illness affecting young persons who presented with feelings of sadness, delusions, bizarre behaviour, a tendency to laugh and tell silly jokes.

It was left to Kraepelin at the end of the nineteenth century to integrate these partial, incomplete clinical descriptions under a single nosological entity termed

\textsuperscript{2} I borrow this phrase from Berrios (2003, 113). Shorter’s (1997, 99-109) description of Kraepelin’s classificatory revolution, and his (unsubstantiated) derision of other contemporary research programs, is a fitting example of the continuity hypothesis.

\textsuperscript{3} The most famous case is that of James Tilly Mathews, described by John Haslam (Cf. Haslam, 1809, 49-51, 64-67). According to historian Edward Shorter (1997, 62-63), Phillipe Pinel also described a similar case (Cf. Pinel, 1809, 182).
‘dementia praecox’. In the sixth edition of his *Psychiatrie* Kraepelin (1899a) brings together Kaulbaum’s catatonia, Hecker’s hebephrenia, and his own “dementia paranoides”, in which delusions and hallucinations were prominent. According to the continuity hypothesis, Kraepelin’s new “dementia praecox” was superior to previous nosological entities since it was based on pure empirical observation. Kraepelin recorded thousands of case histories on special cards, followed his patients up, analyzed the data, and was led to the inexorable conclusion that catatonia, hebephrenia, and paranoid dementia were really different aspects of one kind of disease, and that dementia praecox was also separable from manic-depressive insanity on the basis of differences in course, outcome, and symptoms.

The origin story of the ‘schizophrenia’ concept ends with the publication of DSM-III in 1980. After an all too long hiatus, whereby the concept was used inconsistently amongst researchers and clinicians (because of the undue influence of Bleuer, Meyer, and psychoanalysis on American psychiatry), Kraepelin was rediscovered. While DSM-III opts for Bleuer’s (1908) term ‘schizophrenia’ instead of ‘dementia praecox’, the

---

4 As Berrios (2003, 117, 128) suggests, Kraepelin might have taken the phrase from Arnold Pick, who in 1891 suggested that dementia in young people should be called “dementia praecox.”

5 In the early 1850s Pierre Falret and Jules Baillarger observed that mania and depression were not isolated symptoms, but often alternated over the course of a person’s life, which they called “circular insanity” (Cf. Sedler, 1983). Kraepelin (1899a) re-baptized this and other so-called affective disorders as “manic-depressive illness”. See section 4.5.1 below.
nosological entity itself largely resembled Kraepelin’s, and still does to this day. Included are the subtypes disorganized (hebephrenia), catatonic, and paranoid. And emphasis is placed on course (at least 6 months duration and deterioration from previous levels of functioning).\(^7\)

4.2.2 Kraepelin’s Recantation

Is the continuity hypothesis supported by historical evidence? It is true that the concept of ‘dementia praecox’ was a consequence of empirical observations. Kraepelin, for instance, took great pains to record longitudinal observations of a large number of clinical cases. These empirical observations, however, were not “pure”. While it has been argued elsewhere that the observations Kraepelin recorded on his cards contained insufficient information to draw any statistical conclusions (Weber & Engstrom, 1997)\(^8\), it is more important to recognize that the putative kind dementia praecox was only plausible at the time if one already had preconceived notions about its nature. Kraepelin himself recognized that the phrase ‘dementia praecox’ encapsulated diverse clinical

---

\(^6\) See APA (1980, 181-193; 1994, 273-290; 2013, 99-105). Our current ‘schizophrenia’ concept was also influenced by Schneider’s (1950) first-rank symptoms, which were first incorporated in the Research Diagnostic Criteria (RDC), a prelude to DSM-III (Cf. Spitzer, Endicott & Robins, 1978). Schneider took his first-rank symptoms to be decisive in the diagnosis of schizophrenia. They included: audible thoughts; voices arguing about, or discussing, the patient; voices commenting on the patient’s actions; experiences of influences on the body; thought withdrawal and other interference with thought; thought broadcast (diffusion of thought); delusional perception; and other experiences involving “made” impulses and feelings experienced as caused by an outside agency. As Jablensky (2010) notes, the International Classification of Mental and Behavioural Disorders (ICD-10) places much greater emphasis than the DSM on Schneider’s first-rank symptoms (particularly in allowing for a diagnosis of schizophrenia when only one of Schneider’s first-rank symptoms is present for at least a month), though both classification systems are in essence Kraepelinean.

\(^7\) In the most recent version, DSM-5, the subtypes are dropped, but the same catatonic, hebephrenic, and paranoid delusional symptoms remain. DSM-III-DSM-5 do not strictly require a progressively deteriorating course, but do require a “deterioration from previous levels of functioning”, and schizophrenia is distinguished from schizophreniform disorder and brief psychotic disorder principally on the basis of duration of symptoms (1 month but less that 6 months for the former, and 1 day but less than 1 month for the latter).

\(^8\) But see Jablensky (1999) for a defence of Kraepelin’s statistical conclusions.
presentations⁹; yet, on the basis that they all had a similar course (a general loss of mental efficiency and loss of control over volitional action), he maintained that they were fundamentally the same disorders (Kraepelin, 1899a, 103; 1899b). But why assume that course was indicative of a real kind of disorder and not pathognomonic symptoms, which was largely the default approach at the time? As it happens, Kraepelin was much taken by the idea, popular in biological taxonomy, that natural kinds could properly be identified by close analysis of their natural history (Cf. Berrios & Hauser, 1988; Berrios, 2003; Jablensky et al., 1992). For Kraepelin, natural kinds of diseases were assumed to be biological and evolutionary in origin, having a predictable course resulting in atavistic behaviors.¹⁰

Not everyone was convinced by Kraepelin, however. Carl Wernicke (1906) observed cases of hebephrenia in puberty which resolved over time, leaving one’s mind intact. Bleuler (1911) also believed there to be cases which did not evolve into a terminal state of deterioration, and spoke instead of a “group of schizophrenias”, rather than the

---

⁹ In fact, in successive editions of his Psychiatrie, Kraepelin broadened the category to include nine clinical forms. Alongside catatonia, hebephrenia, and paranoid dementia, Kraepelin added: dementia praecox simplex (general decline in many psychological capacities), depressive dementia praecox (depression followed by slowly progressive cognitive decline, avolition, and sometimes delusions), circular dementia praecox (prodromal depression followed by gradual auditory hallucinations, delusions, fluctuations of mood and impulsivity), agitated dementia praecox (acute onset, perplexity or exaltation, multimodal hallucinations, fantastic delusions), schizophrenia (prominent deterioration of expression in speech, with little impairment otherwise). Cf. Kraepelin (1909).

¹⁰ Degeneration theory (the idea that a hereditary taint could lead to the devolution of an organism from a complex to a simple, less differentiated form) and Faculty psychology (the idea that there are three separable mental functions—the intellectual, emotional, and volitional) served as influential conceptual frameworks in the nineteenth century. As Berrios (2003) notes, the former likely influenced the ‘dementia praecox’ concept, and the latter the distinction between dementia praecox (intellectual insanity), manic-depression (emotional insanity), and psychopathic/antisocial personality (volitional insanity). See also: Berrios and Hauser (1988).
all-encompassing “dementia praecox”\textsuperscript{11}. Most importantly, both Bleuler and Wernicke did not assent to Kraepelin’s \textit{a priori} assumption that dementia praecox was a degenerative disease affecting the brain. Both believed dementia praecox to be fundamentally a dysconnection syndrome. Bleuler, for instance, hypothesized that schizophrenia is a consequence of a “splitting of different psychic functions” (ibid., 1911, 5).\textsuperscript{12} By “splitting of different psychic functions” Bleuler meant that there is a disintegration of different mental functions (particularly the functions of intellect and emotion) which resulted in the loose associations observed in schizophrenia. Wernicke too hypothesized that schizophrenia-like presentations (Wernicke died in 1905 before the term ‘schizophrenia’ ever appeared) are a result of the interruption (“sejunction”) of associative connections in the brain (Cf. Wernicke, 1900). The principal difference between Bleuler and Wernicke is that the former conceived of dysconnection in a functional sense, whereas the latter conceived of dysconnection in an anatomical sense.\textsuperscript{13}

Towards the end of his career Kraepelin all but recanted his earlier view that dementia praecox could be reliably distinguished from other kinds of mental illness. For

\textsuperscript{11} Bleuler did not believe schizophrenia to be a disease \textit{per se}, but rather a group of diseases. As Shorter (1997, 107) has noted, even a quarter of Kraepelin’s own patients recovered, so it is likely that Kraepelin was aware of this issue. Interestingly, when Kraepelin first described dementia praecox in 1896 he believed it to be distinct from catatonia and paranoid dementia, and closer to Hecker’s hebephrenia. It wasn’t until 1899 that Kraepelin decided to group catatonia, paranoid dementia, and hebephrenia under the broader category ‘dementia praecox’.

\textsuperscript{12} In fact, the etymology of schizophrenia from Greek is \textit{skhizein} (“to split”) and \textit{phren} (“mind”). Bleuler introduced the term ‘schizophrenia’ to replace ‘dementia praecox’ because he believed the splitting of psychic functions to be one of the most important features, and he thought that ‘dementia praecox’ had fatalistic undertones, since it was taken by many to mean dementia affecting young people. Cf. Bleuler (1911).

\textsuperscript{13} Indeed, a version of the dysconnection hypothesis was resurrected in 1993 (Cf. Friston and Frith, 1995; Weinberger, 1993) and is still alive and well (Cf. Friston et al., 2016).
instance, of the distinction between dementia praecox and mental manic-depressive illness, Kraepelin states:

[I]t it is becoming increasingly obvious that we cannot satisfactorily distinguish these two diseases…it is conceivable that there may be an overlap in the clinical signs which are normally observed in schizophrenia and manic-depressive psychosis…It is an everyday experience that manic and depressive features can occur transiently in frank schizophrenic illnesses (Kraepelin, 1920/1992, 527-528).

Although Kraepelin was reluctant to give up on the idea that dementia praecox is a degenerative brain disease\(^\text{14}\), he grants that “schizophrenic symptomatology is in no way confined to dementia praecox” (ibid., 523).\(^\text{15}\) As evidence, he cites the fact that schizophrenic manifestations are comparable to some of the behaviours and experiences found in children and in dream states, which are clearly not a consequence of progressive brain atrophy.

In the end, Kraepelin conjectures that there might be three broad forms of mental illness (affective, schizophrenic, and encephalopathic) in which there is a continuum between more and less severe forms (based on the extent of brain damage), but nevertheless a significant overlap in symptomatology. According to this picture, “affective illnesses” are largely psychogenic in origin, having no identifiable brain

\(^{14}\) For example, after admitting that we cannot reliably distinguish between dementia praecox and manic-depressive illness, he says: “In any case, we cannot help but maintain that the two disease processes themselves are distinct. On the one hand we find those patients with irreversible dementia and severe cortical lesions. On the other are those patients whose personality remains intact. This distinction is too overwhelming for us to accept much overlap between the two groups, particularly as we can often predict the course of the two from the clinical signs” (ibid., 527-528). Indeed, throughout this essay it is clear that Kraepelin has not shaken off the influence of degeneration theory (see note 10), frequently conceptualizing mental illness, including schizophrenia, as “vestiges of earlier stages of evolution”, in which animal-like (or child-like) instinctual mechanisms have not been “subjugated to more sophisticated mechanisms” (ibid., 522).

\(^{15}\) It is crucial to note that Kraepelin uses the terms ‘dementia praecox’ and ‘schizophrenia’ separately throughout this essay, suggesting that he still believed dementia praecox to be but one type of Bleuler’s new diagnostic entity “schizophrenia”.
damage, whereas “encephalopathic illnesses” include diseases with clearly identifiable
brain damage, such as neurosyphilis and encephalitis, and “schizophrenic illnesses” fall
somewhere in between these two extremes.

4.2.3 Conclusion: The Fallout

The assumption that the course and prognosis of a disorder was indicative of real
kinds of disease was understandable at the time Kraepelin was writing given the
widespread influence of degeneration theory (see note 10), and the success in identifying
distinct pathological processes in the cases of general paresis of the insane (tertiary
syphilis) and, later, Alzheimer’s disease. Such exceptions notwithstanding, this
conceptual framework has not been successfully applied to schizophrenia. To date, no one
has been able to point to a pathological disease process specific to schizophrenia. And, in
fact, the belief that individuals with schizophrenia get progressively worse appears to be a
myth (Cf. Zipursky & Abid, 2015). Yet, as numerous authors have pointed out (Cf.
Berrios & Hauser, 1988; Jablensky, 2007), we still live in a Kraepelinian world. Our
category of schizophrenia today is, for all intents and purposes, the same one Kraepelin
proposed in 1899 (see note 6 and note 7).

The fact that Kraepelin introduced the concept ‘dementia praecox’ based on a
theoretical guesswork is not a particularly unusual aspect of science. The Austrian
philosopher of science Otto Neurath once famously said that scientists

are like sailors who on the open sea must reconstruct their ship but are never
able to start afresh from the bottom. Where a beam is taken away a new one
must at once be put there, and for this the rest of the ship is used as support. In
this way, by using the old beams and driftwood the ship can be shaped
entirely anew, but only by gradual reconstruction (Neurath, 1921, as cited in
Cartwright et al., 2008, 191).
And so it is with the science of psychiatry. We don’t have a definition to ground a theory, nor a theory to ground a definition. Be that as it may, what is unusual in psychiatry is that, for almost forty years since the publication of DSM-III, the ship has been left largely intact, only a select few beams having been replaced. Perhaps most striking is the fact that our current concept of ‘schizophrenia’ has not been revised despite the fact that, more than a century on, no one has been able validate Kraepelin’s theoretical claims. Far from theories being our best guide to the existence of the kinds of things in the world, it would seem that, in the case of psychiatry at least, theories often take on a life of their own, determining which kinds exist, and thereby constraining what it is that we theorize about.

4.3 The Resurrection of Schizophrenia

Although Kraepelin believed himself to be identifying natural disease entities by way of pure observation, and not conjectural theories about the brain (Cf. Shorter, 1997, 101-109), it is nevertheless clear that his own classification system was influenced by unsubstantiated theoretical presumptions about dementia praecox.

Be that as it may, Kraepelin was anything but dogmatic about his own classification system, even questioning the entire project of psychiatric classification toward the end of his career. It is a wonder, then, that a bit of theoretical guesswork on the part of Kraepelin has had such a widespread influence on contemporary psychiatry. Why is it that our current concept of ‘schizophrenia’, which largely reflects Kraepelin’s ‘dementia praecox’, has not been progressively revised over time in light of new empirical observations? Contrary to the continuity hypothesis, Kraepelin’s schizophrenia was resurrected not because we accumulated further evidence that he was right. Instead,
in response to a professional crisis the architects of DSM-III were eager to accept Kraepelin’s assumptions without reservation.

4.3.1 Psychiatry In Crisis

...
environment could cause mental illness was by no means new\textsuperscript{18}, it had never been extended in any comprehensive way to the domain of psychosis. As Shorter (1997, 176) notes, an important figure in this movement was Harry Stack Sullivan, who began treating individuals he diagnosed as “schizophrenic” after arriving at the Sheppard and Enoch Pratt Hospital in Townson, Maryland in 1922. Drawing upon his therapeutic experiences with six male schizophrenics, Sullivan concluded that schizophrenia was a maladaptive adjustment or reaction to anxiety (Sullivan, 1927, 105), a personality trait which had its origin in early painful interpersonal experiences (Sullivan, 1931). Frieda Fromm-Reichmann—who arrived at the Chestnut Lodge in Rockville, Maryland in 1935, and was to some degree influenced by Sullivan—went a step further, eventually proposing the idea of the “schizophrenogenic mother”, a kind of mother whose domineering influence and maternal rejection could cause schizophrenia:

\[T\]he schizophrenic is painfully distrustful and resentful of other people, due to the severe early warp and rejection he encountered in important people of his infancy and childhood, as a rule, mainly in a schizophrenogenic mother. During his early flight for emotional survival, he begins to develop the great interpersonal sensitivity which remains for the rest of his life (Fromm-Reichmann, 1948, 265; see, also, ibid., 1940).

As Hartwell (1996) argues, the idea of the schizophrenogenic mother was picked up in one way or another by researchers of diverse theoretical backgrounds, including psychodynamic theorists, learning theorists, field theorists, communication theorists. Not

\textsuperscript{18} Already in the seventeenth century Burton implicated environmental and psychological factors in the genesis of melancholy. In the nineteenth century Esquirol (1845)—along with many others—emphasized the psychosocial causation of insanity. Freud, of course, conceptualized a whole range of neuroses psychologically, though he was always ambivalent about whether or not such explanations could be extended to psychosis.
all of these theorists pointed the finger exclusively at the mother$^{19}$, though the idea that psychosis was the consequence of faulty child rearing, environmental stress, or maladaptive communication patterns between family members (so-called “communication deviance”) was widely held in American psychiatry until at least the 1970s.$^{20}$

The insistence that schizophrenia—which some have called the “sacred symbol of psychiatry” or “madness *par excellence*” of our time$^{21}$—was more a “problem in living” than a medical disease in any traditional sense paved the way for widespread criticism of the discipline of psychiatry itself.$^{22}$ After all, if mental disorders were not medical diseases, what were psychiatrists—who fashioned themselves as medical doctors—treating?

Based on his observations of psychiatrists screening patients for involuntary commitment, sociologist Thomas Scheff (1966/1999) concluded that psychiatrists were in the business of labelling and managing deviance, rather than diagnosing and treating disease. Another sociologist, Erving Goffman (1961), did observational research at St. Elizabeths Hospital in Washington and concluded that mental hospitals were more like

---

$^{19}$ For instance, Gregory Bateson and his colleagues (1956) proposed the “double-bind” theory of schizophrenia, arguing that schizophrenia was a consequence of a failure by children to integrate contradictory meanings in parental communications. This idea was later picked up by Laing and Esterson, who, in their influential book *Sanity, Madness, and the Family* (1964/1990), present extensive analysis of interviews with the families of eleven women diagnosed with schizophrenia. They argue that the families of these schizophrenics presented them with confused and impossible commands (e.g., encouraging them to become independent but consistently undermining their attempts to do so), giving credence to the idea of a “schizophrenogenic parent” or “schizophrenogenic family”.


$^{21}$ See Szasz (1976) and Foucault (1973, 374), respectively.

$^{22}$ Cf. Harrington (2019, chapter 4).
prisons or concentration camps than medical facilities. Perhaps the most influential sociological critique of psychiatry came from Michel Foucault. In *Madness and Civilization* (1961/1988) Foucault argues that the history of psychiatry is not a story about the progressive scientific discovery of “mental illness”, but rather a moral and political story about how we have managed to discipline those who were believed to be mad and socially deviant.

Criticism also came from within the psychiatry’s ranks. In 1961 the Hungarian psychiatrist Thomas Szasz published *The Myth of Mental Illness*, arguing that real medicine dealt with disease, and there was simply no proof that mental illness involved any kind of organic pathological disease process. The project of identifying real kinds of mental disease by way of postmortem observations of anatomical abnormalities is the brain—the so-called “brain localization” project, which began in the mid-nineteenth century—had failed.²³ For Szasz, it was high time to give up on the idea that mental illness is an illness like any other.

Schizophrenia was a popular target for criticism at this time. In *The Divided Self* (1960/1990) Scottish psychiatrist R.D. Laing theorizes that, due to dysfunctional family relationships, schizophrenics develop ontological insecurity (i.e. insecurity about one’s existence), which prompts a defensive reaction in which the self splits into separate components, thus generating the psychotic symptoms characteristic of schizophrenia. He later accused families of colluding with bad-faith doctors by turning their children into

---

²³ To qualify Szasz’ position, it’s not that the brain localization project was a complete failure. There were many successes (to cite just one well-known example: the localization of speech production and comprehension—Broca’s area and Wernicke’s area, respectively). It’s just that this project did not pan out for mental illnesses like schizophrenia, manic-depression, hysteria, neurasthenia, etc. Cf. Szasz (1961/1974).
patients, which—to his estimation—they were not.\textsuperscript{24}

As the historian Anne Harrington (2019, 124-126) details, a watershed moment in the history of psychiatry (and, I would argue, the history of the ‘schizophrenia’ concept) was when “On Being Sane in Insane Places” (1973), written by the psychologist David Rosenhan, was published in one of the most influential scientific journals in the world: \textit{Science}. In the article Rosenhan describes a natural experiment he conducted in which eight pseudopatients (including himself) gained admission into twelve different mental hospitals by pretending they were mentally ill. Upon arriving at the institutions the pseudopatients told admissions staff that they had been hearing voices which said “empty”, “hollow”, and “thud”.\textsuperscript{25} All of the pseudopatients were admitted, given diagnoses of schizophrenia (and, in one case, manic-depressive illness). Despite ceasing to simulate any symptoms of abnormality upon admission (as was instructed), telling staff that their symptoms had abated, and generally being very amicable and cooperative, the average length of hospitalization was 19 days, with one pseudopatient being discharged after 52 days. All patients were released with a diagnosis of “schizophrenia in remission”, the presumption being that the patients were symptom-free but still “schizophrenic”.

The natural experiment continued when a research and teaching hospital heard of Rosenhan’s results and challenged him to send pseudopatients their way. Rosenhan obliged, telling the hospital staff that within three months one or more pseudopatients

\textsuperscript{24} Cf. Laing and Esterson (1964/1990) and Laing (1967).

\textsuperscript{25} The pseudopatients reported no other symptoms, and, according to Rosenhan, the symptoms were chosen because they were similar to “existential symptoms…which are alleged to arise from painful concerns about the perceived meaningless of one’s life…as if the hallucinating person were saying, ‘My life is empty and hollow’”—and, further, that existential psychosis had never been recorded in the literature before (Cf. ibid., 251)
would try to gain admission to the hospital. In the end, forty-one patients were
confidently identified as pseudopatients by at least one member of the hospital staff. In
fact, Rosenhan hadn’t sent any. It is clear, Rosenhan concludes, that “we cannot
distinguish the sane from the insane in psychiatric hospitals” (ibid., 257); “psychiatric
diagnoses…are in the minds of the observers” (ibid., 251).

The reaction was vitriolic. Robert Spitzer—who had recently been appointed chair
of the American Psychiatric Association's task force of the third edition of the DSM,
DSM-III—published a two-part criticism of Rosenhan’s study, saying:

> “On Being Sane in Insane Places” is pseudoscience presented as science. Just
as his pseudopatients were diagnosed at discharge as “schizophrenia, in
remission”, so a careful examination of this study's methods, results, and
conclusions leads me to a diagnosis of “logic, in remission” (Spitzer, 1975,
442).26

In the article Spitzer gives a point-by-point criticism of the study, but his main argument
is that Rosenhan’s study is not ecologically valid. The fact that hospital staff incorrectly
diagnosed pseudopatients does not entail that they incorrectly diagnose real patients. After
all, the hospital staff had never diagnosed pseudopatients before—and, under the
circumstances, their diagnosis of schizophrenia was reasonable.

In his concluding remarks Spitzer defends the utility of psychiatric classification
—particularly in that it allows for effective communication between mental health
professionals—but concedes that at present psychiatric diagnoses are woefully
inadequate, and that he is working with a group out of Washington University in St. Louis
to develop an entirely new diagnostic system.

---

26 See also: Spitzer (1976).
4.3.2 The Making of DSM-III

There is widespread agreement that the DSM-III was formulated as a response to a professional crisis (Cooper, 2004, 2005; Harrington, 2019, 126-136; Shorter, 1999, 295-305; Tsou, 2016; Whooley, 2010, 2014). This crisis, as Whooley (2010, 2014) has argued, was initiated by a confluence of factors: dissatisfaction with psychoanalysis; the contention that psychiatric diagnoses were arbitrary and inconsistent; and increasing pressure from third-party medical insurance companies for more clarity in psychiatric diagnosis and treatment. The straw that broke the camels back, however, was the assertion that mental illness was a myth, and that psychiatry was not a branch of medicine at all.

In collaboration with a small group of research-orientated psychiatrists based primarily at Washington University in St. Louis and the New York State Psychiatric Institute, Spitzer chartered a new course, determined to place psychiatry back on a firm scientific footing. In 1972 members of the St. Louis group, led by John Feighner, published the first set of rigorous criteria to be used for psychiatric diagnosis (Feighner, Robins & Guze, 1972). No longer was diagnosis to be based merely on clinical judgement, but, rather, on predetermined sets of inclusion and exclusion criteria, which could be used consistently from clinician-to-clinician, country-to-country. The Feighner criteria were later refined into the Research Diagnostic Criteria (RDC), which would serve as the basis for DSM-III (Cf. Spitzer, Endicott & Robins, 1978).

One common misconception amongst philosophers is the belief that DSM-III was a purely atheoretical, descriptive approach to classification, a modest attempt to make diagnoses more reliable. On this reading, the DSM architects were neo-positivists who
decided to renounce speculative theory altogether in the interest of descriptive adequacy, and this is the reason that (for the most part) real kinds of mental disorder have not been unearthed (Cf. Murphy, 2006, chapter 1 & chapter 9; Tsou, 2011, 2016).

From a historical perspective, this is plainly false. It is true that the DSM-III document itself states that it is “atheoretical with regard to etiology” (APA, 1980, 7). But the drafters of DSM-III were being disingenuous.27 The publication of DSM-III was a remarkable turning point in the history of psychiatry not only because it introduced objective, operationalized diagnostic criteria, but because it effectively marked the end of psychoanalysis, and once again established psychiatry as a medical discipline which dealt with real illnesses, not problems in living or myths. As both Shorter (1997, 300-301) and Harrington (2019, 134) note, the hope was that the new DSM diagnoses would correspond to natural disease entities, and the belief was that biological—not psychosocial—causes would eventually be discovered for all mental disorders.28

Note long before the publication of DSM-III Gerald Klerman fashioned the DSM architects as “neo-Kraepelinians” who were bringing psychiatry back into the fold of medicine, much as Kraepelin did at the end of the nineteenth century. In 1978 he provided a summary of everything that the DSM architects stood for: psychiatry is a branch of

---

27 As Spitzer would later admit: “With its intellectual roots in St. Louis instead of Vienna, and with its intellectual inspiration derived from Kraepelin, not Freud, the task force was viewed from the outset as unsympathetic to the interests of those whose theory and practice derived from the psychoanalytic tradition” (Bayer & Spitzer, 1985, 188). This attitude was evident early on. In a highly influential article, Robins and Guze (1970), members of the St. Louis group, argued that a valid scientific classification in psychiatry is one which comprised of classes of “real illnesses”, much like the rest of medicine. The architects of DSM-III originally planned to include a statement in the manual indicating that the mental disorders are medical conditions (Cf. Spitzer and Endicott, 1978), but were met with opposition from psychologists, social workers, and counsellors who believed this statement to be a declaration that only psychiatrists with medical training were qualified to treat mental disorders (Mayes and Horwitz, 2005).

28 As Shorter (2013, 8) says, “Robert Spitzer…couldn’t wait to get rid of hysteria, neurasthenia, and the rest of the psychoanalytic baggage”.
medicine; mental illnesses are not myths, but real illnesses that require treatment; there is a natural boundary between the normal and the sick; there are many illnesses and natural boundaries between them; and classification ought to be based on investigation into the causes of mental illness, particularly biological causes (Klerman, 1978).

4.3.3 *Schizophrenia Resurrected*

Out of the foregoing crisis came the concept of ‘schizophrenia’ as we know it today. ‘Schizophrenia’, in particular, was used wildly inconsistently in the decades prior to the publication of DSM-III. For instance, psychiatrists in America diagnosed people with schizophrenia much more frequently than those in Britain, even when symptomatology appeared to diverge widely from what we now consider “schizophrenia” (Cf. Kendall et al., 1971). There was some rationale for this, however. Psychoanalytically-oriented psychiatrists, as we have seen, believed schizophrenic-like symptoms to be particularly morbid psychological reactions to the social world—on the extreme end of a continuum ranging from mild neuroses to severe psychoses—and so never had much interest in nosology. After all, if psychotic symptoms were a consequence of how one adjusted to the world, and each individual’s world and reaction to it were different, why pigeonhole individuals into broad, abstract categories?

It goes without saying that neo-Kraepelians did not share this perspective, and, as historian Edward Shorter has nicely detailed, they would make certain that the idea of a non-chronic, etiologically psychogenic kind of schizophrenia was expunged from the record:

Yet once the disease-designers were at the negotiating table, their approach
more resembled horse-trading than admiration for science. Take “schizophrenia,” which emerged from the published DSM-III as seemingly rock-solid a diagnosis as mumps. As psychologist Theodore Millon, member of the DSM-III Task Force, pointed out to Spitzer in 1978, this label—for what is essentially chronic psychosis—had begun life in DSM-I as “schizophrenic reaction” and was simplified to “schizophrenia” in DSM-II in 1968. Then, as the DSM-III Task Force began to meet, they converted it in 1974 into “schizophrenic disorders” in recognition, said Millon, that the term represented “a spectrum, if you will, that is, a heterogeneous syndrome, etiologically biogenic in some cases, psychogenic in others, and most likely interactive in the majority.” Then, to Millon’s irritation, in 1978 Spitzer unilaterally reconverted the label to “schizophrenia,” as it remained in the published version…So, schizophrenia, was it one disease or a haphazard catchall? The DSM system was all over the map (Shorter, 2013, 11).

As was the case with Kraepelin’s ‘dementia praecox’, our contemporary concept of ‘schizophrenia’ was based on the assumption that schizophrenia is a chronic, neurodegenerative disease which is affected little, if at all, by the environment.

Initially, DSM-III was hailed as a crowning achievement of psychiatry. Klerman (1984, 542) called it “a victory” for science. In her widely read book, The Broken Brain: The Biological Revolution in Psychiatry (1984), Nancy Andreason said that the biological revolution in psychiatry would finally put an end to mental illness stigma and psychoanalytic speculation. In 1989 the psychiatrist Samuel Guze wrote an essay entitled: “Biological Psychiatry: Is There Any Other Kind?”, effectively marking the end of psychiatry’s historical, and quite rapid transition from psychoanalysis to biological psychiatry.

Fast forward a decade and the leaks in the dam were already beginning to show. In 1991 clinical psychologist Mary Boyle wrote Schizophrenia: A Scientific Delusion?, arguing that the putative illness is far less coherent than it has long been taken to be.
Nancy Andreason herself, who had been involved in DSM-III working group, started questioning the category, contending that we seem to have fallen prey to the ahistorical fallacy (Andreason, 1994), and that DSM-III led to many unintended but negative consequences, such as the loss of “careful clinical evaluation that is targeted to the individual person’s problems and social context and that is enriched by a good general knowledge of psychopathology” (Andreason, 2007, 108).

Perhaps the most damning statement concerning the current state of psychiatric classification came from Thomas Insel, who at the time was director of the National Institute of Mental Health (NIMH). In anticipation of the release of the most recent version of the DSM, the DSM-5, Insel released his own scathing report of the so-called “Bible” of Psychiatry (Insel, 2013). According to Insel, the final version of the DSM-5 “involves mostly modest alterations of the previous edition”, and that, more than thirty years after the neo-Kraepelineans thought they had put psychiatry back on solid scientific foundations, not one of the DSM’s diagnostic categories were based on “objective laboratory measures”. Of schizophrenia Insel says: “[W]e still do not have a basic understanding of the pathophysiology of the disorder and therefore lack the tools for curative treatment or prevention needed” (Insel, 2010, 187). Translation: we can’t even say that schizophrenia is indeed a real kind, let alone a real kind of neurodegenerative disease.

4.3.4 Conclusion

The history of the ‘schizophrenia’ concept in particular, and psychiatric concepts in general, has not been kind to philosophers and scientists who would fashion themselves
as realists about kinds of mental disorder. If it is true, as Hilary Putnam argues, that scientific realism “is the only philosophy that doesn't make the success of science a miracle” (Putnam, 1975c, 73), then it is equally true that the failure of psychiatry to meaningfully prevent, cure, and explain mental disorder makes scientific realism about mental disorders incomprehensible.

What is more, realists about natural kinds claim that our theories about the world are our best guide to the existence of real kinds, and that, given the success of our scientific theories, it would be utterly miraculous if they did not pick out real kinds. The case of schizophrenia turns this argument on its head. The theoretical assumption that schizophrenia is in fact a real kind of disease has led to some successful discoveries. To cite one well-known example, we know that increased dopamine levels are particularly associated with the positive symptoms of schizophrenia, and, on this basis, we can successfully treat at least some individuals diagnosed with schizophrenia with antipsychotic medication. Yet, given that the category of ‘schizophrenia’ was formulated on the basis of shaky theoretical foundations, and not knowledge of its causal structure, it would be utterly miraculous if schizophrenia turned out to be a real kind.

As Kornblith (1993, 51) has noted, there was a time in the early classification of both chemical and biological kinds when color played an important role in distinguishing between different kinds of thing. In light of theoretical developments in chemistry and biology, however, color proved to be an unreliable indicator of real kinds.

Psychiatry, in many ways, is still awaiting its own theoretical revolution. While I have repeatedly said that what is true of concepts such as ‘depression’ and ‘schizophrenia’
does not necessarily apply to the whole psychopathological domain, there is a general consensus that most of our psychiatric concepts have been constructed similarly. As I contended in the previous chapter, psychiatric concepts are best interpreted as “useful heuristics” in the sense that they provide us stable “objects” to observe and study, and in some instances might even be both construct valid or predictively valid. Be that as it may, they remain “hypothetical constructs” in the sense that they do not (so far as we know) have unobservable causal structure which explains why the observable properties appear to co-occur.

4.4 Diagnosing the Problem of Psychiatric Classification

As we have seen, in psychiatry our best theories are not necessarily reliable guides to which kinds do or do not exist. The question that presents itself to us now is whether or not we should expect that our current psychiatric categories will eventually come to reflect natural kinds. Is this normative ideal an appropriate or useful goal for psychiatry, or should we abandon it altogether?

The historical record should make us skeptical. Why is it that, more-or-less two centuries since psychiatry emerged as a distinct discipline, real kinds of mental disorder have not been unearthed, despite our best efforts? Some philosophers have argued that psychiatry has not progressed because, since DSM-III, psychiatric classification has been based on atheoretical observational criteria, and this has discouraged etiological explanations of mental disorders (Cf. Murphy, 2006, chapter 1 & chapter 9; Kendler et al., 2011; Tsou, 2008, 2015, 2016).

The problem with this explanation is that it’s historically false. First, as we have
seen, at least some DSM mental disorders were based on theoretical speculation concerning their nature. Second, while it is true that psychiatric *classification* has been based in large part on observable signs and symptoms, scientists did not abruptly stop formulating and testing theories of mental disorders once DSM-III was published. In fact, the DSM architects fully believed and hoped that future scientific investigation would sort out which kinds of mental disorder were real and which were not, and that psychiatric classification would be revised accordingly.

A more plausible explanation is that the way in which the DSM is employed in research makes the discovery of the true causes of mental disorder difficult, thereby frustrating our attempts at identifying real kinds of mental disorder. As Tabb (2014) has argued, owing to its widespread use in the framing of scientific hypotheses about mental disorder, the DSM has largely determined the objects of inquiry (in particular, DSM diagnostic criteria are widely used to gather test populations for studying mental disorders).\(^{29}\)

While there is some truth to this, this explanation is too simplistic. For example, as Jablensky (2010) has detailed, since DSM-III many attempts have been made to revise the ‘schizophrenia’ concept. Shortly after the publication of DSM-III, “Type I (or positive)” and “Type II (or negative)” schizophrenia were proposed based on the predominance of either positive or negative symptomatology (Cf. Crow, 1980, 1985; Andreason & Olsen, 1982). The rationale for this was that positive symptoms (hallucinations, delusions, and formal thought disorder) were presumed to be related to

\(^{29}\) This is essentially the same argument made by Hyman (2010) and Insel (2013), and has served as motivation for the Research Domain Criteria (RDoC) project.
dopaminergic dysfunction, while negative symptoms (social withdrawal, loss of volition, affective flattening, and poverty of speech) were presumed to be associated with structural brain abnormalities.

Other attempts have been made: deficit-nondeficit schizophrenia, based on the presence or absence of enduring negative symptoms (Carpenter, Heinrichs & Wagman, 1988; Kirkpatrick, Buchanan, Breier & Carpenter, 1993; Kirkpatrick, Buchanan, Ross & Carpenter, 2001; Buchanan et al., 1994); family-sporadic schizophrenia, based on the presence or absence of a family history of schizophrenia (Lewis et al., 1987); different subtyping based on factor and cluster analyses of which symptoms tend to co-occur (Farmer, McGuffin & Spitznagel, 1983; Johnstone & Frith, 1996; McGrath et al., 2004). All of these different methods have resulted in different possible ways of further refining “schizophrenia”, but thus far nothing has stuck.

The failure of psychiatry to identify natural kinds of mental disorder has not been for a lack of trying. This opens up a number of possibilities. First it is possible that we’re searching for a phantom, and there are no real kinds of mental disorder. Second, it is possible is that we just haven’t got there yet, and in due time improved investigative techniques and theoretical developments will reveal to us which kinds of mental disorder are real and which aren’t. And, finally, it is possible that psychopathological phenomena are in fact much more complex and chaotic than was once anticipated, and the project of classifying real kinds of mental disorder by formulating hypothetical constructs is not a useful one.

In the end, I take the latter hypothesis to be the most plausible, and will argue for
it in the remainder of this chapter. It is true, of course, that science is full of surprises and it is impossible to know exactly where science is headed based on knowledge of science’s past. Be that as it may, over the past two centuries we have accumulated enough evidence to suggest that psychiatric categories are best conceptualized as convenient pigeonholes, attempts to impose order on a disorderly, chaotic nature.

As we have seen, many psychoanalysts once shared a similar intuition, and as a consequence were not much interested in psychiatric classification. While psychoanalysts were clearly wrong to suggest that psychopathological symptoms are merely morbid reactions to the social world, and do not involve biological factors\(^\text{30}\), their insistence that psychopathological phenomena could not be neatly cut up into natural kinds in virtue of clustering signs and symptoms appears to be borne out by close analysis of the empirical evidence—or so I will argue.

### 4.5 From Order to Chaos

The project of classifying natural kinds of mental disorder has historically proceeded under the assumption that one kind of disorder can be distinguished from another on the basis of course and outcome, signs and symptoms, and causes. This research program was a gamble, one that bet that amongst the many diverse clinical presentations that psychiatrists encountered in practice could be found natural kinds of mental disorder. To date, this conceptual framework has failed.

In my mind, the principal reason why this conceptual framework has failed is that the psychopathological domain itself is much more complex and chaotic than was once

\(^{30}\) It should be noted that Freud himself did not think this, but many of his followers did.
anticipated. First, as has long been recognized, there is significant overlap in symptomatology between individuals, which strongly suggests that the properties of mental disorders do not cluster together in useful ways. Second, the properties of mental disorders are unstable inasmuch as they both change over time and space, and from individual to individual.

Taken together, this evidence strongly suggests that the project of classifying real kinds of mental disorder by way of identifying clusters of properties, one aspect of what Gold and Gold (2014, 229-234) have called “aspirational psychiatry”, is an aspiration ill-suited to the psychopathological phenomena.

4.5.1 Clustering Signs and Symptoms

As we saw in the previous chapter, realists about mental disorder believe that at least some kinds of mental disorder are natural kinds because they are characterized by a cluster of properties that regularly co-occur, and the co-occurrence of these properties can be explained by the homeostatic mechanisms which bring them about. As I have already argued (see section 3.4), the claim that there are causal mechanisms which do in fact explain how the superficial properties of, e.g., schizophrenia and depression cluster together is based on a superficial understanding of the scientific literature. Be that as it may, there is an additional assumption that I did not address in chapter 3 and simply granted for the sake of argument—namely, that psychopathological phenomena can be characterized by clusters of naturally co-occurring properties.

There are both historical and empirical reasons for doubting this assumption. From a historical standpoint it would be entirely serendipitous if the properties which
characterize, e.g., psychotic disorders do in fact cluster together. As we have already seen, in 1899 Kraepelin combined the previously distinct dementia praecox, catatonia, and paranoid dementia based on longitudinal analysis, not cross-sectional or cluster analysis. Kraepelin fully recognized that these diverse clinical forms did not cluster together neatly, but made a bet that they were in fact diverse manifestations of the same disease.

Consider also manic-depressive illness, the other psychosis in Kraepelin’s bipartite distinction. According to Berrios (1988, 2004), for roughly two millennia the terms ‘mania’ and ‘melancholia’ were more-or-less synonymous with ‘madness’, and stood for a panoply of observed signs and symptoms, including fury, despondency, agitation, confusion, delusions, excitation, aggression, and decreased motility. By the early nineteenth century, however, the terms ‘mania’ and ‘melancholia’ were increasingly being used to refer to counterposing affective states such as fury/agitation/excitation and despondency/listlessness/sadness, respectively. For example, in 1820 Esquirol—who was acutely aware of the fast-and-loose usage of ‘melancholia’ in the past—coined the term ‘lypemania’, which he says is

a disease of the brain characterised by delusions which are chronic and fixed on specific topics, absence of fever and sadness which is often debilitating and overwhelming. It must not be confused with mania which exhibits generalised delusions and excited emotions and intellect nor with monomania that exhibits specific delusions and expansive and gay emotions, nor with dementia characterised by incoherence and confusion of ideas (Esquirol, 1820, 151-152; emphasis mine).

According to Esquirol, mania, lypemania, and monomania were different kinds of mental disease which were distinguishable in part on the basis of polar opposite mood states.

The world of so-called affective illnesses was further complicated in the mid-
nineteenth century. At that time French clinicians began describing patients who were neither exclusively manic nor exclusively melancholic, but appeared to go through periods of both mania and melancholia (see note 5). They proposed that mania and melancholia were not isolated symptoms, but often alternated over the course of a person’s life, which they variously called “circular insanity”, “intermittent insanity”, “periodic insanity”, and “insanity of double form” (Cf. Sedler, 1983). While French clinicians generally considered circular insanity to be a chronic mental illness which progressively weakened the mind, ultimately resulting in dementia, in 1882 the German physician Kahlbaum conjectured that at least one type of circular insanity affected only mood and did not lead to dementia. He called this kind of circular insanity “cyclothymia”, suggesting that the manic states of this disorder be called “hyperthymia” and the melancholic states “dysthymia”.

By the late nineteenth century diagnosing affective disorders was a complex affair. Some clinicians adopted the French understanding of circular insanity, in which chronic mental impairment (including delusions) was a central feature. Others adopted Kahlbaum’s understanding of circular insanity, in which mood was central and not intellectual degeneration. In addition to circular insanity were the concepts of ‘mental depression’ and ‘melancholia’— which increasingly came to replace ‘lypemania’ (Cf. Berrios, 1988)—as well as ‘simple mania’ and ‘simple melancholia’, which referred to conditions uncomplicated by delusions or incoherence, and did not necessarily terminate in profound dementia (Cf. Hare, 1981).

See Kahlbaum (1882).
In 1899 Kraepelin took stock of all this—much as he did with the so-called intellectual disorders hebephrenia, catatonia, and paranoid dementia—and concluded that these different clinical forms were but different manifestations of one and the same disease, which he termed ‘manic-depressive illness’:

[M]anic depressive insanity as it is to be described in this section, includes on the one hand the whole domain of so called periodic and circular insanity, on the other hand simple mania, the greater part of the morbid states termed melancholia and also a not inconsiderable number of cases of amentia…I have become more and more convinced that all of the above-mentioned states only represent manifestations of a single morbid process (Kraepelin, 1921, 2).

In the end, Kraepelin sided with Kahlbaum in thinking that circular insanity, melancholia, and the like, primarily affected mood and had a similar course and outcome which did not terminate in dementia. On this basis he also proposed that dementia praecox and manic-depressive illness were separate diseases, even though he recognized that there was significant overlap in symptomatology between these putatively distinct diseases (see section 4.2.2 above).

Manic-depressive illness, like dementia praecox, was decidedly not constructed on the basis of an observed cluster of regularly co-occurring properties, and this fact was not lost on critics of Kraepelin’s classificatory system. While Kraepelin originally subsumed melancholias (unipolar depressions) under the category of ‘manic-depressive illness’, depressive disorders have since survived as separate kinds of mental disorder in each successive version of the DSM. Furthermore, in recognition of the fact that many individuals presented with symptoms of both schizophrenia and manic-depressive illness,

32 Cf. Chaslin (1912), who argues that melancholias should be kept in a separate group from manic-depressive illness.

in 1933 Kasanin introduced the \textit{ad hoc} category of ‘schizoaffective disorder’ to account for these “in-between” patients.\footnote{Cf. Kasanin (1933). In DSM-IV and DSM-5 \textit{schizoaffective disorder} is distinguished from \textit{schizophrenia} on the basis of the presence of a mood episode during the active-phase symptoms of schizophrenia, and is distinguished from a \textit{depressive or bipolar disorder} with psychotic features based on the presence of prominent delusions and/or hallucinations for at least two weeks in the absence of a major mood episode. The assumption here is that the psychotic features of depressive or bipolar disorder primarily occur during mood episodes. See APA (1994, 292-296; 2013, 89, 105-110). Schizoaffective disorder is also listed in the first three editions of the DSM, but is not given clear operational criteria. See APA (1952, 83; 1967, 35; 1980, 202).}

Given that significant heterogeneity was built into the constructs ‘schizophrenia’ and ‘bipolar disorder’ from the start, it would be entirely fortuitous if the properties which characterize both of these putative kinds do indeed cluster together.

As a matter of fact, there are good empirical reasons for thinking that they don’t. For instance, factor analyses suggest that the symptoms of schizophrenia load on three latent factors: “psychomotor poverty” (i.e. negative symptoms, such as poverty of speech), “reality distortion” (i.e. positive symptoms, such delusions and hallucinations), and “disorganization” (i.e. thought and speech disorder) (Liddle, 1987)\footnote{This result has been replicated in both European and non-European populations (Cf. Arndt, Andreasen, Flaum, Miller & Nopoulos, 1995; Emsley et al., 2001; Gureje, Aderibigbe & Obikoya, 1995; Johnstone & Frith, 1996; Smith, Mar & Turoff, 1998). More recently, McGrath et al. (2004) have identified five factors (positive, negative, disorganized, affective, and early onset/developmental) and Cuesta and Peralta (2001) proposed a hierarchical 10-dimensional model.}. Other statistical techniques (e.g., cluster analysis and latent factor analysis) have revealed anywhere from two to six latent classes (Dollfus et al., 1996; Farmer et al., 1983; Johnstone & Frith, 1996; Kendler, Karkowski & Walsh, 1998 McGrath et al., 2004).

Interestingly, grade of membership (GoM) analysis—a kind of latent structure analysis which allows individuals to be members of more than one disease class and
represents the latent groups as “fuzzy sets”\textsuperscript{36}—was applied to the symptom profiles of 1065 individuals in a WHO International Pilot Study of Schizophrenia, and revealed eight classes, five of which were related to schizophrenia, and two of which were related to affective disorders (Murray et al., 2005).

The GoM technique is perhaps most congruent with the definition of natural kinds that I’ve been working with in this dissertation, and this evidence, while not conclusive\textsuperscript{37}, suggests that, in addition to not having a known causal structure, schizophrenia is not a natural kind because \textit{in all likelihood} the properties of schizophrenia don’t even cluster in the way that one would expect if it was a natural kind.

Compare this with both natural kinds of species and natural kinds of disease. Members of the same species \textit{tiger} undoubtedly differ in many of their properties (see section 1.4.2)\textsuperscript{38}, but there is a sort of “internal coherence” whereby members share most of their important properties in common and hence “cluster together”. Similarly, individuals diagnosed with a \textit{duodenal ulcer} do not share all of their properties in common, but nevertheless display a similar range of causes and effects (see section 1.4.3).

“Schizophrenia” is not internally coherent in this sense. As we have seen, not only is there no coherent unobservable causal structure (e.g., a network of causal mechanisms)

\textsuperscript{36} The GoM model simultaneously extracts from the data matrix a number of latent “pure types” and assigns to each individual a set of numerical weights quantifying the degree to which that individual resembles each one of the identified pure types.

\textsuperscript{37} It should be noted that the results of factor analyses, cluster analyses, and latent factor analyses are highly sensitive to the initial selection of symptoms and measurement methods. For instance, studies using different rating scales for schizophrenia end up “finding” different factors or clusters (Cf. Jablensky, 2006). Given this, statistically-derived symptom dimensions or clusters ought to be viewed with some caution. That being said, the \textit{overall body of evidence} from factor, cluster, and latent factor analyses at the very least suggests that the property-cluster which currently characterizes “schizophrenia” can come apart.

\textsuperscript{38} For example, intraspecific differences such as body size, pelage coloration, striping patterns, skull dimensions, craniological details, genetic, and molecular structures.
which explains how the signs and symptoms of schizophrenia arise, factor, cluster, and latent structure analyses suggest that the observable signs and symptoms of “schizophrenia” come apart.

Although schizophrenia might not be a kind unto itself, it is of course possible that further causal investigation might substantiate one of the aforementioned latent class models. This conclusion, however, is complicated by the fact that there is not only significant overlap in symptomatology within diagnostic boundaries, but also across diagnostic boundaries.

In practice, it is quite common to find individuals who meet the diagnostic criteria for two or more disorders\(^{39}\), which is referred to in the literature as “comorbidity”. For instance, a prospective cohort study of more than 2.5 million persons born in Denmark after 1954 found significant overlap between schizophrenia, schizoaffective disorder, and bipolar disorder (Laursen, Agerbo & Pedersen, 2009). Schizophrenia is also comorbid with depression, anxiety, suicidality (Cassidy, Yang, Kapczinski & Cavalcante Passos, 2018), obsessive-compulsive symptoms (Sharma & Reddy, 2019; Won-Gyo Shin et al., 2018), social phobia (Vrbova, 2017), panic symptoms (Ulas et al., 2007), adult ADHD (Donev et al., 2011), and substance-use disorder (Regier et al., 1990). In addition, bipolar disorder is comorbid with panic disorder (Mackinnon et al., 2002), obsessive compulsive disorder (OCD) (Amerio et al., 2015, 2019), borderline personality disorder (Parker et al., 2016), and anxiety disorders (Amerio et al., 2016).

\(^{39}\) For example, Kessler et al. (2012) report that 27.9% of adolescents 13 to 17 years of age met the diagnostic criteria for two or more disorders. Other community surveys have estimated that 45-54% of individuals who meet the criteria for one mental disorder will have one or more additional lifetime diagnoses (Cf. Andrews, Slade & Issakidis, 2002; Bijl, Ravelli & van Zessen, 1998).
In many instances the overlap between two putatively distinct disorders is considerable. For example, it is estimated that as many as 57% of people with schizophrenia have comorbid depression, 10–15% have comorbid panic disorder, 12–29% have comorbid posttraumatic stress disorder (PTSD), and 12–23% have comorbid OCD (Buckley et al., 2009). Moreover, an average of 18% of individuals diagnosed with bipolar disorder will also meet the criteria for obsessive compulsive disorder at some time in their life (Amerio et al., 2015). Perhaps the highest rate of comorbidity reported in the literature is that between anxiety disorders and mood disorders. Brown and Barlow (2002) found that 55% of people concurrently met the diagnostic criteria for both an anxiety and mood disorder. This rate skyrocketed to 76% when lifetime diagnoses were taken into consideration.

Historically, studies of comorbidity have restricted analyses to a subset of disorders. Recently, however, Plana-Ripoll et al. (2019) undertook one of the most comprehensive epidemiological studies on comorbidity to date, examining all major mental disorders recorded for individuals born in Denmark from 1900 to 2015. They found that all mental disorders were associated with an increased risk of all other mental disorders.

The accumulated evidence from comorbidity studies indicates that the so-called “comorbidity problem” in psychiatric diagnosis is pervasive. What are we to make of this? Sometimes it is suggested that comorbidity is not really a problem. After all, there is

---

40 These included organic disorders, substance use disorders, personality disorders, schizophrenia, mood disorders, neurotic disorders, eating disorders, intellectual disabilities, developmental disorders, and behaviour disorders.
nothing contradictory about one person developing two separate disorders. Some people (a lot, it turns out) are just unlucky. The problems with this line of argumentation is that it implicitly assumes that current psychiatric categories are natural kinds, a claim which is unwarranted given the evidence adduced in this dissertation.

Another explanation is that comorbidity is a function of a poor classificatory system. Once we further refine our categories, or introduce new categories, comorbidity should all but vanish. For example, it might be the case that some kinds of disorders—e.g., depression and anxiety disorder—have a similar underlying causal structure, which would explain why they often co-occur in one individual. It might also be the case that anxiety is produced by the negative appraisal of some life event, and thereby gives rise to the symptoms characteristic of depression (see section 2.3.2). In theory, further investigation should be able to sort out whether depression and anxiety should be lumped into one kind, or whether anxiety should be thought of as part of the causal pathway leading to depression.

Be that as it may, revising psychiatric classification in this way doesn’t really solve the comorbidity problem writ large. For instance, even if we lump anxiety and depression together, what are we to make of the fact that depression and numerous anxiety disorders are also comorbid with schizophrenia, and that schizophrenia itself is comorbid with OCD, ADHD, and substance use disorder? Do all of these conditions have similar underlying causes, or do some mental states, such as delusional states, tend to cause other mental states, such as depressive and obsessive-compulsive states? Does being schizophrenic often lead to depression because schizophrenics have trouble getting
on in the world?

Of course, we may very well introduce new kinds such as “schizo-obsessive disorder”, “schizo-depressive disorder”, “schizo-panic disorder”, and “schizo-PTSD disorder”, but this is bound to be an arbitrary, *ad hoc* solution. After all, if it is true that having one disorder increases the risk of developing *any* other disorder, then there is clearly a relationship of some kind between all kinds of mental disorder. This, however, opens up the possibility that the clusters of symptoms characteristic of current psychiatric kinds can be recombined in endless ways. For example, given that there is significant overlap between schizophrenia, depression, and various anxiety disorders, “schizo-depressive-anxiety disorder” would not be an unreasonable suggestion.

But how likely is it that “schizo-depressive-anxiety disorder”, which collects together diverse array of signs and symptoms, will have a common causal structure? In my mind, comorbidity does reflect a problem in psychiatric classification. The problem, however, is not that we’ve happened to pick out the wrong clusters. The problem is that psychiatric classification has proceeded under the assumption that psychopathological phenomena do in fact cluster together in such a way that reflects the true causal structure of the world.

This approach has not only been unsuccessful in unearthing natural kinds of mental disorder, but it is unlikely to be successful. Consider just one symptom of schizophrenia: hallucinations. As Oliver Sacks (2012) has convincingly argued, both auditory and visual hallucinations arise for a whole host of causal reasons. For example,

---

41 This is a serious consideration. Cf. Poyurovsky et al. (2003), Bottas et al. (2005), and Buckley et al. (2009).
both simple and complex hallucinations can arise from loss of sight (Charles Bonnet Syndrome), hearing loss, sensory deprivation, Parkinsonism, intoxication, migraines, epilepsy, narcolepsy, and PTSD. If hallucinations—which are just one symptom of schizophrenia—do not have a common causal explanation, should we really expect schizophrenia and similar hypothetical constructs to have one?

It must be said that comorbidity does not strictly imply that kinds of mental disorder cannot be natural kinds. It is well known, for instance, that species significantly overlap either because they are part of the same evolutionary line, because of interbreeding (e.g., tigrons and ligers), or because of evolutionary pressures. Be that as it may, given the likelihood that different kinds of signs and symptoms are likely to involve different causal processes, the project of classifying real kinds of mental disorder by way of identifying clusters of properties is bound to be unhelpful.

4.5.2 Looping Effects, Placebo Effects, and Nocebo Effects

In chapter 2 I considered three arguments for the conclusion that mental disorders cannot in principle be natural kinds. As you might recall, one common thread in each and every one of those arguments was the presupposition that the “instability” of (at least some) mental disorders precludes them from being natural kinds. For example, multiple personality disorder is assumed by Tsou and Hacking not to be a natural kind because it is a “moving target” in the sense that the properties associated with the kind rapidly change over time as a consequence of looping effects. Natural kinds, so the argument goes, are “relatively stable objects of study across time” (Tsou, 2008, 66), and multiple personality disorder is not.
Other philosophers of psychiatry also appear to treat stability as a precondition of natural kinds. For instance, Kendler et al. contend that mental disorders are natural kinds in virtue of being produced by “stable patterns of complex interaction between behavior, environment and physiology that have arisen through development, evolution and interaction with the environment” (Kendler et al., 2011, 1147; emphasis mine). And, further, they claim: “the identity of [a psychiatric] disease across time and across cultures is grounded in the similarity of the complex, mutually reinforcing network of causal mechanisms in each case” (ibid., 1148; emphasis mine).

The preoccupation with stability on the part of philosophers of psychiatry demands further exploration. As I mentioned in several places in chapter 2, while the fact that the properties of mental disorders are continually changing does not entail that they cannot possibly be natural kinds, such a fact, if true, would imply that the project of classifying real kinds of mental disorder by way of identifying clusters of properties will inevitably be confounded.

Suppose for a moment that it is true that being labelled with ‘multiple personality disorder’ can initiate a feedback process whereby individuals so labelled continually change their properties, and at a relatively rapid rate. The project of classifying real kinds of mental disorder by way of identifying clusters of properties is unlikely to be very useful if these cluster keep coming apart.

_Hysteria_ is another plausible example. In the history of psychiatry the putative kind hysteria has been one of the most intractable. In the fin de siècle of the nineteenth century, when hysteria was at its height, somatic symptoms in the absence of a
demonstrable organic basis—such as paralysis, tremors, dystonia, tremors, choking, amnesia, etc.—were fairly common. Today, however, such complaints are few and far between. As I mentioned in chapter 2, because hysteria is no longer with us philosophers often dismiss it as a mythical construct. But there is another explanation: namely, that the signs and symptoms of hysteria keep changing. For example, Scull (2009) has suggested that hysteria has not necessarily disappeared, but may be manifesting in different ways. Although dramatic displays of paralysis, tremors, dystonia, tremors, choking, amnesia, etc., are no longer the norm, somatic symptoms in the absence of a demonstrable organic basis still are: e.g., chronic fatigue syndrome, conversion disorder.

Of course, none of this implies that there is no kind—namely, multiple personality disorder or hysteria—to have knowledge about. It does imply, however, that there is no stable kind to have knowledge about (Cf. Hacking, 1995a, 61), which clearly has implications for the project of classifying real kinds of mental disorder by way of identifying clusters of properties.

As I outlined in both chapter 2 and chapter 3, some philosophers have tried to avoid this consequence by insisting that looping only alters the superficial or stereotypical properties of a psychiatric kind, leaving the important properties unchanged. For instance, Hacking posits that schizophrenia, though subject to looping effects, might be a natural kind in the sense that it is ultimately produced by some underlying “biological pathology P” (see section 2.2.3), which itself is not subject to looping effects. Tsou makes a similar

---

42 See both Scull (2009) and Goldstein (2001) [1987] for a historical account of the rise and fall of hysteria in the nineteenth and twentieth century, respectively.
claim about schizophrenia, maintaining that schizophrenia has a “stable biological causal structure” (see section 3.4.1), which remains stable in spite of looping effects.

One must question, however, whether the distinction between properties which are subject to looping effects and properties which are not is too neat. After all, if there is a significant change in an individual’s experience and behaviour as a consequence of looping effects, how could it be that there is no corresponding change at the level of biology? If looping effects were not driving real changes in the biology of an organism—“all the way down”, so to speak—then, insofar as biological change (e.g., a change in brain activity) is necessary for psychological or behavioural change, we should not even expect looping effects to alter the superficial or stereotypical properties of a kind.

Although no research to date has specifically examined how being labelled or classified as, e.g., “schizophrenic” or “depressed” might result in biological changes in an individual, the well-known placebo effect phenomenon (the ameliorative effect of taking an inert substance or undergoing a sham procedure) provides evidence for the view that one’s beliefs can affect both their symptomatology and biology. Consider the case of depression. Individuals with depression are particularly susceptible to strong placebo responses. According to some estimates, up to 82 percent of the improvement in mood, as measured by the Hamilton Depression Scale, can be duplicated by patients taking a placebo pill instead of an antidepressant (Kirsch, Moore, Scoboria, & Nicholl, 2002; Kirsch et al., 2008). While most research on the neurobiological underpinnings of the

---

43 Although these results were initially controversial, they have since been replicated repeatedly (Cf. Fountoulakis & Möller, 2011; Fournier et al., 2010; NICE, 2004; Turner et al., 2008). Even the Food and Drug Administration (FDA) found similar results in its own meta-analysis of all of the antidepressants that they have approved (Khin et al., 2011).
placebo effect have been done with analgesia (to reduce pain), recent evidence suggests that the belief that one is taking an active antidepressant increases mu-opioid receptor activity in areas of the brain that have previously been linked to depression, such as the subgenual anterior cingulate cortex, nucleus accumbens, and amygdala (Pecina et al., 2015).

Although the concept of a placebo effect is more narrow than the concept of a looping effect, the strong placebo response found amongst depressed patients arguably would not be present if these individuals were not at some level interacting with ideas about themselves. As Wager and Atlas (2015) note, placebo responses are elicited by an individual’s expectations that they will get better. These expectations, in turn, depend on the complex context in which the treatment is delivered. Part of this context are ideas about what depression is and whether or not it is treatable. If it were not the case that an individual came to believe that depression is the kind of thing (e.g., a chemical imbalance) that is amenable to treatment with pills, then it is highly unlikely that the mere belief that one is taking an active antidepressant would result in expectations of improvement.

In fact, recent evidence suggests that the interaction between an individual and

---

44 In fact, individuals who are able to release more endogenous opioids in response to placebo showed more improvement, suggesting that the capacity for an individual to release endogenous opioids when they receive a substance is related to their improved state. Although Pecina et al. do not test this hypothesis, there is evidence from research on the analgesic effects of placebos indicating that the regulation of endogenous opioids via placebo is a top-down process, whereby the expectation of relief first activates the prefrontal cortex which thereby “tells” the midbrain to release opioids. Cf. Wager et al., (2004, 2011, 2013).

45 That is to say, one’s expectations of treatment will depend on how they interpret a vast array of contextual information, both external and internal: being in a doctor’s office, verbal cues (e.g., the suggestion that one will feel better after treatment), social cues (e.g., a physician’s body language, tone of voice, eye contact), treatment cues (e.g., a needle puncture, syringe, a pill, or a prescription), emotions (e.g., “I feel less anxious”), and previous memories (e.g., past positive and negative experiences regarding treatment).
their very own classification can also lead to the expectation that one will not get better, resulting in nocebo responses.\textsuperscript{46} For instance, the increasing public belief that illnesses such as depression and schizophrenia are genetic in origin or “diseases of the brain”\textsuperscript{47} is strongly associated with the belief that mental illness is “hard wired” and that people with mental illness are dangerous, incompetent, and will not recover (Phelan, Cruz-Rojas & Rieff, 2002; Read & Harré, 2001). The internalization of this public belief or stigma\textsuperscript{48} can have deleterious effects on people suffering from mental illness, including a decrease in self-efficacy and self-esteem (Corrigan, Kerr & Knudsen, 2005; Corrigan, 2007; Corrigan et al., 2016), a decrease in well-being (Cruwys and Gunaseelan, 2016), an exacerbation of symptoms of mental disorder (Schrank, Amering, Hay, Weber & Sibitz, 2014), and worse prognoses (Corrigan, 2016; Phelan, Cruz-rojas & Reiff, 2002).\textsuperscript{49}

Beliefs about oneself and one’s condition might also be able to explain (at least in part) one of the most puzzling findings in schizophrenia research—namely, that people with schizophrenia in developing nations do better than those living in industrialized

\textsuperscript{46} A “nocebo response” is defined as new and worsening symptoms that are caused only by negative expectations on the part of an individual. A “nocebo effect”, on the other hand, is defined as new or worsening symptoms that occur during sham treatment (e.g., in a clinical trial patients develop “side effects” of a drug as a consequence of the deliberate or unintended suggestion and/or expectation that these side effects will occur). Cf. Häuser, Hansen, and Enck (2012) for a comprehensive review of the nocebo phenomenon in medicine.

\textsuperscript{47} For instance, in their review of sixteen population studies between 1990 and 2006, Schomerus et al. (2011) found a significant increase in the belief that schizophrenia and depression are brain disorders (e.g., 55% of respondents in 1990 agreed that schizophrenia was a brain disorder, compared to 75% of respondents in 2006). Despite this change, stigma (public prejudice and discrimination) surrounding depression did not change, and actually got worse in the case of schizophrenia.

\textsuperscript{48} In the literature public stigma is defined as public prejudice and discrimination directed at those suffering from mental illness, whereas as self-stigma is the internalization of public prejudice and discrimination.

\textsuperscript{49} Indeed, studies show that over the last half century or so the world (especially in the United States and Europe) has steadily adopted the medical model of mental illness, believing that the causes of mental illness to be “chemical imbalances”, “brain diseases”, or “genetics”. In this same time period stigma surrounding the mentally ill has actually increased.
nations (Cf. Sartorius, Gulbinat, Harrison, Laska & Siegel, 1996; Craig, Siegel, Hopper, Lin, & Sartorius, 1997; Hopper, Harrison, Janca & Sartorius, 2007). While many different explanations have been offered for this phenomenon (e.g., religious differences/attitudes toward life, less demanding and stressful conditions, more opportunity to feel productive, less highly charged attention and criticism or “expressed emotion”), some ethnographic research reveals that the belief that schizophrenia is a chronic, debilitating disease of the brain is notably absent in developing nations such as Zanzibar and Tanzania (Cf. Watters, 2010, chapter 3). It would seem that, when the label ‘schizophrenia’ carries with it the connotation the one has a chronic illness and will not get better, they often don’t.

The hypothesis that being labelled or classified as, e.g., “schizophrenic” or “depressed” changes the very nature of an illness itself undoubtedly demands further investigation. Even so, the existence of placebo and nocebo responses is enough to cast doubt on the assumption that looping effects only alter the superficial or stereotypical properties of a kind.50 At the very least, the interaction between an individual and their very own classification can change the course and progression of their illness as well as their biology. If this is right, then looping effects are not merely superfluous, causally inert aspects of a mental disorder. Rather, they are part of the causal story of a mental disorder—a part of the story that happens to make psychiatric kinds in some respects unpredictable and unstable clusters of properties. After all, inasmuch as individuals’ beliefs about themselves and their classification can differ in myriad ways, and inasmuch

---

50 In fact, according to Häuser, Hansen, and Enck (2012, 461), the same neurobiological mechanisms seem to be implicated in both placebo and nocebo responses.
as these beliefs are causally efficacious, one should not expect that psychiatric kinds will be identical across time and across cultures.51

4.5.3 Unstable Patterns of Complex Interaction

Although it is plausible to suppose that psychiatric kinds are in some ways unstable and unpredictable as a consequence of looping effects, some philosophers and scientists still maintain that the fact that schizophrenia presents similarly across cultures (Cf. Jablensky et al., 1992) implies—or at least strongly suggests—that individual members of the putative kind share a stable causal structure (Kendler et al., 2011, 1145).

But is it true, or even likely, that schizophrenia has a stable causal structure? Consider the evidence from genetics research over the past half century or so. It is undoubtedly clear that genes play an important role in schizophrenia, the overall heritability being between 65-80% (Lichtenstein et al., 2009; Sullivan, Kendler & Neale, 2003).52 Yet, even when one identical twin has schizophrenia, the chances of the other twin also having the illness are only around 40-50% (Onstad, Skyre, Torgersen & Kringlen, 1991). The risk also diminishes quite dramatically when one considers non-identical twins, siblings, and children of individuals diagnosed with schizophrenia. In fact, as Gottesman (1991, 99) notes, most parents of individuals diagnosed with

---

51 Hacking appears to make a similar suggestion when he contends that feedback in natural kinds is different because “it works not at the level of individuals but through a great many generations, be it for microbes or mammals” (Hacking, 1992, 190). While I do think that this prevents mental disorders from being natural kinds (as I argue in chapter 2), if it is true that the properties of human kinds shift unpredictably as a result of feedback operating on an individual level, then this would imply that human kinds are not particularly useful natural kinds.

52 Compare this with depression, in which the overall heritability is between 30-40%. See section 2.3.2.
schizophrenia do not have schizophrenia (~6%), or are even conspicuously abnormal.\textsuperscript{53}

What is more, recent estimates suggest that there are upwards of 600 genes involved in schizophrenia, each conferring a small genetic risk (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Schizophrenia, in other words, is not a “genetic disorder” in the sense that \textit{Down’s Syndrome} and \textit{Huntington’s Disease} are genetic disorders. At best, schizophrenia exhibits a non-Mendelian pattern of inheritance in which there are no rare mutations with deterministic causal effects, but, rather, many single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) which each have a very small effect on the total risk of developing schizophrenia (Cf. Harrison, 2015).

What is more, recent evidence suggests that the genetic factors associated with schizophrenia might not even be specific to schizophrenia. In a population-based study examining more than two million families in Sweden between 1973 and 2004, Lichtenstein et al. (2009) found that first-degree relatives of individuals with either schizophrenia \textit{or} bipolar disorder were at increased risk of developing any one of these disorders. They estimate the comorbidity between the disorders due to additive genetic effects to be around 63%.

The massively polygenic and genetically comorbid\textsuperscript{54} nature of schizophrenia (at least as “schizophrenia” is currently conceived), coupled with the complex neurochemical

\textsuperscript{53} In his review of family and twin studies from 1920-1987, Gottesman (1991) places the risk for an individual developing schizophrenia when their identical twin has schizophrenia at 48%. For non-identical twins it’s 17%; 9% for siblings of a schizophrenic; 13% for children of a schizophrenic parent; and 6% for parents of a schizophrenic child.

\textsuperscript{54} Although I do not have the space to review the literature here, evidence is emerging that there is “causal comorbidity” between kinds of mental disorder. For instance, in chapter 2 (section 2.3.2) I suggested that the hippocampus (responsible for both emotional memory and spatial navigation) was implicated in depression. Recent evidence suggestions that it is also implicated in schizophrenia. Cf. Allen, Fung & Weickert (2015).
processes and variable anatomical findings discussed in chapter 3 (section 3.4.1), do not paint a picture of a kind which has a “stable biological structure”. Furthermore, once we expand our search to include known environmental risk factors of schizophrenia—maternal infections; maternal stressors, obstetrical complications; childhood trauma; social isolation; living in a city; and belonging to a minority ethnic group—we open up the possibility that the symptoms of schizophrenia are not predictable epiphenomenal effects of genes which alter neurodevelopment which thereby alters brain chemistry and circuitry. It is quite possible that there are complex gene-environment interactions (so-called “epigenetic factors”) in which the environment itself exerts its causal influence via epigenetic mechanisms (Cf. Gavin & Sharma, 2010; Roth et al., 2009; Rutten & Mill, 2009; van Os, Rutten & Poulten, 2009).

For instance, during neuronal migration cortical neurons begin their life near the ventricles and must make their way through the white matter to the outer regions of the neural tube. The first cortical neurons that migrate start forming the cortex, while neurons that migrate later must “crawl through” the existing layers of cortical neurons until they reach their final destination. This is a complex and delicate process which, in theory, could be disrupted by any number of factors: abnormalities in genes that sustain neural migration or help neurons grow and form connections, activity of neurotransmitters responsible for guiding migration and synapse formation, and even exposure to stressors (e.g., maternal infection). While the epigenetics of mental disorders are thus far poorly understood, the hypothesis that brain circuitry might be affected at different critical stages in brain development would explain why it is that both genetic and pre- and post-natal
environmental factors are strongly associated with schizophrenia.

Kendler et al. concede that these complex, dynamic processes may be at work in schizophrenia when they say that “risk genes work in cells that are parts of physiological systems that shape behaviors and environments…which, at all of these levels, feed back to influence genetic regulation and protein expression” (2011, 1145). Yet, if this is indeed the case, then it unclear in what sense schizophrenia is defined by “stable patterns of complex interaction between behavior, environment, and physiology”. After all, if it is true that many different, dynamic causal processes are at work in producing the symptoms of schizophrenia, and these causal processes might differ on an individual basis depending on when brain development is affected and depending on what kind of genes, stressors, and neural mechanisms are involved in particular cases, then it would seem that schizophrenia is not at all a stable entity or process in nature, but something that is continually morphing or shape-shifting in step with an ever-changing and adaptable human organism, from the level of genes to behavior.

4.5.4 Conclusion

It has long been thought that, if any kind of mental disorder were to be found to be a stable entity or process in nature, it would be schizophrenia. More than a century of research on schizophrenia has not borne out this conclusion. Indeed, as I have argued here, there are reasons to think that the signs and symptoms of schizophrenia in fact change across time and space and from individual to individual because of complex causal processes—including looping effects—that could have scarcely been imagined when the term ‘schizophrenia’ was first coined by Eugen Bleuler in 1908.
The hypothesis that psychopathological phenomena do not cluster together neatly should not be taken to imply that there are no natural kinds of mental disorder. After all, real kinds of mental disorder might just be a lot more complex than we originally thought, as I suggested in chapter 2.

Be that as it may, it is unlikely that hypothetical constructs such as “depression” and “schizophrenia” are going to capture this complex reality. For one thing, the project of classifying natural kinds of mental disorder by way of identifying stable clusters of properties has proven unsuccessful. Furthermore, given that a diverse cluster of observable properties is unlikely to have a shared causal explanation—and given that the clusters characteristic of many kinds of mental disorder often “come apart”—it is doubtful that this strategy will generally prove useful in the future.

4.6 Conclusion

Natural kind realists believe that the world is divided up into kinds—viz. naturally occurring clusters of observable properties undergirded by an unobservable causal base—and that we can know which clusters are real and which are bogus by consulting our best theories about the world. In this chapter I argued that these twin pillars of natural kind realism are simply not congruent with what we know about mental disorders, both historically and empirically. Not only has the project of classifying natural kinds of mental disorder by way of identifying clusters of properties failed historically; it’s likely to fail as a general strategy for psychiatric classification given the nature of psychopathological phenomena.
CONCLUSION

Throughout this dissertation I have argued that psychiatric kinds are not natural kinds, since we cannot even explain causally how it is that the signs and symptoms that are characteristic of putatively distinct kinds of mental disorder are produced. Without knowing the true causal structure of a particular kind of mental disorder, it is impossible to know whether the properties definitive of a psychiatric kind do indeed cluster together, or whether these groupings are merely a reflection of our predilections to group certain properties together.

Indeed, this conclusion is further strengthened by the fact that many of our current psychiatric kinds, such as schizophrenia, were constructed not on the basis of well-formed theories about the mind/brain, or the nature of psychopathological phenomena, but on theoretical hunches made over a century ago.

Compare this situation with that of biology. While there are still deep disagreements in biology about how species ought to be classified—e.g. whether one should appeal to morphological similarities or genetic similarities (Cf. Khalidi, 2013, 75; Kornblith, 1993, 51)—biology has an overarching theory from which to work. Outside of psychoanalysis, psychiatry does not have an overarching theory—and, as we have seen, theories of particular kinds of mental disorder that have been proposed over the years have presupposed the existence of the very thing they are attempting to theorize about.

While most philosophers have argued that psychiatry has not progressed because psychiatric classification has for the most part been based on atheoretical observational criteria, this diagnosis does not get things quite right. There is a sense in which the failure
of psychiatry to meaningfully progress was built in from the start. First, the discipline of psychiatry has historically proceeded in accordance with the motto: classify first, ask questions later. Second, nosologists have traditionally assumed that identifying clusters of co-occurring properties would carve up nature in such a way that natural kinds of mental disorder would be forthcoming.

There are ample reasons to believe that this “top-down” approach to the psychopathological domain is wrongheaded. For one thing, it hasn’t worked historically. Second, evidence for the complex and chaotic nature of psychopathological phenomena (i.e. comorbidity, instability, and causal complexity) suggests that, as a general strategy, it will likely not work in the future. In his late-life reflections on the status of psychiatric classification Kraepelin expressed similar doubts about the way in which he and his contemporaries were going about trying to identify natural kinds of mental disorder:

It has been repeatedly stated that research in clinical psychiatry has almost reached a dead end. It has been the practice for some time now to separate one disease from another with respect to causes, clinical findings, course, outcome and pathological findings. This practice has now outgrown its use. We must explore new avenues (Kraepelin, 1920, 509).

Today, we are faced with the very same question that Kraepelin pondered almost a century ago: which avenues should we explore? As far as classification is concerned, there are many different approaches one could take. The fact that a variety of psychopathological phenomena do not cluster together neatly would seem to suggest that dimensional, rather than categorical, approaches to classification are most appropriate. But mixed categorical/dimensional models are also plausible. Consider depression, for instance. It might be the case that most instances of clinical depression fall on a
continuum with normal unhappiness, but that the most severe form of clinical depression—which has variously been called *melancholia, endogenous depression*, or *nuclear depression*—is a kind unto itself (Cf. Haslam, 2002, 211; Horowitz & Wakefield, 2007).

There is also the question of utility. Even if categorical approaches to psychiatric classification do a poor job of capturing the complex reality of psychopathological phenomena, they might still be useful. For example, the category of ‘hypertension’ is arbitrary in the sense that it falls on a continuum with normal blood pressure variation within the general population. Nevertheless, given the serious health risks associated with high blood pressure, there are pragmatic reasons for drawing a categorical distinction between normal blood pressure and hypertension. A similar case can be made for psychopathological phenomena which are dimensional in nature. Although clinical depression might very well exist on a continuum with normal sadness, surely not all varieties of sadness require treatment.

As has been argued at length elsewhere (Kendler & Jablensky, 2003; Zachar, 2000, 2002, 2014a, 2014b), psychiatry is as much an applied science as a theoretical one, and an enduring problem in psychiatric classification is to strike a delicate balance between the needs of researchers, clinicians, patients, and third-party insurance companies. Although one might suggest that, for far too long, psychiatric classification has been in the service of too many masters, this is unlikely to change in the near future given the indispensability of current psychiatric categories to both clinicians and patients. After all, even if paradigmatic psychiatric kinds such as schizophrenia and depression do not reflect real kinds in nature, patients still need a diagnosis in order to receive treatment
and disability, and clinicians still need to make treatment decisions while they wait for the science to catch up.

In my mind, it is high time to do something that, outside of psychoanalysis, the discipline of psychiatry has historically been reluctant to do: table the question of psychiatric classification for the moment and instead direct our attention back to the phenomena that are still in dire need of explanation. In particular, a more tractable approach for psychiatric research would be to first explain how particular kinds of experiences and behaviours—e.g., hallucinations, delusions, disorganized thinking, mania, low mood, etc.—are produced. Once we have figured this out there may be no “schizophrenia” or “bipolar disorder” left to explain.

This approach is “bottom-up” rather than “top-down” in the sense that it suspends judgment concerning whether or not certain psychopathological properties (e.g., delusions and hallucinations) cluster together, and instead focuses on explaining one thing at a time. The advantages of this approach are clear. For one thing, explaining, e.g., delusions—or even a particular type of delusion—is much more manageable than explaining all of delusions, hallucinations, disorganized thinking, catatonic behaviour, and comorbid states of depression, anxiety, and personality. This line of attack not only sidesteps the problem of heterogeneity and comorbidity—at least at the level of symptomatology—but is also much more likely to yield promising causal explanations given that, on the face of it, delusional states are distinguishable from hallucinatory states, as well as abnormal behaviours such as disorganized speech and catatonic behaviour.
Perhaps the greatest advantage of this approach, however, is that it is not predicated on the specious assumption that differences between psychopathological phenomena and non-psychopathological phenomena are differences in kind, not degree. This opens up new theoretical avenues in which explanations of psychopathological phenomena can proceed from more general theories about how the mind/brain operates, a conceptual framework which has been perspicaciously lacking in psychiatric research.

Although a complete illustration of this kind of explanatory approach is beyond the scope of this dissertation, an example will suffice as a bookend to this essay. In *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought* (2014) Phillip Gerrans offers a novel explanation of delusion. Gerrans’ principal thesis is that delusions arise when “*default cognitive processing, unsupervised by decontextualized processing, is monopolized by hypersalent information*” (ibid., 38; emphasis in original).

According to Gerrans, delusions are “default thoughts” which represent the operation of the Default Mode Network (DMN). The DMN is a cognitive system which allows all human beings the ability to simulate/imagine experiences (both past and future) in the absence of stimuli. It evolved in order to allow an organism to simulate/imagine scenarios in response to “salient information”—namely, that which is relevant to an organism at any given time—for the purpose of planing, deliberation, and reflection. In particular, default circuitry enables human beings to make salient information intelligible by providing a *subjectively adequate* autobiographical narrative for experience.
Under normal waking conditions default cognitive processing is supervised by higher cognitive processes, which makes possible “decontextualized processing”—i.e. the capacity to treat one’s beliefs as hypotheses open to falsification. In cases of delusion, however, decontextualized processing is inoperative either because the neural mechanisms responsible for decontextualized processing (e.g., the right dorsolateral prefrontal cortex) are damaged or hypoactive, or because decontextualized processing is overridden by a hyperactive default network. The default network can become hyperactive when the salience system, which is underwritten by dopaminergic activity, makes some kinds of information “hypersalient”—i.e. extremely relevant or significant to an organism, dominating attention, working memory, and executive function.

For example, Gerrans argues that Capgras delusion (the delusion that a familiar person has been replaced by an imposter or lookalike) is the consequence of the default network incorporating an anomalous experience (in this case, the experience of seeing a familiar person in the absence of an autonomic affective response) by producing the delusion: “My wife has been replaced by a lookalike”. Conversely, Fregoli delusion (the delusion that one is being followed by a familiar person in disguise) is the consequence of the default network incorporating the anomalous experience of seeing a stranger and having a strong affective response characteristic of seeing a familiar person.¹ Both of these strange delusions begin as default thoughts meant to make an anomalous experience

¹ According to Gerrans (ibid., 103-104), in the former case the anomalous experience is caused by a lesion to pathways connecting face recognition circuitry (the fusiform gyrus) and affectivity (the amygdala), and in the latter case the anomalous experience is caused by hyperactivity in the pathways that typically trigger an autonomic affective response to a familiar person.
intelligible to an individual, and, due to compromised decontextualized processing, are further entrenched in an agent’s psychology.

Gerrans also suggests that his account can be extended to delusions which are common in “schizophrenia”, such as paranoid delusions (the delusion that others are out to harm you), referential delusions (the delusion that external events or objects hold a special significance), and grandiose delusions (the delusion that one is exceptional or powerful). In these instances default thoughts are triggered by experiences which have been intensified by the salience system, effectively hijacking cognitive, behavioural, and metabolic resources, and thereby preventing decontextualized processing.

One question that arises from Gerrans’ account of delusion is why it is that delusions have the content that they do. In delusions of misidentification (e.g., Capgras delusion and Fregoli delusion) it is fairly obvious that the experience which triggered a default thought, though elaborated upon, is more or less reflected in the delusional content. In schizophrenic-like delusions, however, what kinds of anomalous experience or hypersalient information could lead one to produce the default thought that they are being followed by the Central Intelligence Agency (CIA)?

Interestingly, in Suspicious Minds: How Culture Shapes Madness (2014) Joel Gold and Ian Gold postulate that human beings possess a “suspicion system” which evolved over time for the purpose of enabling adaptive responses to environmental threats, such as others’ malign intentions. In the case of delusions—which they argue are invariably about others or oneself in relation to others—the suspicion system is in overdrive, causing one to see malign intent where there is none (e.g., believing that one is
being followed by the CIA). To put this hypothesis in the language of Gerrans’ account, a hyperactive suspicion system can make certain kinds of information (e.g., body language, facial expressions, speech, others’ behaviour) hypersalient. As a result, default thoughts related thematically to oneself vis à vis others are produced, and, if unsupervised by decontextualized processing, eventually become part of an agent’s psychology.

While the claim that delusions are default thoughts might turn out to be false, the attractiveness of this approach to psychiatric explanation is that it seeks to explain a known phenomenon (delusions) rather than an unknown, hypothetical disease (schizophrenia). For reasons already mentioned, we are much more likely to make theoretical progress this way.
References


Fromm-Reichmann, F (1940). Notes on the mother role in the family group. *Bulletin of*


(originally published in 1964).


Linnaeus, Carolus (1758). *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis* (10th ed.). Stockholm: Laurentius Salvius.


four-class, four-dimension model of schizophrenia and affective psychoses.


Sullivan, H. S (1931). Environmental factors in the etiology and course under treatment


CURRICULUM VITAE

NAME:

Nicholas Slothouber

POST-SECONDARY EDUCATION & DEGREES:


Bachelor of Arts. Brock University. Honors Philosophy and Psychology Programme. Degree conferred June 2009, with first-class distinction, Distinguished Graduating Student in Philosophy.

HONOURS AND AWARDS:

2016 Student Essay Prize, 1st place, Canadian Philosophical Association Annual Congress, Calgary, AB, Canada
2012-2015 Western Graduate Research Scholarship, University of Western Ontario
2012 Dean’s Entrance Scholarship, University of Western Ontario
2009-2010 Graduate Fellowship, University of Toronto
2009 Distinguished Graduating Student in Philosophy, Brock University
2008-2009 Doris Senior Award in Philosophy, Brock University
2005-2009 Deans Honours Certificate, Brock University
2004-2007 Brock Returning Scholars Award, Brock University

RELATED WORK EXPERIENCE:

2019 Bioethics (PHIL 2P95), Brock University (teaching assistant, four sections) (Michael Berman)
2019-2020 Introduction to philosophy: problems of philosophy (PHIL 1F94), Brock University (teaching assistant, two sections) (Xavier Scott)
2016-2020 Brain and behavior II (NEUR/PSYC 2P37), Brock University (course coordinator and laboratory demonstrator, several sections) (Julie Baker & Dawn Good)
2015-2019 Brain and behavior I (NEUR/PSYC 2P35 & 2P36), Brock University (course coordinator, laboratory demonstrator, and teaching assistant, several sections) (Dawn Good)
2018 Bioethics (PHIL 2P95), Brock University (teaching assistant, four sections) (Brian Lightbody)
2016-2018 Introduction to logic (PHIL 2P95), Brock University (teaching assistant, three sections) (Michael Berman)
2016 The ethics of professional relationships (marker-grader), Western University (Barry Hoffmaster)
2015 Questions of the day (PHL 1305G), Western University (teaching assistant, one section) (Dean Proessel)
2015 Philosophy of psychiatry (PHL 2044G) (occasional lecturer), Western University (Louis Charland & Gillian Barker)
2014-2015 Introduction to critical thinking (PHL 1200), Western University (teaching assistant(marker-grader, online course) (Chris Viger & David Bourget)
2013-2014 Introduction to critical thinking (PHL 1200), Western University (teaching assistant, one section) (Ryan Robb)
2012-2013 Introduction to philosophy (PHL 1020), Western University (teaching assistant, one section) (John Thorp)
2012 Personality and Individual Differences (PSYC 2P25), Brock University (teaching assistant, four sections) (Michael Ashton)
2011 Counseling and Psychotherapy (PSYC 4P79), Brock University (teaching assistant, one section) (Dorothy Markiewicz)
2011 Learning and Behavior (PSYC 2P45), Brock University (teaching assistant, two sections) (William Bradley)
2011-2012 Abnormal Psychology (PSYC 3F20), Brock University (teaching assistant, two sections) (Andrew Dane)
2011-2012 Introduction to Philosophy: Human Nature (PHIL 1F91) (teaching assistant four sections) (Calvin Hayes)
2011-2012 Introduction to Philosophy: Problems of Philosophy 1F94 (teaching assistant, two sections) (Brian Lightbody)
2007-2011 Brain and behavior I (NEUR/PSYC 2P36), Brock University (lab demonstrator and teaching assistant) (Dawn Good)
2008-2012 Brain and behavior II: neuroscience/psychology 2P37, Brock University (lab demonstrator and teaching assistant, several sections) (Dawn Good)
2010 Probability and Inductive Logic, University of Toronto (teaching assistant, one section) (Colin Howson)
2009-2010 17th and 18th Century Philosophy, University of Toronto (teaching assistant, one section) (Marleen Rozemond & Ulrich Schloesser)