UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

REALISM, NATURAL KINDS, AND ATTENTION DEFICIT HYPERACTIVITY DISORDER

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

DAVID M. SPINDLE
Norman, Oklahoma
2017

REALISM, NATURAL KINDS, AND ATTENTION DEFICIT HYPERACTIVITY DISORDER

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PHILOSOPHY

BY

_____
Dr. Stephen Ellis, Chair

_____
Dr. Lauren Ethridge

_____
Dr. James Hawthorne

_____
Dr. Neal Judisch

_____
Dr. Martin Montminy

Acknowledgments

I am indebted to many people at the University of Oklahoma. I thank James Hawthorne and Stephen Ellis for their patience and assistance in completing this dissertation. I am grateful to Neal Judisch and Martin Montminy for serving on my committee since its inception, and to Lauren Ethridge for agreeing to do so at the very last minute. I also thank Sherri Irvin, who many years ago worked with me on my first foray into the philosophy of psychiatry. Dr. Irvin's guidance sharpened my then ill-formed thoughts about the field, and helped me move forward with my own research. Finally, I thank my colleague and friend Madeline Martin-Seaver, a consistent source of solidarity throughout this endeavor.

I must acknowledge Dr. Brint Montgomery of Southern Nazarene University, the first philosopher I ever knew. My thanks as well to Dr. Joel Nigg, Director of the Division Psychology at the Oregon Health and Science University, who took time out of his busy schedule to correspond with me about ADHD research.

Finally, I express my deepest gratitude to my family. First, my parents Debra and Randall Spindle, for their steadfast support; and my wife – Angela Jerome Spindle, who sacrificed far too much while I toiled away on this project, but never hesitated to assure me that I would ultimately see it through.

Table of Contents

List of Illustrations

Abstract

Realism about mental disorders is a perennial area of dispute, but the controversy burns especially intensely for Attention Deficit Hyperactivity Disorder (ADHD). In this dissertation, I clarify what is at issue in these debates, surveying how realists have typically argued for mental disorder realism: the definitional debate about health and illness. I argue that the realist need not be committed to the terms of the definitional debate and recommend that a better approach is to show that mental disorders are natural kinds. While there are many accounts of kind-hood on offer, I adopt Richard Boyd's homeostatic property cluster (HPC) theory of kinds, which I interpret through the philosophy of neuroscience literature on mechanisms. In sum, I conclude that if ADHD is a natural kind – and thus *real* – then individuals diagnosed with the disorder should be sufficiently similar with respect to an underlying cognitive neurobiological mechanism. To determine whether ADHD individuals are similar in this way, I consider the question through Russell Barkley's Executive Function Model of ADHD. Relying primarily on the cognitive neurobiological research, I argue that there is now reasonable evidence to conclude that the DSM classification of ADHD corresponds not to a single natural kind, but several. Thus, ADHD is thus *real*.

Chapter 1: Introduction and Survey of the Project

1.1: Introduction

According to the predominant medical understanding of the condition, ADHD is a neurodevelopmental disorder affecting between 5-11% of children and 8% of adults in the United States, and between 2.5-5% of children and 1.2–7.3% of adults worldwide (Hinshaw and Scheffler 2014 and Faraone et al. 2015). Beyond this, much is contested about ADHD. Some of these controversies (e.g. vague diagnostic boundaries, high co-morbidity) aren't unique to ADHD, but could very well apply to any condition in the Diagnostic and Statistical Manual of Mental Disorders (DSM). But if we had to select a thematic core in the controversies about ADHD, it would most certainly include the following items: a) it is one the most common medical conditions in children, b) by far, males compose the majority of ADHD diagnoses, c) teachers are frequently involved in initiating, and therefore steering, families towards seeking medical treatment, d) stimulant medication, which under certain conditions exerts similar psychoactive effects as cocaine, is often prescribed to curb ADHD type behaviors, and e) the high rate of prevalence in the United States. While all of these are involve important questions that are no doubt worthy of attention, in the present project I am

concerned with what I believe to be the *fundamental* question: whether ADHD is *real*.

Doubts the reality of ADHD have persisted ever since the public at large became aware of the disorder. While we cannot say this about many other mental disorders (e.g. depression, schizophrenia), for ADHD we can date precisely when that occurred: June 29th, 1970. On that day, the *Washington Post* reported on an academic enhancement program called "STAR", which was supposedly implemented by the Omaha public school system. In addition to providing tutoring and other types of educational assistance, the story claimed that school administrators were coercing parents into dosing their children with "behavior drugs", including Ritalin and Dexedrine. Robert Maynard, who penned the story, estimated that "5-10%" of the district's 62,000 students were taking psychoactive medication (Ross and Ross 1976, 20). The *Post* story created such a public outcry, that the House Committee on Government Operations convened a public hearing to investigate whether the federal government had any role in dispensing behavior modification drugs to public schools. The committee's investigation discovered a number of inaccuracies in Maynard's story, but that came too late to stop the feeding frenzy in the press (Mayes et al. 2009, 64).

In the aftermath of the *Post* story, Americans learned about a condition called "minimal brain dysfunction" (MBD), which pediatricians had been diagnosing since the 1940's. MBD was presumably biological in origin, and its symptoms included excessive motility, inattention, and impulsiveness. Questions about this "new" disorder immediately arose. In an August 1970 letter to the *New York Review of Books*, for instance, educational reformer John Holt argued that MBD was no more than, "…currently fashionable quackery, which blames on the nervous systems of children the stupidities and inhumanities of our schools…" (Holt 1970). Five years after Holt's letter, the journalists Peter Schrag and Diane Divoky published the first book length critique of hyperactivity *The Myth of the Hyperactive Child and Other Means of Child Control* (1975). Schrag and Divoky decried what they characterized as the growing "therapeutic state", a network of mental health and educational professionals seeking to exercise ever more control over children's behavior by "making up" new medical disorders. That same year, sociologist published *The Discovery of Hyperkinesis: Notes on the Medicalization of Deviance* (1975). Although more measured than Holt, Schrag and Divoky, Conrad nevertheless implied that hyperkinesis reflected not a property internal to a child, but rather was product of the social environment, "If we focused our analysis on the school system we

might see the child's behavior as symptomatic of some 'disorder' in the school or classroom situation, rather than symptomatic of an individual neurological disorder," (Conrad 1975, 20).

During the 1990's, at the height of the "Ritalin wars", these and other claims about the dubious "reality" of ADHD were repeated *ad nauseum*. Thomas Armstrong inflammatorily compared ADHD to drapetomania, a 19th century medical diagnosis "discovered" by southern physician Samuel Cartwright. Drapetomania referred to the "irrational" tendencies found among some slaves to run away from their owners (Armstrong 1995). Several conservative intellectuals argued that ADHD was the product of a feminist conspiracy to "androgenize" young boys (Will 1999, Fukuyama 2001) and suppress their natural exuberance (Sowell 2001). In *The Hyperactivity Hoax* Sydney Walker blamed the rise of new ADHD diagnosis on the parents' unwillingness to properly discipline their children (Mayes et al. 2009). Fred Baughman's *The ADHD Fraud* argued that greedy pharmaceutical companies had concocted ADHD in order to sell medications and the disorder lacked any consistent biological or neurological basis (Baughman and Hovey 2006). These debates continue into the 21st century, and show no sign of stopping. In 2003, for example, UK psychiatrist Sami Timimi claimed that since there were no, "…specific cognitive, metabolic or neurological markers and no

medical tests for ADHD," the disorder was best understood as a "cultural construct" (Timimi 2004, 8-9). As recently as 2013, neurologist Richard Saul published a volume unsubtly titled *ADHD Does Not Exist* (2013).

Despite this long-standing controversy, philosophers - both within the philosophy of psychiatry and in other relevant fields – have remained largely silent about ADHD and broader questions concerning its "reality." When I started research for this dissertation in the summer of 2013, there was no philosophical literature about this issue to consult. Since then two works have appeared that are worth mentioning at the outset. I will discuss these briefly and then go on to outline the scope and aims of this project.

The first is Susan Hawthorne's monograph *Accidental Intolerance* (2014), which is study about how ADHD is understood across different institutional contexts – clinical, research, educational – and the ways in which the values embedded in those contexts lead to the unintentional stigmatization of individuals with the disorder. These are important issues and Hawthorne's treatment of them is excellent. But my interest in ADHD is different from hers. Save for the naturalist-normativist debate in the philosophy of medicine in chapter 2, I do not have much to say about ADHD's ethical implications, whatever they might be.

More relevant is the brief discussion in Muhammad Ali Khalidi's *Natural Categories and Human Kinds* (2013, 188-199). Both Khalidi and I

cover similar ground, although our purposes are different. Khalidi's overall aim is to provide general theory of natural kinds. His book makes use of several case studies (e.g. lithium, polymers, cancer) in order to clarify particular aspects of his theory. While he does conclude that, "…there is at least some evidence to suggest that [ADHD] is a psychiatric natural kind" (2013, 199) the primary purpose of his discussion of ADHD is to show, "…that the standards and principles for identifying natural kinds in a social science like psychiatry are similar in their general features to those deployed in the other sciences that have been examined," (199).

My discussion of ADHD in chapter 4 is a detailed study of the ADHD research, particularly from a cognitive neurobiological perspective. I argue that, given the present state of the field, there is sufficient evidence to indicate that ADHD *is* a natural kind, or more precisely several natural kinds lumped together by the DSM. So, qua natural kinds, Khalidi's aims are much wider than my own; but qua ADHD, my aims are much more specific and detailed than Khalidi's interest the subject.

1.2**: Overview of Chapter 2**

As we observed, many controversies surrounding ADHD concern its "reality." This is a perennial problem for *all* mental disorders. I begin chapter 2 to by clarifying what is meant by the assertion that this or that condition is (or is not) "real." A standard way to parse that claim is in terms of mind-independence. But without further clarification, simple mind-independence poses a special problem for distinctively *mental* phenomena. I argue that George Graham's notion of "act-of-classification" (AC) independence is the relevant sense behind the realism-irrealism debate. I then proceed to outline the standard way in which philosophers have pursued realism about mental disorders, which I call the definitional debate about disease.

The definitional debate concerns whether our disorder judgments are value-laden or "objective." The standard realist position in the definitional debate is that both mental and physical disorders involve a dysfunction and that dysfunction is an entirely theoretical (i.e. value-free) notion. Therefore, whether mental disorder realism can be defended within the terms of the definitional debate depends on whether there is a value-free analysis of function-dysfunction to be had. I will show that the reason mental disorder realism's fortunes are yoked to the definitional debate is largely a matter of historical circumstance. Specifically, two crises in the

1970's that threatened to undermine institutional psychiatry's legitimacy as a medical discipline: the anti-psychiatry movement and the gay liberation movement. The American Psychiatric Association (APA) tried to resolve both of these crises by formulating a general definition of disorder. This, in turn, influenced the debate within the philosophy of medicine. I then proceed to outline three main positions in the definitional debate: simple naturalism, two-stage views, and simple normativism. Along the way, I will show how various participants in this debate largely assume that AC-independence is the relevant criteria for determining whether a given mental disorder is "real." I conclude chapter 2 by arguing that nothing in AC-independence requires us to conceive of disorders in terms of function-dysfunction, and so whether or not those concepts can be given a value-free analysis is orthogonal to the larger realist enterprise. If sound, that argument frees us up to pursue other approaches, which I proceed to do in chapter 3.

1.3**: Overview of Chapter 3**

Another way to meet AC-independence for mental disorders is to argue that they (or some subset thereof) are *natural kinds*. Other than citing paradigmatic examples of natural kinds – e.g. gold, tigers, humans, and electrons – it is difficult to specify what else natural kinds are like because philosophers have invoked them to settle many different sorts of

issues, including those involving meaning and reference, explanation, induction, the purpose of science, the nature of laws, the fundamental ontological structure of the universe, and much else besides. As Paul Churchland correctly observes, "…different writers confront a different 'problem of natural kinds' depending on which background issue is for them the principal issue at stake," (1992, 281).

So to get clear on the notion of natural kinds, I begin with essentialism, which is the traditional framework in which philosophers have understood natural kinds. Originating in Aristotle's work, the two core tenants of essentialism say that natural kinds are a) classes of individuals united by a shared set of properties, or *essences* and b) essences are necessary and sufficient for membership to the kind. In the modern era, philosophers have imposed several further conditions about what a kinds essences are supposed to be like.[1] I will consider two. One says that kind-essences are identical to its microstructural properties; another says that a kind's essential properties are intrinsic.

I argue that essentialism is not a suitable theory of natural kinds. Before recapitulating argument here, I say something about the perspective informing my assessment of the view. I approach the question

[1] This is not intended to suggest there is consensus among philosophers about what these further conditions might be. There isn't.

of natural kinds from broadly naturalistic perspective, which is not so much a commitment to a core doctrine, but rather a particular method for evaluating philosophical accounts of kind-hood. Khalidi provides a particularly clear statement of this method (2013).

Khalidi argues that a theory of natural kinds should aim towards a sort of reflective equilibrium between our scientific and philosophical commitments. This suggests a set of defeasible constraints assessing whether a particular theory of natural kinds is adequate. First, on balance, philosophical accounts of kind-hood that admit scientific categories we paradigmatically count as natural kinds are more adequate than those that do not. Second, philosophical accounts of kind-hood that take on board, as much as possible, traditional philosophical claims made on behalf of kind-hood are preferable to accounts that do not. There will inevitably be instances when these two constraints pull in different directions, but reflective equilibrium recommends that in such cases, we revise our prior scientific and philosophical beliefs until we can reconcile our commitments about kinds (Khalidi 2013, 3-4).

Now, there is always a danger that in articulating such a criteria for assessing different theories, one inadvertently begs some important question against their opponents. Perhaps one might say that philosophy is and should be autonomous from the natural sciences, and so would reject

any scientific constraints about what is acceptable philosophical doctrine. So, perhaps there are limits in how inclusive the approach outlined above really is. I would just add that with respect to essentialism, the method of reflective equilibrium is not *obviously* question begging.

Consider that every single one of the classic examples of *essentialist* kinds – lemons (Putnam 1975, 158-159), tigers (Kripke 1980, 120-121), gold (1980, 123-125) and water (1980, 128) – are components of actual scientific taxonomies in biology and chemistry. Arguably, essentialism provides a metaphysical account of, say, why members of these kinds resemble one another in certain respects. Namely, kinds have a set of essential properties that reliably lead to the co-instantiation of other properties, which is why members of the same kind are similar. Thus, we can presumably discover what these essences are and how they sustain these systematic resemblances among its members. But since any item in a scientific taxonomy is not immune to further revision, it is also true that our beliefs about it could turn out to be wrong. If a philosophical theory of kinds purports to make sense of actual scientific cases, then perhaps that view would similarly require revision, should new facts about the scientific cases emerge. And this is *exactly* what the method of reflective equilibrium demands.

A standard critique of essentialism is that many scientific categories we think of as kinds, such as species, do not meet its demands. This is true, but so stated that critique implies that essentialism leaves us with an impoverished ontology. Many versions of essentialism are actually essentialism plus microstructuralism. For individual species (e.g. homo sapiens) essentialism plus microstructuralism rules all of those categories out, but further implies that each individual member – or, what we thought of as an individual member (e.g. Jim, Martin, Steve) – are actually natural kinds in their own right. This suffices to make essentialism about kinds a non-starter, since it is radically at odds with how science presently carves up the world. I also consider some more specific critiques such as the notion that a kind's essential properties must be intrinsic. For one version of essentialism – Brian Ellis's "scientific essentialism" – his account of intrinsicality actually undermines his flagship example of essentialist kind: chemical kinds. But the larger problem for that thesis is that simply too many scientific categories are individuated by properties that *no* account of intrinsic property would satisfy.

I then present a better view of natural kinds, Richard Boyd's homeostatic property cluster (HPC) theory. Boyd explicitly identifies his account as a competitor to essentialism, "The natural kinds that have unchanging definitions of intrinsic necessary and sufficient

conditions…are an unrepresentative minority of natural kinds (perhaps even a minority of zero)," (Boyd 1999, 169). An HPC is cluster of contingently associated properties that tend to be reliably co-instantiated by its members, or instances, due to a causal mechanism (or network of causal mechanisms). The property clustering is often imperfect, so unlike essentialism, the HPC-theory to declines to identify any single subset as necessary and sufficient for membership. However, HPC kinds are still fit for scientific investigation and can support causal generalizations, since the presence of the mechanism tends to keep the cluster in a relative state of homeostasis. It is not clear whether the HPC-theory works as a "general" theory of kind-hood, but it is widely agreed that it accommodates several actual categories in the "special sciences", which suffices for our purposes.

Boyd provides a lot of interpretive leeway in understanding how an HPC-kind is "put together" and which properties count in establishing the identity of a given HPC. I thus further clarify the sense in which ADHD is, or might be, an HPC by consulting cognitive neurobiology. Roughly, this view holds that complex cognitive capacities – such as those likely involved in the production of mental disorders – are implemented by mechanisms, defined as an organized hierarchical physical system composed of components and their activities. On this view, ADHD is an

HPC-kind if its symptoms result from, or are sustained by, a similar organizational structure, specified in terms of a hierarchically organized cognitive neurobiological mechanism. I conclude chapter 3 by surveying how cognitive neurobiologists explain, model, and investigate mechanisms. This will set the context for interpreting the discussion of Russell Barkley's Executive Function (EF) model of ADHD in chapter 4.

1.4**: Overview of Chapter 4**

Chapter 4 brings us to the central question of this study: whether ADHD is "real" in the sense that the symptoms associated with that category are causally sustained by similar cognitive neurobiological structure. I will consider this question through Russell Barkley's EF-model of ADHD, but I first begin with a brief presentation of the DSM classification of ADHD. I then consider an objection against the DSM categories that poses a particular problem for my project.

This argument has been made in many different ways, but following Kathryn Tabb, we can refer to it as the problem of diagnostic discrimination (Tabb 2015). According to Tabb, diagnostic discrimination refers to an assumption on the part of psychiatric researchers that the DSM categories groups individuals together in such a way that enables the discovery of facts about causal mechanisms. She goes on to argue that given the DSM's descriptive and operational approach to classification,

this assumption is most assuredly false. This poses a problem for my argument because it relies on the cognitive neurobiological literature on ADHD, all of which relies on the DSM categories to select populations of individuals to study. While I am sympathetic to this general line of critique, I believe the objection's force is limited by a few considerations. First, the DSM categories are a varied bunch, and so whether any single one can function as a useful target of scientific research is probably better decided on a case-by-case basis. Perhaps as our understanding of the mechanisms underwriting mental disorders increases, a future psychiatry will cut the pie up differently than we would on the basis of our present (and no doubt) imperfect understanding. But that doesn't mean there is nothing to be learned from the epistemic circumstances we are now in.

That said, with respect to ADHD, I argue that three lines of evidence make it prima facie reasonable to treat it a suitable object of scientific inquiry. First, the disorder is highly familial, if not highly heritable. Second, factor analysis and other statistical measures consistently show that the two symptom dimensions in ADHD – inattention and hyperactivity-impulsivity – are separate, yet highly correlated in individuals diagnosed with the disorder (Willcutt, et al. 2012). Third, the symptom complex manifested in ADHD can be reliably differentiated from other closely related disorders, such as Oppositional

Defiant Disorder (Taylor and Sonuga-Barke 2008). While by no means sufficient to conclude that ADHD is an HPC-kind, this serves as defeasible evidence warrant moving forward.

Russell Barkley's EF-model of ADHD is a highly influential account of the disorder. As such, it has been extensively studied and provides a useful vehicle for navigating the already voluminous research literature on ADHD. According Barkley, ADHD is primarily due to a prefrontally mediated, maturational delay in the behavioral inhibition system, which in turn causes a suite of cognitive and behavioral problems in the executive system. After specifying the basic outlines of the EF-model in more detail, I show how it satisfied many of the properties of a mechanistic model, which we discussed in chapter 3. I then go on to assess its plausibility as a *general* explanation for all (or most) of the individuals diagnosed with ADHD. After surveying various parts of the research literature that are supportive of the view, I conclude that a subset ADHD individuals maybe fairly characterized as implementing the mechanism described by the EF-model, or something near enough. However, the EF-model fails to describe majority of individuals with ADHD, which suggests that the present DSM classification of the disorder does not correspond to a single natural kind, not even in the flexible sense specified by Boyd's HPC-theory. I then survey another line of research, which

provides compelling evidence to suggest the presence of alternative mechanism underwriting the symptoms associated with ADHD. Broadly speaking, this deficit implicates the reward and motivational systems. Although there are competing models of the motivational components involved in ADHD, will focus on Edmund Sonuga-Barke's Delay Aversion (DAv) model. There is good evidence to suggest the presence of another mechanism involved in the production of ADHD and, so, evidence to suggest the presence of a distinct HPC-kind: one implicating the executive-inhibitory systems, the other the reward-motivational system.

The narrow conclusion from this study is that the DSM classification of ADHD is not a natural kind. However, that provides no comfort to the irrealist. The extension of ADHD isn't empty; it's not like phlogiston. A more apt comparison is between ADHD and, say, declarative and procedural memory (Craver 2002; Squire 2009), or sensory and affective pain processing (Hardcastle 1999). In each of these cases, investigators began with the assumption that the extension of the category picked out a single natural kind, only to later find out that the antecedent taxon classified distinct kinds together. Insofar as these - albeit still imperfectly understood - capacities are considered to be *real* structures within the natural firmament, ADHD fairly earns its place among them.

Chapter 2: Realism and the Definitional Debate

2.0: Introduction

I begin this chapter with a discussion of realism about mental disorders, the purpose of which is to get a rough idea of the sense in which mental disorders are claimed to be real (or unreal). I then discuss the primary vehicle through which philosophers have adjudicated that issue, which I call the definitional debate about disease. I will briefly describe the historical circumstances surrounding the definitional debate, which will go some way towards explaining why the case for realism is thought to hinge on providing an analysis of notions such as disorder, function, dysfunction, and so on. I then outline three main positions in the definitional debate: simple naturalism, two-stage views, and simple normativism. Along the way, I demonstrate how the sense of realism I lay out in the next section animates these debates. After surveying the definitional debate and describing how it operates in the literature, I conclude this chapter by showing how the case for realism within the definitional debate is tied to deriving a non-evaluative definition of the function-–dysfunction distinction. I will argue that the realist need not worry about whether that issue can be resolved, which should suffice to motivate another way of arguing for mental disorder realism. I take steps towards pursing that argument in the next chapter.

2.1**: Realism: Mind-Independence**

For some category (or concept, or item, etc.) **x**, mind-independence is a perfectly natural way of understanding the assertion "**x** is real" – i.e. **x** is not just an idea in our minds, but exists independently of our ideas. That said, two intuitively plausible interpretations of mind-independence settle the issue of mental disorder realism in a rather trivial fashion. Let us begin with a simple statement of it as an ontological thesis: **x** is real if its existence does not depend on minds. Since we are talking about mental disorders here, one can see how this rough and ready formulation is inadequate. Whatever the precise relationship between the two consists of, mental disorders would seem to depend on minds in that there can be no mental disorders if there are no minds to be disordered (Samuels 2009). Another reading of mind-independence is causal independence: **x** is real if it is not causally sustained by the existence of minds (Page 2006). A great many things are causally sustained (in part) by our mental activity – climate change, border collies, artifacts (Samuels 2009, Magnus 2012). So too are psychological states. Again, the causal reading of mind-independence settles the matter trivially.

As we will see, the relevant sense of "mind-independence" in debates about mental disorder realism is what George Graham calls "act-of-classification" (AC) independence (2014). While not offering anything

like a comprehensive definition, the notion is intuitive enough: AC-independence requires that, for some condition **x**, **x** is a real if there is a positive body of fact about it that exists independently of our labeling conventions. Graham illustrates the idea by contrasting Father's Day and molecules. Father's Day is AC-dependent on us. There is nothing outside of our social conventions that makes claims about Father's Day (e.g. Father's Day is on the third Sunday of June) true or false. Molecules, by contrast, are not dependent on our conventions in this way. There is a positive body of fact about molecules that are independent of anybody's labeling conventions, which in turn can render statements about them true or false. So molecules are AC-independent (Graham 2014, 126). As we will see, questions about the reality of mental disorders usually turn on precisely this issue. That is, when we wonder if a mental disorder is real, we ask if there is anything in the world aside from our labeling practices that answers to the label "mental disorder."

2.2**: Definitional Debate: Historical Background**

In the early 1970's, the main professional body of psychiatrists in the United States, the American Psychiatric Association (APA), commenced preparations for drafting the third and forthcoming edition of its officially recognized nosology, the Diagnostic and Statistical Manual (DSM-III). As part of that effort, the APA convened a special task force in

1973, with the purpose of crafting a definition of disorder for inclusion in the DSM-III. It's certainly not unheard of for working scientists to quibble about how to define the boundaries of their domain. But why the APA decided to officially sanction the enterprise *is* curious. The story of post-war psychiatry in the United States was, by disciplinary considerations anyway, one of unequivocal success. Fueled by NIMH-funded training grants, the APA's membership roles grew from 5,000 to 27,000 between 1948 and 1976 (Hale, 1998, 246).[2] More to the point, neither of the two previous editions of the DSM contained a definition of disorder. What explains the decision to include a definition for DSM-III?

As has been amply documented, the APA's decision was a reaction to a converging series of crises that collectively threatened to undermine its legitimacy *as* a medical discipline (Bayer 1981, Shorter 1997, Cooper 2005, Decker 2013). One source was a change in the nature of social circumstances during the 1960's, whereby old patterns of authority and deference increasingly came under question (Shorter 1997, Grob 2011). Institutional psychiatry was not exempted from this upheaval. In this climate emerged a loose-nit group of intellectuals that has subsequently been called the anti-psychiatry movement. Consisting of philosophers

[2] Grant funding grew from $4.25 to $111 million between 1948 and 1978 (Scully et al. 2000, 124).

(Foucault), sociologists (Conrad, Scheff, Goffman), lawyers guilds (ABA), scientologists (Citizens Commission on Human Rights), and even psychiatrists (Szasz, Laing) the anti-psychiatrists are only a "movement" in the most generous sense (Grob 2011). Nevertheless, one finds many thematic similarities in their various critiques. A reoccurring allegation, expressed in different ways, is that disorder attributions were merely a covert means to control "deviant" behavior.

These wider social forces set the conditions for a very specific internal controversy that eventually prompted the APA to set out on the definitional endeavor. Between 1970 and 1972, the APA and various regional psychiatric organizations faced continual pressure from the gay rights activists to remove homosexuality from the DSM-II. The APA eventually agreed to consider the issue at their national convention in 1973. One party to these debates was Robert Spitzer, who would go on to chair the drafting committee for DSM-III (Bayer 1981). Out of the 1973 discussions, a faction of psychiatrists – mainly those who favored deleting homosexuality from the DSM - recognized the need for a general definition of disorder (Decker 2013). This would, in effect, justify the decision to de-medicalize homosexuality, but also provide a framework for revising the DSM-III. These were the immediate circumstances that led the APA to convene the special task force. In 1978, the APA published

the results of these discussions in a volume titled *Critical Issues in Psychiatric Diagnosis*, edited by Spitzer and colleague Donald Klein (1978).

In the intervening years between 1973 and 1978, philosophers of medicine took note of these public controversies. In 1975 and 1976, Christopher Boorse published two papers, *On the Distinction Between Health and Illness* and *What a Theory of Mental Health Should Be*. The first paper argued that 'health' and 'disease' were well-entrenched theoretical (i.e. descriptive) concepts in physiological medicine that had no larger normative implications (1975). The second paper argued that while psychiatry had not yet assimilated these concepts into practice, there was no in-principle reason why it could not do so once it developed a mature theory of function and dysfunction for psychological phenomena (Boorse 1976). Also in 1975, philosopher Tristram Engelhardt published *The Concepts of Health and Disease* that, among other things, argued that all attributions of disorder were essentially value-laden (1975). These papers clearly had an impact on the APA's deliberations; both Boorse and Engelhardt's works are cited throughout *Critical Issues* (Aucouturier and Demazeux 2014). Interestingly, an essay by Spitzer and Jean Endicott detail how the definition task force moved from the narrow objective of defining mental disorders, to the wider task of deriving a general

definition for *all* medical disorders (Spitzer and Endicott 1978). In effect, this definition would subsume psychiatric *and* somatic conditions under a single concept, thereby fending off the anti-psychiatry movement's critique and justifying psychiatry's place among the medical sciences (Wakefield 2007). After the publication of DSM-III in 1980, psychiatry's political position stabilized, and the issue of defining medical disorders became less pressing (Cooper 2005, 41-42). However, in the philosophy of psychiatry – and more generally, the philosophy of medicine – the search for a suitable definition of disorder, both mental and somatic, persisted. It is this long-running argument that I call the definitional debate.

2.2**: The Definitional Debate: Three Positions**

The definitional debate concerns our *concepts* of health and disease -- i.e. what we mean when we employ these terms. Dominic Murphy provides a useful taxonomy of the main orientations under which most of the positions fall: simple naturalism, simple normativism[3], and hybrid views (2006).

Simple naturalism is the view that all true disorders are real,

[3] What I call "simple normativism" Murphy calls "simple constructionism." My label hews more closely to the terminology within the philosophy of medicine.

phenomena of nature. It holds that the concept of disease is value free and thus, classification of a condition as a disorder is solely an empirical matter. While the term "simple naturalism" might suggest realism about mental illness, a curious fact about this position is that the only people who ever held it were anti-psychiatrists. Take, for example, Thomas Szasz, who was a central figure in the anti-psychiatry movement. Szasz was a realist about somatic conditions, but an irrealist about mental disorders. Many of his arguments appealed to an alleged difference between the sorts of considerations guiding our attributions of physical and mental illness, respectively. The former, Szasz argued, are guided by medical theory, which contains a value-free conception of normal and abnormal functioning. Attributions of mental illness, by contrast, were guided by social norms rather than any theory of normal function or dysfunction. Elselijn Kingma (2013) reconstructs Szasz's reasoning from his influential *The Myth of Mental Illness* (1962):

> P1: What constitutes a disorder is a dysfunction or lesion at the structural, cellular, or molecular level.
> P2: "Mental disorders" present without such a physical lesion.
> C: Mental disorders do not exist. (Kingma 2013, 365)

Szasz's claim is not that the behaviors associated with mental disorders do not exist. Rather, it's that psychiatry doesn't employ a non-evaluative criteria in judging which types of behaviors and thoughts count

as mental disorders. This argument fits comfortably within the AC-independent template for realism (although Szasz doesn't put it in these terms). For Szasz, the only thing sustaining the category of psychiatric disorders is our social conventions.[4]

Hybrid views, or 'two-stage' models of disease (Murphy 2006), are easily the most popular approach among philosophers and medical practitioners; even the DSM at one time presumed a version of it (Cooper 2005). Two-stage models distinguish between a dysfunction and a disorder. A necessary condition for a condition to count as a disorder is the presence of a dysfunctional mechanism – specified either physically or functionally – within the patient. Like simple naturalists, two-stage theorists hold that a dysfunction is an objective, AC-independent theoretical concept, and that determining the presence or absence of a dysfunction is straightforwardly (at least in principle) a task for medical science. Where two-stage theorists depart form simple naturalism is with respect to whether a dysfunction is sufficient for a condition to count as a

[4] One might think that Szasz would be open to the idea that some future theory of psychological function would alleviate this concern. But Szasz seems to have been a kind of dualist about mentation and a structuralist about functions. That is, he seems to have believed that minds were non-physical stuff and that functional analysis could only be given for physical stuff. Whatever the merits of this position, I think it safe to assume that Szasz did not pay very careful attention to philosophical views about the nature of mind during the 1970's, when he was at the peak of his influence.

disorder. Two-stage theorists deny that the mere presence of a dysfunction is sufficient for a condition to count as a disorder. A further normative judgment is required, to the effect that the particular dysfunction is harmful (Wakefield 1992), or a 'bad thing to have' (Cooper 2005). In this way, the two-stage model tries to balance scientific descriptions of biology with normative considerations of harm (Murphy 2006).

There are two main versions of the hybrid view: Christopher Boorse's[5] biostatistical theory (BST), Jerome Wakefield's harmful dysfunction analysis (HDA). It is important to note a difference between these two accounts with respect to what they're trying to explain. For Boorse, the explanatory target is medical theory. The BST purports to give a rational reconstruction (an explication) of how concepts like 'health', 'disease', 'pathology' etc. operate in theoretical medicine. Aside from his

[5] Boorse (2014) explicitly rejects the two-stage label, which was originally proposed by Dominic Murphy. Murphy wrote that Boorse was committed to the "two-stage picture of the foundations of psychiatry," (2006, 19), which holds that our disorder judgments are a hybrid concept, one part factual the other evaluative. Narrowly speaking, Boorse is correct. Boorse holds that at the foundations of medical science – including, presumably, psychiatry – is constituted by his theory of function and dysfunction, which is value free. However, Boorse does distinguish between various "disease-plus" concepts in medicine, which are hybrid concepts consisting of both fact and evaluative components. These concepts guide clinical practice, including judgments about when somebody is ill. It is in *this* regard that I classify Boorse's view as a two-stage theory.

1975 and 1976 papers, Boorse says very little about mental disorders, preferring to keep his focus on somatic conditions. Wakefield, by contrast, is almost exclusively concerned with psychiatric disorders, although he thinks his view applies to both mental and physical conditions. However, unlike Boorse, Wakefield's explanatory target is our folk concept of disorder. He appears to believe that disorder judgments are guided by a universal – perhaps innate – hybrid concept, with one component consisting of harm and the other of dysfunction (Wakefield 2010, 276). Wakefield's account of the function-dysfunction theory looks not to medical theory, but to evolutionary theory. These differences aside, the success of the BST and HDA accounts as *realist* positions regarding the nature of mental disorders depends on whether they can successfully articulate a non-evaluative theory of function and dysfunction.

According to the BST, functions and dysfunctions are defined in terms of their statistically typical-atypical contribution to an organism's inclusive fitness; the ability to survive and reproduce. An individual is in a pathological (i.e. dysfunctional) condition insofar as some internal component departs from the statistically normal range of operation relative to a reference class, such as the group of all individuals belonging to the same age group, sex and race of an individual (Boorse 1977). So, my heart is functional insofar as it operates within a certain range of performance

relative to people who are like me (in the same reference class as me) in the relevant ways. My heart is dysfunctional if it departs from this range, thus impeding my prospects for survival and reproduction.

Jerome Wakefield's HDA is easily the most popular version of the two-stage view, at least among psychiatrists. According to Wakefield, evolutionary theory provides the most scientifically rigorous account of function and dysfunction. On this view, an item's function is identified with its naturally selected *effect*. So the function of, say, an eye, is to discriminate between light and shade, since that is presumably the reason eyes were selected in the first place. A definition of dysfunction straightforwardly follows: for some physical or psychological system **s**, **s** is dysfunctional if and only if **s** fails to perform the function for it was selected (Wakefield 2007).

To illustrate how the two-stage view is supposed to work in actual cases, let us consider Wakefield and sociologist Allan Horwitz's analysis of depression. Wakefield and Horwitz argue that humans possess an innate, universal, psychological module that regulates sadness response. These modules are usually activated in circumstances of loss, such as the death of a loved-one. Depression arises when something goes wrong in the loss-response module. There are, of course, many ways this might happen, but a typical example would be when the module triggers in the absence of

the appropriate environmental input (i.e. when no loss has occurred) (Horwitz and Wakefield 2007, 17).

There is a third version of the two-stage view, which differs from Boorse and Wakefield in terms of how its proponents specify the function-dysfunction distinction. Proponents of these views rely on a "systematic", or mechanistic, account of function (and dysfunction). Mechanistic accounts of functions get their sense from mechanistic functional models they employ, a practice deeply rooted in medicine and biology (Bechtel and Richardson 2010). Mechanistic approaches model biological systems as machines, systems of causally interrelated parts. Modeling a feature of the system involves identifying the relevant sub-parts, their properties and functions, and how this system conspires to produce the phenomena of interest. Mechanistic models make two assumptions about the systems they seek to explain. The first says that the system's behavior is the product of a limited set of subordinate functions (decomposition). The second identifies sub-functions with the activities of the system's component parts (localization). A mechanistic function, then, just is the typical causal contribution a component makes to the overall functioning of the system of which it is a part (Bechtel and Richardson 2010, 23-24). A dysfunction, on the other hand, is the failure of some component of the system to make this contribution (Murphy 2006). Like Boorse and

Wakefield, proponents of this view argue that whether a particular system (a human body, or a human cognitive system) conforms to a given model is a perfectly objective, empirical issue, open to scientific investigation in the usual way. And it is similarly a perfectly objective hypothesis that some system of subsystem is failing to function properly, i.e. is dysfunctional, according to a given functional model. But again, according to this kind of hybrid (two-stage) view, any such *dysfunction* will count as a *disorder* only if it is judged to be harmful.

Simple normativism is more difficult to characterize than either simple naturalism or two-stage views. As Carel and Cooper note, "All [simple normativist] accounts agree that "disease" is a value-laden concept and that diseases are essentially bad, but as authors disagree in their accounts of what is bad, and what other criteria might be essential for a disease, this agreement hides much disagreement," (2014, 4).

At least among sociologists, early proponents of simple normativism tended to accept that physiological medicine is guided by a value-free concept of function and dysfunction, whereas the function-dysfunction concept employed in psychiatry is value-laden. For example, Erving Goffman, an influential figure in the early anti-psychiatry movement, remarks on this difference in his now classic essay, "The Insanity of Place" (1971). Goffman argued that somatic conditions were

medicalized because they involve a departure from biological norms of functioning– norms, he says, that, "…have no moral or social connotation" (1971, 345). Mental disorders, by contrast, involve a departure from social, or culturally established norms. Sociologist Thomas Scheff accepted a similar distinction. But he was even more explicit that Goffman about the *purpose* of mental disorder concepts. As a category, Scheff argued that mental disorders were contrived as a mechanism to control deviant behavior (Scheff 1970).

Fellow sociologist Peter Conrad[6] echoed Scheff on this score in his *The Discovery of Hyperkinesis: Notes on the Medicalization of Deviance* (1975). There, Conrad discusses the "relatively recent phenomenon" of hyperkinesis, a diagnosis increasingly given to children who exhibit patterns of excessive motor activity, short attention spans, and emotional lability. Conrad remarked how the medicalization of hyperkinesis provides several social benefits to all parties involved, while also strongly implying that these benefits are the *only* reason for medicalizing hyperkinesis:

> Hyperkinesis minimizes parents' guilt by emphasizing "its not their fault, its an organic problem" and allows for nonpunitive (sic)

[6] In *The Discovery of Hyperkinesis*, Conrad did not say whether he thought a non-value involving analysis of function-dysfunction for physical disorders could be had. In a follow-up work co-authored with Joseph Schneider, *Deviance and Medicalization: From Badness to Sickness*, he denied that any such distinction for either physical or mental disorders could be formulated (1980/1992).

> management or control of deviance. Medication often makes a child less disruptive in the classroom and sometimes aids a child in learning. Children often like their "magic pills" which make their behavior more socially acceptable and they probably benefit from a reduced stigma also.
>
> [Conrad 1975, 17]

Peter Sedgwick, an intellectual and activist on the British left argued that *all* disorder attributions are value-laden. In a paper criticizing Goffman's distinction between biological and social norms, Sedgwick charged him with implicitly giving somatic medicine an undeserved pass from critique[7]:

> [The] 'biological norms' to which physical medicine appeals and the 'social norms' which back up psychiatry, dissipates into nonsense as soon as we are brought to see that the medical enterprise is from its inception value-loaded; it is not simply an applied biology, but a biology applied in accordance with the dictates of social interest.
>
> [Sedgwick 1974, 28]

He goes on to argue that disorder attributions follow from a gap between the observable state of the person and a social expectation of well-being, which likely varies from culture to culture. According to Sedgwick, what individuates disorders from other disvalued states (e.g. criminality) is that explanation of the state is limited to, "…a relatively restricted set of causal factors operating within the boundaries of the

[7] Goffman was not the only anti-psychiatrist Sedgwick criticized. Szasz was also a frequent target of Sedgwick's ire. In a humorous passage from his polemic *Psycho Politics*, Sedgwick (a lifelong Marxist) dismisses Szasz's (a fervent libertarian) critique as "only seldom militant" (Sedgwick 2012).

individual human being," (1972, 216). Rachel Cooper, who is no anti-psychiatrist, also argues for normativism about disorders. She defines a disorder in terms of three individually necessary and jointly sufficient conditions: a condition **x** is a disorder if and only if **x** is a bad thing to have, those afflicted with **x** are unlucky, and **x** is potentially medically treatable (Cooper 2005, 4).

These three orientations – simple naturalism, hybrid views, and simple normativism – capture most of the positions in the definitional debate. In what follows, I consider some of the main elements of the debate between these three camps. I'll begin with a word about the over-all structure of the debate.

2.3**: The Definitional Debate as Conceptual Analysis**

The definitional debate is straightforwardly a project of conceptual analysis. Robert Audi provides a good working definition of this activity: "Let us simply construe [conceptual analysis] as an attempt to provide an illuminating set of necessary and sufficient conditions for the (correct) application of a concept," (1983, 90). Proposed definitions thereby assume the form of '**x** is a disorder if and only if…' where conditions for **x**'s application are placed after the ellipses. It is worth noting that in these debates, the application conditions are almost always relatively simple, consisting of no more than two or three conjuncts. Recall Wakefield's

HDA: **x** is a disorder if and only if **x** is a result of an evolutionary dysfunction and **x** is judged harmful. Rachel Cooper's purely normative account exemplifies a similar structure: **x** is a disorder if and only if **x** is a bad thing to have, **x** is such that the afflicted person is unlucky, and **x** is potentially medically treatable (Cooper 2005, 22). Objections to candidate definitions are adduced by devising a counterexample, undermining either their necessity or sufficiency.

For instance, necessity objections to simple naturalism and the two-stage view point to a number of widely accepted medical conditions (e.g. hypertension, appendicitis) that do not straightforwardly result from a dysfunctional mechanism (Murphy and Woolfolk 2000, Cooper 2005). Simple normativism suffers from worries about sufficiency, since not all disvalued states – states we think of as 'bad things to have'- are considered medical disorders (e.g. poverty, ugliness, etc.) (Cooper 2005). Sedgwick seemed to have been aware of this problem, and proposed to get around it by stipulating that a condition is a disorder if and only if we disvalue it *and* its explanation (in whole or in part) is located within the individual. But as Wakefield notes, while this rules out some problematic cases (e.g. poverty), it doesn't rule out all of them; for instance, teething and illiteracy are disvalued *internal* states and yet not considered medical disorders (1992).

The exclusive reliance on conceptual analysis has prompted many to doubt whether the definitional debate is a useful project to pursue. Maël Lemoine, for instance, argues that the definitional debate follows an internal sort of logic that makes it incapable of settling its core objective – deciding whether our concept of disorder is descriptive or value-laden (Lemoine 2013). Other criticisms point to the implicit fact-value distinction underlying the definitional debate. These critics question whether these two things are neatly separable in actual practice (Douglas 2009, Aucouturier and Demazeux 2014), and note that what is a value-laden classification at one time (e.g. gold is precious metal) may eventually result in a non-evaluative classification at a later time (e.g. gold is the chemical element AU) (Lemoine 2013). That is, even if values play a role in in classifying a condition as a disorder, there maybe further facts to discover about it. Thus, whether our initial categorization judgment was determined (in whole or part) by some normative criteria, that alone is not sufficient to disqualify the category as real. I am sympathetic to these critiques, but for our purposes, the most relevant feature of the definitional debate is the sub-argument surrounding the notions of function and dysfunction, which motivate two-stage views of mental illness.

2.4**: Function and Dysfunction**

Although defined differently, both hybrid accounts of disorder

crucially depend on deriving a non-evaluative definition of function. Both Boorse and Wakefield's distinct versions of function have been subject to numerous epicycles of counter-example, reply, and renewed counter-example. Dominic Murphy and Robert Woolfolk, for example, criticize Wakefield's reliance on evolutionary function on the grounds that it would rule out conditions such as appendicitis (2000). Contra Boorse, Germund Hesslow has argued that 'health' and 'disease' concepts play no significant *theoretical* role in medical theory (1993). An important line of critique seeks to show that Boorse's notion of function and dysfunction *are not really non-evaluative* after all. On Boorse's view, function and dysfunction are relative to a reference class. Many have argued that norms are ineliminably involved in selecting the appropriate features of the reference class, so Boorse's view actually smuggles in values through the back door (Kingma 2007 and 2014).

If our interest is in arguing that mental disorders are *real* and we choose to pursue that line of argument through something like the two-stage view, or even just within the general terms of the definitional debate, then – as a practical matter – the question of AC-independence reduces to finding a non-evaluative account of the function-dysfunction distinction. But that is a much stronger commitment than is required by realism, understood in terms of AC-independence. All AC-independent realism

requires is that the facts about mental disorders don't ultimately bottom out on mere labeling conventions. Perhaps a non-evaluative notion of function-dysfunction is to be had, and perhaps many disorders (physical/mental) involve a dysfunction. Or perhaps there isn't any such notion; or maybe there is, and some conditions involve dysfunctions while others don't, and we can only decide which to include in our medical categories on a case-by-case basis. However this issue shakes out, tying mental disorder realism to the project of objectifying the function-dysfunction distinction is gratuitous. It is not required for either establishing or defending realism about mental disorders. If these considerations are plausible, then that frees us up to pursue an alternative approach.

2.5**: Conclusion**

In this chapter, we articulated the sense in which mental disorders may be real in terms of Graham's notion of AC-independence (act-of-classification-independence). We examined the most common way in which philosophers have pursued realism about mental disorders: the definitional debate about disease. I demonstrated how the project of looking for definitions has been intimately tied to the social and political circumstances institutional psychiatry found itself in during the 1970s. We observed how, even after this context faded away, the debate about mental

disorders proceeded largely in terms set by those debates. After showing how the definitional debate typically proceeds, we considered a few critiques of that enterprise. I concluded by arguing that the mental disorder realist need not be constrained by the definitional debate, and suggested that questions about realism can legitimately proceed in another way.

Chapter 3: Two Accounts of Natural Kinds

3.0: Introduction

In chapter 2 I defined realism in terms mind-independence, and mind-independence in terms of AC-independence (act-of-classification-independence): **x** is AC-independent if and only if there is some positive body of fact about **x** that holds independently of our classification practices. Thus, when **x** is AC-independent, the reality of **x** follows from there being a positive body of fact about **x** that exists independently of our labeling conventions. I noted that the traditional vehicle for defending mental illness realism is the definitional debate. In that context, realists attempt to establish the AC-independence of mental illnesses by articulating a non-evaluative theory of the function-dysfunction distinction. Whether or not such a theory can be had, I argued that nothing about AC-independence *per se* required such a theory. If that argument is sound, then we are free to pursue alternative ways of defending realism about mental disorders. Here, I will suggest that mental disorders are, or might be, understood as natural kinds -- that is, as categories that "exist in nature", independently of our ideas about them. Thus, natural kinds are AC-independent categories. Arguably, natural kinds and their causal properties are the primary subjects of study within scientific disciplines. Fundamental physics studies such kinds as electrons, quarks, and photons.

Chemical kinds include the individual chemical elements (e.g. hydrogen, gold, uranium) and molecules (e.g. H_2O, benzene, etc.). Biological kinds may include specific kinds of cells (e.g. erythrocyte, osteoclast, and various specific types of neurons), specific kinds of organs (e.g. hearts, livers, lungs), specific kinds of organisms (i.e. species), as well as specific cognitive systems and their subsystems (i.e. attention systems, memory systems, various sensory systems, etc.). I will argue that mental disorders are best understood in terms of natural kinds within cognitive systems.

In this chapter I survey two ways in which philosophers have characterized natural kinds. The first, long associated with traditional metaphysics, is essentialism. I interpret contemporary essentialism as a commitment to the following four theses:

1. All natural kinds have essences (i.e. essential properties).
2. Essences are necessary and sufficient for kind membership.
3. A kind's essential properties are microstructural properties.
4. A kind's essential properties are intrinsic (never relational).

I then critique essentialism, arguing that it ought to be rejected since it commits us to an ontology out of step with most of the natural sciences.

The second approach to natural kinds I will explicate is Richard Boyd's homeostatic property cluster (HPC) theory natural kinds. While I

do not know if Boyd's view is successful as a general theory of natural kinds, it is far superior to essentialism in terms of its fidelity to the scientific understanding of *special science* kinds. After briefly outlining the view and contrasting it with essentialism, I then provide a more precise what an HPC-kind is, relying on the philosophy of neuroscience literature to aid in my explication. I argue ADHD qua HPC-kind should be understood in terms of a cognitive neurobiological mechanism. After specifying what that understanding entails, I then consider how cognitive neurobiological mechanisms are explained, investigated, and modeled. My purpose here is two fold. In the next chapter, I will show that Russell Barkley's EF-Model of ADHD is a *mechanism* sketch. So I want to clarify what that means in order to provide a framework for interpreting Barkley's EF-Model, as well as how we should interpret the evidence for it.

3.1**: The Essentialist Account**

The term "natural kind" first entered the philosophical lexicon in the 19th century (Venn 1866, 246), but the basic idea reaches back to antiquity. Within the western tradition essentialism is the predominant way in which philosophers have thought about natural kinds. The core of the view holds that each individual possesses a set of properties that make it the kind of object that it is, its *essential properties*. Aristotle's

metaphysical works (1963, 2016)[8] constitute the foundation for the approach, and for the ensuing millennium and a half, essentialism remained tethered to this basic framework. Aristotelianism conceives of the world as one populated by composite entities called substances. An individual substance – a rock, a dog, a human, etc. – is composed of matter and form, the former individuating it numerically from others of the same kind, while the latter determines its essential properties.[9] A substance's essential properties, in turn, determine the kind to which it belongs. So, the first essentialist thesis is simply the attribution of an essence, a set of *kind-determining* properties possessed by the members of a kind. While the form/matter ontology out of which essentialism is no longer a widely held position, the idea that membership to a kind depends on an individual possessing a set of essential properties remains. The second thesis says that a kind's essential properties are necessary and sufficient for kind membership. To put the point somewhat technically,

[8] See especially VII pt. 4 and XII pt. 3 in *Metaphysics* (2016).

[9] I am glossing over a particularly knotty interpretive dispute concerning how Aristotle and the subsequent Aristotelian tradition conceived of the relationship between essences and forms. Loux (1998) argues the relationship is one of identity: a substance's form and its essential properties are one and the same thing. Robert Pasnau (2011), on the other hand, construes the relationship merelogically: a substance's essence is a composite consisting of form and "common matter", which refers to the general physical features a substance shares with other members of the same kind.

this condition means that for each kind **K** there is an associated set of essential properties {**x**, **y**, **z**} such that any individual **i** belongs to **K** if and only if **i** instantiates each of {**x**, **y**, **z**}. As it stands, these first two essentialist theses tell us virtually nothing about what a kind's essential properties are like. We can think of the last two essentialist theses as filling that gap.

The third thesis identifies a kind's essential properties with its microphysical, or microstructural, properties.[10] The use of "micro" in this context is, perhaps, misleading since not all instances of this claim refer to the microphysical level of reality as understood by physics, although some do. For instance, atomic number (the number of protons in each atom) is commonly designated as the essence for chemical element kinds; while for biological kinds such as a particular species, essentialists favor identifying essences with a species-specific genotype. The salient feature of this doctrine isn't the identification of essences with microstructures (at whatever physical "level" they reside), but the particular causal-explanatory function that essential properties are supposed to play.

[10] Microphysicalism is not, per se, an essentialist doctrine. For example, essentialist David Oderberg argues that "real essentialism" does not privilege a kind's microscopic properties over its macroscopic properties (Oderberg 2007, 15-16). Oderberg's position, however, is a minority view among essentialists. I will thus retain the traditional connection between microphysicalism and essentialism.

Microstructural essences are claimed to causally produce a kind's stereotype (Putnam 1975, 169), the observable (i.e. non-essential), macro-level properties we naïvely associate with members of the kind before we know anything about its essence. To cite a standard example, if water is a natural kind, then H_2O is plausibly the essence of that substance, and it is the micro-physical properties (such as charges) together with the physical molecular arrangement of hydrogen and oxygen atoms that causally produces water's macro-level properties, such as transparency and liquidity.

This causal-explanatory take on essentialism became fashionable during the 20th century due to Putnam and Kripke's work on natural kind semantics. Neither offered anything like a systematic statement of essentialism. Rather, they explicated it with a suite of examples involving particular natural kind terms: "lemons" (Putnam 1975, 158-159), "tigers" (Kripke 1980, 120-121), "gold" (1980, 123-125) and "water" (1980, 128). In a short 1974 paper, J.L. Mackie argued that Locke anticipated the Kripke-Putnam approach to kinds (Mackie 1974). And so, it is now commonplace in the literature to credit Locke as the progenitor of microphysicalism (Dupré 2001).[11]

[11] Mackie and others are right to credit Locke, but it seems to me they credit him for the wrong accomplishment. Locke was certainly not

The last of the four essentialist theses says that a kind's essence is solely constituted by intrinsic, rather than extrinsic, properties. David Lewis provides a frequently cited example of an intrinsic property, "A thing has its intrinsic properties in virtue of the way that thing itself, and nothing else, is." (1983, 111-112) Beyond this, the notion remains the subject of a long-standing debate among philosophers. I cannot hope to provide a comprehensive survey of the various positions here, so an example will have to suffice. Brian Ellis defines an intrinsic property as one an object possesses solely in virtue of *internal* causal forces, "Essential [intrinsic] properties distinctive of a natural kind must be independent of the histories, locations, and surroundings of its members," (Ellis 2002, 51).

---

the first philosopher to articulate the basic logic of the assumption – that of a dependency relationship between a kind's essential and non-essential properties. Rather, Aristotle already did so; and that basic connection persisted relatively intact throughout the long scholastic tradition that followed in his wake. By the time Locke published the first English edition of the *Essay* in 1689, this causal-explanatory conception of essences was, as Robert Pasnau puts it, "…the [absolute] standard seventeenth-century notion of what an essence is," (2011, 658). So, in terms of the function essences perform – as causally underwriting the characteristic, yet non-necessary properties of a kind – Lockean essences were not terribly different from traditional essences.

Locke's innovation is the way in which he gave a material interpretation to the Aristotelian account of the essential-non-essential properties distinction, resituating it within the new corpuscular-mechanistic science. Commandeering a bit of Aristotelian terminology, Locke even refers to these microstructural properties – which include figure, size, shape – as "real essences" (Locke 1996, 185-187).

3.2**: Critique: Essentialism**

Most contemporary philosophers do not consider essentialism an acceptable theory of kind-hood. Due to its rather stringent metaphysical commitments, essentialism leads to an ontology bereft of many well-motivated scientific categories simply because they do not conform to the (largely a priori) strictures of the view. This merits jettisoning the essentialism as a viable account of natural kinds.

Let us consider some example from the special sciences. With respect to biological kinds, essentialism results in metaphysical extravagance. As Magnus (2012, 34) notes, in developing their own versions of this view Wilkerson (1988, 1993, and 1995) and Ellis (2001 and 2002) independently arrive at the conclusion that individual organisms must constitute *distinct* natural kinds. This is because Wilkerson and Ellis argue that the essence of a species is defined by the total genotype each individual organism has, which is ultimately responsible for it having the biological properties, dispositions, and behaviors that it has. But since there just so happens to be a lot of interspecific genetic variation - that is, no two conspecifics (not even monozygotic twins) are genetically identical – the result is that each individual genotype constitutes a distinct kind unto itself. So, whereas a canine biologist – or, you or me, or any other right-thinking person – would classify my pets, Watson and Charlie, as

members the *single* kind canis lupus familiaris, Wilkerson and Ellis's accounts classify them as members of two *distinct* kinds, defined by the ineliminably idiosyncratic composition of their respective genomes! This bizarre outcome follows straightforwardly from the assumption that essences correspond to necessary and sufficient properties and microstructural assumption.[12] As Wilkerson writes, "The real essence which is in part causally responsible for the behavior of the genetically unique individual is the essence of the *kind* of which the unique individual is the only member" (Wilkerson 1993, 16; emphasis original).

The above example suffices to disqualify essentialism as a serious doctrine regarding kinds (as appropriate to the special sciences, at least). But further problems beset the view. Take the assumption that essences must be intrinsic. Ellis stipulates that, "Essential properties distinctive of a natural kind must be independent of the histories, locations, and surroundings of its members," (2002, 51). So stated, it is not even clear whether this definition can accommodate Ellis' own flagship example of natural kinds: chemical elements. For, in keeping with the microstructural

[12] From vantage point of evolutionary theory, this result is wholly unsurprising. Essentialism predicts, at least, a set of traits universally present in the species and unique to each species. Forces like mutation and genetic drift militate against universality, while common ancestry and parallel evolution militate against uniqueness (see: Ereshefsky 2004, 95-102).

assumption, he also wants to say that a chemical element's macroscopic properties are causally produced by its intrinsic essence. But as Holly Vande Wall argues, it is simply not the case that the macroscopic properties of chemical substances flow neat and tidily from its intrinsic, microstructural properties; to a significant extent, an element's macroscopic properties depend on extrinsic properties as well. Concerning the well-worn essentialist example, H_2O, Vande Wall writes, "In order to behave like water, the H_2O molecules and ions must be in contact…here it seems that *where* the molecule is found (in contact with other molecules and ions of oxygen and hydrogen and not, for example, isolated in the interstices of a crystal structure) is an *extrinsic* quality," (Vande Wall 2007, 51; emphasis original).[13]

Although Ellis's analysis of intrinsic properties is but one of many on offer, I am doubtful that any of them could ultimately save the intrinsicality assumption about kinds. The basic obstacle is this: there are just too many clear-cut examples of kinds defined by properties that *no* account of intrinsicality could plausibly capture. Returning to our previous example, part of the reason that canine biologists classify my dogs Watson

[13] Another problem for Ellis' view is that chemists do not consider molecular shape an intrinsic property, but a response property that varies according to the context the substance is in (Ramsey 1997).

and Charlie as conspecifics is that they share an ancestor that diverged from wolves about 35,000 years ago (Skoglund, et al. 2015). Presumably, no reasonable analysis can sanction shared ancestry as an intrinsic property.

3.3: Homeostatic Property Clusters

Richard Boyd first introduced the HPC-theory as part of a defense of moral realism (1988). He argued that many traditional philosophical and scientific categories – reference, goodness; species – could be fruitfully understood as HPCs. Over time, (1989, 1991, 1999, 2000, 2003, 2010) he developed the HPC-theory into a full-blown, general view of kind-hood.

Boyd identifies natural kinds with HPCs. An HPC is a cluster of contingently associated properties that tend to be reliably co-instantiated by its members, or instances, due to a causal mechanism. The qualifier "tend" is important here, as it singles out an important difference between HPCs and the essentialist approach to kinds. Essentialism holds that for each kind, there is a set of properties that all and only members of the kind possess. The HPC-theory makes no such demand. The clustering of properties associated with the kind is often "imperfect" in that no single property need be necessary for kind membership (Boyd 1989, 15). As Magnus quips, on the HPC account, "A one-legged mallard will not walk

like a duck, and a mute mallard will not quack – but both are nonetheless mallards," (Magnus 2014, 149-150).

As other philosophers have noted (e.g. Griffiths 1999), the HPC-mechanisms seem to perform a role similar to essences (i.e. the essential properties involved in essentialist accounts) in that they both are supposed to causally produce the stereotypical properties and behaviors associated with the kind. Samuels goes so far as to call the HPC-mechanisms "causal essences" (Samuels 2009, 56). There are similarities between the essentialist and HPC accounts, but they shouldn't be overstated. First, as already mentioned, on the HPC account no property need be strictly necessary for membership in a kind. Second, unlike essentialism, the HPC-theory imposes no a priori constraints on the types of properties (processes, etc.) that may compose the homeostatic mechanism; nor does it posit a privileged physical level where such properties have to reside (Boyd 1999). The upshot is that a property (or process etc.) earns inclusion within the HPC-mechanism if it causally contributes to the stability of the cluster – that is, if it makes a causal contribution to the co-instantiation of the property cluster associated with the kind.

This looseness should not be construed as a total lack of constraints. Not just any cluster of properties counts as a kind. Some philosophers (Wilson et al. 2007, 198) have interpreted the HPC-theory as

sanctioning disjunctive kinds, but that is a mistake. Boyd argues that one reason natural kinds are epistemically valuable is due to their role in facilitating scientific investigation and inductive inference (Boyd 1999, 146). The HPC-theory *explains* why natural kinds can play this role in our epistemic practices: the collection of properties associated with a given kind reliably cluster as they do *for a reason*. In the case of a species – at least, those that are sexually dimorphic – the primary stability generating mechanism of a kind is its spatio-temporally continuous reproductive history.[14] If a kind were truly disjunctive in having *no* properties (e.g. no morphological, behavioral, historical, etc. properties) in common, the corresponding kind-category could function neither as a stable target of investigation nor feature into explanations in any reasonable way.

Boyd proposes that *all* natural kinds are HPCs, but other philosophers are skeptical. One issue concerns the fact that many fundamental physical kinds do not seem to have their properties in virtue of a causal mechanism (Magnus 2014, Slater 2015). Whether that objection is fatal to the HPC-theory as a general account of kinds, or merely points out a special limiting case of the notion of kinds, need not concern us here. The importance of the HPC-theory for our purposes is

[14] The full story is probably more complex than this simple sketch. There are likely other mechanisms (developmental, environmental) at work in maintaining the stability of a species overtime.

that it does seem to accurately capture a large number of special science kinds. In particular, as other philosophers have noted (e.g. Samuels 2009, Beebee and Sabbarton-Levy 2010, and Tsou 2013) the HPC-account rather nicely accommodates psychological and psychiatric kinds.

3.4: Psychiatric Kinds as Cognitive Neurobiological Kinds

Boyd provides a lot of interpretive leeway in how we might understand how an HPC-kind is "put together" and which properties count in establishing the identity of a given HPC. One interpretation holds that an HPC is simply a self-sustaining property cluster, whereby causal relations between the properties implement the "mechanism" itself. Another separates the mechanism and property cluster such that the property cluster is sustained distinct causal process. Boyd seems to understand species in this way; that is, a species is a collection of individuals bearing various physiological, morphological, and behavioral similarities to one another and individuals tend to instantiate all (or most) of these properties due to causal processes such as the reproductive history of the species, gene flow, and so on.

For mental disorders, I propose to interpret them as HPC-kinds and explicate the sense in which they *are* HPC-kinds by resorting to the

philosophy of neuroscience literature on mechanisms.[15] In sum, I assert that ADHD is an HPC-kind if members of the kind instantiate the same *cognitive neurobiological* mechanism. To be clear, Boydian "mechanisms" and the sense of mechanisms articulated below have distinct senses, and ought not be confused. Boydian mechanisms refer to the causal processes that keep an HPC in homeostasis. Depending on the kind at issue, they may or may not be part of *defining* an HPC-kind. The sense of mechanism I will employ is narrower. Roughly, a mechanism in my sense is an organized physical structure instantiated in the nervous system and is responsible for producing the set of symptoms associated with a given mental disorder. In this restricted sense, the mechanism establishes the identity of a mental disorder, defined as an HPC-kind. Which is to say, I intend my particular interpretation of mechanism to individuate the HPC-kind.

Before explicating this restricted sense of *cognitive neurobiological* mechanisms, allow me to head off a potential objection to this approach. In chapter 2, we discussed mechanistic explanations as a way to cash out the dysfunction-function distinction. One may, therefore,

[15] This construal of HPCs was first proposed by psychiatrist Kenneth Kendler, psychologist-philosopher Peter Zachar, and the philosopher of neuroscience Carl Craver (2011). The account I offer here is similar to Kendler et al., but more restricted.

justifiably suspect that the proposal to understand psychiatric kinds in mechanistic terms just returns us to the argumentative ground of the definitional debate. That is, we use mechanistic explanations to ground a dysfunction-function distinction, and then argue that mental disorders are real since they are sustained by *mechanistic* dysfunctions. Not so! For, the definitional debate takes the function-dysfunction distinction as primary, and attempts to define that distinction in a way that is independent of any specific disease (or any specific functional mechanism). We are not engaging in that project here. Rather, my goal is to address the issue of realism for specific mental disorders; namely, ADHD. While the account I consider, Barkley's EF-model, *does* suppose that ADHD involves a dysfunctional mechanism, we can remain agnostic about the mechanism is "really" dysfunctional, or whether it is *merely* different. Establishing a *difference* is sufficient for establishing AC-independence.[16]

Cognitive neuroscience seeks to explain cognition and behavior in terms of the organization of the nervous system. For complex cognition – e.g. memory, attention, and perception – it assumes, by and large, the theoretical framework of cognitive psychology, which views the mind as a

[16] Another difference between this project and the definitional debate is one of scope. Unlike realists in the latter, I have no additional commitment to the notion that *every* mental disorder *must* involve a dysfunction of a mechanism.

set of information-processing capacities operating over representations (Crastley and Samuels 2013, Sullivan 2015). The primary difference between these two disciplines is that cognitive neuroscience marries functional analysis to structural analysis (Boone and Piccinini 2015). So, to a first approximation, cognitive neuroscience aims to elucidate the physical mechanisms underlying the cognitive capacities recognized by cognitive psychology.

As understood here, a mechanism is a hierarchically decomposable physical system of causally interacting entities and activities. An entity is a component (i.e. part) of the mechanism; an activity is what the entity does. A mechanism implements a capacity in virtue of the causally organized relations of its components and their activities (Craver 2007, Craver and Darden 2013). As Boone and Piccinini note, neurocognitive mechanisms exemplify an iterative structure, wherein, "…each component of the mechanism is in turn another mechanism whose capacities are explained by the organized capacities of *its* components; and each whole mechanism is itself a component part that contributes to the capacities of the larger whole," (2016, 1515; emphasis original). In other words, what counts as a single mechanism at one level of organization is, at a more detailed level of organization a hierarchical system of nested mechanisms.

A word about "levels" is warranted here, since that term is invoked

across a number of argumentative contexts and admits different interpretations. In mechanistic contexts, levels are understood in terms of two notions: composition and organization (Craver 2007, Craver and Bechtel 2007). The former expresses a structural fact about mechanisms. Namely, mechanisms are composed of part-whole relations between the "higher-level" mechanism implementing a capacity and its "lower-level" component parts and their activities. We can understand the meaning of "organization" by contrasting mechanisms with aggregates (Wimsatt 1997, Craver 2007). An aggregate is an unorganized collection of components whose properties are literally the sum of its component parts (e.g. the mass of a heap of sand just is the sum of the mass of its grains) (Craver 2009, 395). Mechanisms are more than aggregates because, in the first instance, their components must stand in specific causal relations to one another in order to implement a capacity. For example, a mere heap of cardiomyocytes cannot push blood through the circulatory system; rather, they must be suitably organized (e.g. causally, temporally, spatially, etc.) in order to do so. In addition, mechanisms – unlike aggregates – possess causal powers their components do not. This should not be read as a commitment to emergent properties, at least in any significant metaphysical sense. It merely expresses an intuition, widely embedded in common sense and scientific practice, that organized physical systems can

do things none of their parts can do on their own (Craver 2007, 217) – e.g. hearts are cardiomyocytes organized in a way that can pump blood, whereas no cardiomyocyte can itself pump blood. Kim characterizes this feature of macro-micro wholes as "macrocausation" and, similarly, argues that it is a metaphysically unproblematic notion:

> This baseball has causal powers that none of its proper parts, in particular none of its constituent microparticles have, and in virtue of its mass and hardness, the baseball can break a window when it strikes it with a certain velocity. The shattering of the class was caused by the baseball and certainly not by the individual particles composing it.
>
> [Kim 2005, 56]

A mechanism will exhibit different degrees of organizational complexity depending on the phenomena in exhibits. It is largely assumed that the mechanisms underwriting thought and behavior will be quite complex; involving multiple levels of organization– i.e. molecular, cellular, circuit, network levels. Nothing in this interpretation requires complex cognitive mechanisms to be *strictly* localized in the sense of being confined to a relatively well-circumscribed area of neural tissue. As physical structures, mechanisms have to be localized *somewhere*, but "somewhere" implies continuum of localizability (Bechtel and Richardson 2010). For example, the components of the mechanism underwriting long-term potentiation across the synapse (a persistent increase in synaptic strength) are confined to the components of individual cells, thereby

occupying the more strongly localizable end of the continuum; the components of neural networks are, by contrast, much more weakly localizable and thus, reside at the opposite end of the continuum (Buckner 2015, 3936).

We can now state how ADHD understood as an HPC-kind: ADHD is an HPC-kind if its symptoms result from, or are sustained by, a similar organizational structure, specified in terms of a hierarchically organized cognitive neurobiological mechanism. What remains to be determined is how we would go about finding out if individuals with ADHD met this condition. For that, I know turn to provide a brief account of the nature of mechanistic models and their role in the explanation and investigation of the workings of cognitive systems.

3.5**: Mechanistic Explanations, Models, and Investigation**

Complex cognitive capacities are exhibited in hierarchically arranged physical mechanisms. A mechanism exhibits a capacity in virtue of the organized collection of its component parts and their activities. A mechanistic explanation can be cast in many different representational formats, but for the sake of simplicity let us subsume these under the broad heading of "mechanistic models." Mechanistic models lie along a continuum in terms of completeness, ranging from sketches, to schemas, to fully adequate structural-functional models (Craver 2007). Mechanism

sketches are incomplete specifications of the explanandum mechanism (the mechanism to be explained), usually relying on gaps and filler terms for entities and activities that are not yet known (Weiskopf 2011, Craver and Darden 2013). Mechanism schemas are abstract specifications of mechanisms that can be filled in to yield specific causal processes (Craver 2007, Craver and Darden 2013). Most models of complex cognitive phenomena lie somewhere between sketches and schemas. Nonetheless, they are more useful than mere phenomenal models. For, whereas phenomenal models merely describe the mechanism's observable behavior, it includes no details of its underlying means of operation, whereas mechanism sketches provide *constraints* on the space of possible mechanisms that might be responsible for the phenomena of interest (Craver 2007, 228).

As Craver notes, scientists rarely consider the full space of possible mechanisms the outset of any attempt to model a cognitive mechanisms. Rather, it is usually constrained by what the researcher already knows about the phenomena at issue (Craver 2007, 247). Although not every mechanism implemented by the nervous system spans several levels, complex cognitive capacities do, and so pruning this space requires information about causal relations between components operation at several distinct organizational levels. Cognitive neuroscientists have a

number of experimental protocols for determining whether a putative component[17] causally contributes to the operation of the mechanism. (Craver 2007, 198)[18] I'll mention two, but both comprise variations on a single theme: intervene on the variable of interest, and track the (resulting) changes elsewhere in the system. This can proceed in either a bottom up or top-down fashion. Lesion studies and gene knock-out experiments are bottom-up interlevel experiments, where some lower-level component is manipulated to see whether that produces a change in the behavior of the whole. Top-down experiments look at the behavior of the components in order to determine if they are active during the mechanism's operation. Examples include task-based functional magnetic resonance imaging (fMRI) studies: a subject performs a task believed to tap the relevant cognitive capacity while changes in blood-flow or oxygenation are tracked (via an fMRI machine). The same experimental design can be

[17] The use of the term "component" is intended widely, to include both "small" items (e.g. cells) and "big" items (e.g. circuits) as well as their activities.

[18] Stated formally: for (a) system *S*, (b) capacity *y-ing* exhibited by *S*, (c) putative component *x* and sub-capacity *f*-ing, causal relevance can be shown if and only if (i) *x* is contained within *S*, (ii) some ideal interventions on the *x*'s *f*-ing changes the phenomenon *S*'s *y*-ing, and (iii) some ideal interventions on *S*'s *y*-ing changes *x*'s *f*-ing. This is sufficient to establish that *x* is a component in the causal mechanism. (Craver and Tabery 2017).

implemented with different measuring apparatuses, including single and multi-cell recordings (Craver and Darden 2013, 125-128).

Mechanistic investigations proceed in an iterative fashion across multiple levels of organization, where each field uses its own specialized experimental techniques and models. The picture is one of *relative*, but not total autonomy among fields. The overall goal is to identify mechanisms for a particular capacity by shrinking the space of possible mechanisms. At minimum, this involves determining roughly *where* the mechanism's boundaries lie, and what kinds of components causally contribute to its operation. The best, and perhaps, only way to accomplish that is through the coordination of models across different organizational levels (Craver 2007).

To sum up, complex cognitive capacities are implemented by hierarchically arranged physical structures. Mechanist models explain these capacities to the extent that they can articulate how the organized activity of the structure's components produce the phenomena of interest. Models vary according to the level of detail they include. But most models of complex cognitive capacities lie somewhere between sketches and schemas. Nonetheless, whatever their degree of completeness, each mechanistic model of complex cognitive capacities is designed to constrain the space of possible mechanisms for the phenomena of interest.

Moreover, all such models are evaluable by the same set of evidential criteria, namely, the degree to which they accounts for observable behaviors and physical processes.

Although only a partial guide to the functioning of a mechanism, mechanism sketches often contain enough organizational structure to endorse realism with regard to a particular mental disorder. The mental disorder realist faces no special problem here. Most models of complex cognitive mechanisms are likely to be mere sketches of a complex underlying physical process. Unless one is a dogmatic irrealist about the mind, the evidential support for mechanism sketches should provide good reason to believe that the systems and components they model are real features of human cognitive systems.

3.6**: Conclusion**

I have argued for the rejection of essentialism about natural kinds. My arguments against both appealed to scientific practice. I argued that essentialism commits us to an ontology out of step with the scientific worldview and is thus, not a suitable philosophical view of natural kinds. I presented an alternative account, Boyd's HPC-theory, and argued that its flexibility made it a superior view to essentialism. I then provided a more precise interpretation of HPC-kinds, relying on the philosophy of neuroscience literature about mechanisms. I argued a mental disorder is an

HPC-kind if set of symptoms associated with the disorder is causally sustained by a similar organizational structure across instances, specified in terms of a hierarchically organized cognitive neurobiological mechanism. I then specified how cognitive neurobiologists constructed their models, distinguished between mechanism sketches and schema, and further showed how models are tested in order to make sense of my discussion of Barkley's EF-model in the next chapter.

Chapter 4: Realism and ADHD

4.1: Introduction

In this chapter I address the primary question of this dissertation: is ADHD an HPC-kind and thus, by the dictates of AC-independence, is it *real*? I will pursue that question through Russell Barkley's Executive Function (EF) model of ADHD. If Barkley's model is more or less right – that is, if (some or all) of individuals diagnosed with ADHD instantiate the mechanism specified by the EF model - then ADHD does indeed count as a *real*. In section 4.3, I provide brief survey of the EF-model, its components, and the way in which accounts for ADHD behavior, I will show that Barkley clearly intends that the EF-model to be a mechanistic model of ADHD. So, we may evaluate it as we would any mechanistic model. I will then assess a few distinct lines of evidence in its favor. I argue that while the mechanism described by the EF-model (or some very similar model) is most likely instantiated in a subset of individuals with ADHD, not all individuals with ADHD have an EF deficit. There is compelling evidence to suggest the presence of an alternative mechanism that produces a class of the symptoms associated with ADHD. I will survey one of them, Edmund Sonuga-Barke's Delay Aversion (DAv) model. I conclude that ADHD, as classified by the DSM, fails to be natural kind, but not in the sense its critics often suppose. Rather *failing* to

pick out a *single* natural kind, the DSM classification of ADHD corresponds to several distinct natural kinds. Thus, ADHD earns its status as *real* insofar as it picks out distinct cognitive neurobiological mechanisms (i.e. natural kinds).

Before I address these questions however, I begin with the DSM classification of ADHD. Along the way I will address the argument that, due to its operational and descriptive approach to classification, the categories identified by the DSM are terminally unsuitable for research into their underlying mechanisms. If sound, this argument would seem to rule out any possibility of assessing whether ADHD is, or might be, an HPC-kind. I will show that this criticism, at least with respect to ADHD, can be met, and that it is prima facie reasonable to treat the DSM category for ADHD as a suitable object of scientific study.

4.2**: ADHD and the DSM-5**

The DSM-5 classifies ADHD in the superordinate category "Neurodevelopmental Disorders" (APA 2013). This tacit acknowledgment of the role played by neurobiology in the disorder represents a slight departure from previous DSM classifications of ADHD, which were entirely atheoretical and acausal in their characterization. Nevertheless, we shouldn't overstate the difference between the DSM-5's classification of ADHD and the classification provided in previous editions. The new entry

still contains no specific information about etiology or neuropathology, and the disorder and symptoms are still exclusively described in terms of observable behavior.

ADHD is characterized as, "A persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development" (APA 2013, 59). The manual distinguishes between three "presentations" of the disorder: ADHD-PI (predominately inattentive), ADHD-PH (predominately hyperactive-impulsive), and ADHD-C (combined presentation) (APA 2013, 60). This tripartite division was originally introduced in DSM-IV, although that edition called them "subtypes" (APA 1994, 80). Subsequent longitudinal studies showed these distinctions were essentially nominal. The subtypes did not show significant differences in treatment, course, or outcome and, moreover, they were temporally unstable. This is primarily because hyperactive behavior tends to remit in late-adolescence, so a young child diagnosed with ADHD-C might later qualify for an ADHD-I diagnosis as an adult (Willcutt et al. 2012). The DSM-5 language of clinical "presentations" is intended to signal this fluidity (DuPaul and Stoner 2014).

For children 16 and younger, a diagnosis of ADHD-PI, ADHD-PH, or ADHD-C requires:

- The presence of six of nine symptoms in the inattentive (ADHD-PI) or hyperactive-impulsive domain (ADHD-PH) or both (ADHD-C)
- The symptoms must have persisted longer than six months
- The symptoms must be inconsistent with developmental level (APA 2012, 59-60)[19]

There are three further criteria an individual must meet for a diagnosis. First, the symptoms must cause impairment in at least two settings (e.g. social, academic, or occupational). Second, there must be evidence that "several" of the symptoms were present and impairing before age 12. Third, the symptoms must not exclusively occur during the onset of schizophrenia or some other disorder, and must not be better explained by another mental disorder (APA 2012, 60).

4.3: ADHD, the DSM and Diagnostic Discrimination

Psychiatric researchers use the DSM criteria to select groups of individuals to study, so it plays a crucial role in the scientific investigation into the causal underpinnings of mental disorders. But the DSM's classification criteria are also subject to an enormous amount of criticism. Here I consider an objection that poses a particular problem for the argument I pursue here. The objection has been spelled out in many different ways, but the gist of it says that the DSM's categorical,

[19] The DSM-5 is the first edition of the manual to include age-specific symptom thresholds (five out of six) and descriptions for adults (APA 2012, 59-60).

operational, and descriptive-atheoretical approach to classification renders its categories terminally unsuitable as a guide to scientific research (Poland and Van Eckardt 2013).

Kathryn Tabb provides a reasonably clear statement of this critique. Tabb argues that psychiatrists err in assuming that the DSM categories will facilitate diagnostic discrimination. That is, psychiatrists who use the DSM assume its categories lump patients together in such a way as to, "…allow for relevant facts about mental disorder to be discovered." (Tabb 2015, 1047-1048) For the present case (ADHD), the "relevant facts" are those pertaining to the mechanisms underlying the symptoms. Tabb then provides a sustained argument that the DSM does not group patient populations in a way that facilitates successful investigation into the underlying causal mechanisms of their disorders.

Although I have no desire to mount a defense of the DSM, I believe the force of this criticism is limited. The individual conditions classified by the DSM are clearly a varied bunch, so as a *general* critique, this argument may cut against some DSM categories more than others. Harold Kincaid argues that it is better think of the DSM categories as potentially exemplifying "heterogeneous validity" (Kincaid 2017, 279), whereby some categories might pick out reasonably homogenous groups at some interesting theoretical level of description, and some of them

might not. The larger point is that if we want to know whether a DSM category can support fruitful scientific research – that is, research into underlying mechanisms – then thinking about the issue in the abstract is probably not the best way to answer that question. Rather, we need to look at individual conditions and make that assessment on a case-by-case basis.

With respect to ADHD, there are several reasons that prima facie suggest it is a suitable category for research into underlying mechanisms. First, ADHD type behaviors are highly familial, if not highly heritable. Family studies consistently demonstrate that parents and siblings of children with ADHD have a two to eight-fold increase of qualifying for the disorder (Faraone and Biederman 2014, 52849). A meta-analysis of twin studies from populations in the United States, Scandinavia, Australia, and the European Union yielded a heritability estimate of 76%, which suggests a strong genetic contribution to the ADHD phenotype. Adoption studies further support the inference of a genetic contribution, since biological relatives of ADHD children are found to have the condition at a higher rate than adoptive relatives of ADHD children; the risk of ADHD in adoptive relatives is similar to what one would find an a non-ADHD control group (Faraone et al. 2005). Second, factor analysis and other statistical measures consistently show that the two symptom dimensions in ADHD – inattention and hyperactivity-impulsivity – are separate, yet

highly correlated in individuals diagnosed with the disorder (Roberts et al. 2015). In other words, the symptom pattern in ADHD individuals consistently "hangs together." Third, the symptom complex manifested in ADHD can be reliably differentiated from other closely related disorders, such as Oppositional Defiant Disorder (Taylor and Sonuga-Barke 2008).

While none of evidence just cited is sufficient to show that the DSM classification of ADHD is an HPC-kind, it is sufficient to demonstrate that the DSM category picks out a reasonably consistent population of individuals -- i.e. a *stable* object for further investigation. Whether or not that population will be homogenous in terms of some further underlying mechanisms remains to be seen. But it is clearly *not* the case that the DSM classification of ADHD has nothing going with respect to diagnostic discrimination. Its classification systematically corresponds to several interesting patterns, and that is enough to justify its use for research, whatever limitations the DSM's operational and descriptive approach to classification may have more generally.

4.4: The Executive Function Model of ADHD

In the following discussion I assess whether the DSM classification of ADHD identifies an HPC-kind. I've previously clarified the notion of an HPC-kind in cognitive neurobiological terms. To recap, the DSM classification of ADHD is a natural kind if the members of that

category are sufficiently similar with respect to their underlying cognitive neurobiological structure and function, as captured by an appropriate mechanistic model (or mechanism sketch). Given the DSM's atheoretical stance, we can only pursue this question by engaging the research literature on causal models of ADHD. This is not a simple, straightforward proposition, given the volume of scientific papers about ADHD.[20] So, I will focus on one fairly well-studied model of ADHD, Barkley's executive function (EF) model. My selection of the EF-model is pragmatic; it's not the only model of ADHD, nor it is the newest. But Barkley's model is very important in the history of ADHD research, since it organized a large body of descriptive data about the disorder and linked it to a well-founded neuropsychological construct: EF. Thus, it attracted an enormous amount of research attention, and therefore provides us with a convenient means of efficiently navigating the literature.[21]

---

[20] One bibliometric study estimated that between 1980 and 2005, over 5,000 scientific papers have been published about ADHD (López-Muñoz, et al. 2008).

[21] Barkley thinks that the DSM classification of ADHD picks out two distinct (disorders). Specifically, he thinks ADHD-PI is a separate disorder, which he calls Concentration Deficit Disorder (CDC) (Barkley 2014). Though not as thoroughly researched as ADHD, the extant data tentatively indicates that individuals with ADHD-PI (or CDC) suffer from a different pattern of impairments than individuals with ADHD-C or ADHD-PH (Becker et al., 2014). I will not pursue the question of whether CDC constitutes a distinct natural kind. I mention it to note that the EF-model applies only to ADHD-C and ADHD-PH, which in any case

How to precisely define EF remains an ongoing controversy in the neuropsychological literature, but it is standardly assumed to be comprised a set of top-down, cognitive supervisory processes involved in abstract thought, planning, goal-directed behavior (Carlson, Zelazo, and Faja 2013).[22] I begin with a short summary of how the EF-model attempts to explain ADHD.

On the EF-model, ADHD is primarily a deficit in behavioral inhibition, which produces secondary deficits in executive functioning. The dysregulated interaction between the inhibitory and executive systems, in turn, produces a myriad of cognitive and behavioral deficits observed in ADHD individuals. Barkley argues that the behavioral and executive systems are implemented in the prefrontal-striatal network.

Barkley identifies four components of the executive system compromised in ADHD: spatial working memory, verbal working memory, reconstitution, and the self-regulation of affect. While slightly different in the details, Barkley's model of working memory is essentially the same as Alan Baddeley's influential neuropsychological model.

---

constitute the vast majority of individuals diagnosed with ADHD (Nigg 2006, 9).

[22] Barkley's model purports to be a *general* model of EF, although a precise definition of EF remains controversial in the neuropsychological literature. I mention this controversy only to point out that even if the particular facets of Barkley's EF-model are not exactly right, the overall construct of executive functions is very well supported.

Baddeley's model splits working memory into a central executive and two subsidiary components: the phonological and visuospatial storage/rehearsal subsystems. The phonological system is responsible for the temporary storage of verbal information, while the visuospatial system stores spatial representations. The central executive is responsible for selective attention, manipulation and retrieval of task-relevant information from long-term memory (Baddeley 2012). The phonological loop corresponds to Barkley's spatial working memory; the visuospatial system corresponds to nonverbal working memory; and the central executive corresponds to reconstitution (as best I can tell). Reconstitution, according to Barkley, operates on the contents of spatial/verbal working memory by decomposing and reassembling its contents into novel representations and/or behavioral sequences.

Self-regulation of affect involves two processes. First, it involves the regulation of emotional responses – both the subjective experience (e.g. anger) and any associated motor behavior (e.g. brow-furrowing when angry). Secondly, it involves the capacity to modulate one's arousal level in the service of goal-directed behavior, especially in the absence of external rewards (Barkley 1997).

Barkley argues that ADHD individuals exhibit deficits in most, if not all, of the aforementioned components of the executive system. But

the foundational deficit is a maturational delay in the brain's behavioral-inhibitory capacities. This system is causally necessary for the operation Executive Functions (Barkley 1997, 51). This inhibitory system is composed of three interrelated processes: halting a prepotent response, interruption of ongoing response, and interference control. Halting a prepotent response is the most important component. Barkley defines it as the capacity to arrest a learned behavioral pattern in the presence of a stimulus with which positive or negative reinforcement is associated. Interrupting an ongoing response refers to the capacity to halt a particular motor sequence in response to environmental feedback. Both of these processes create a temporary period of delay between environmental input and motor output, during which time executive processing can occur. Interference control is the capacity to protect the delay period created by latter two functions from either competing motor responses or task-irrelevant stimuli (Barkley 1997).

Let's now frame the EF-model in terms of the HPC-theory of natural kinds (as I've interpreted it). The EF-model says that the symptoms associated with ADHD are sustained by a particular cognitive neurobiological mechanism, located roughly in the prefrontal-striatal network. This mechanism, or better yet, network of mechanisms, may be specified coarsely as a pattern of interaction between two systems: the

behavioral inhibition system and executive system. In order to operate efficiently, the executive system causally depends on the inhibitory system to regulate automatic behavioral responses to incoming stimuli. The EF-model maintains that the development of the behavioral-regulatory system is *delayed* in individuals with ADHD. Since the inhibitory and executive systems involve several different components that are differentially related to one another, the EF-model obviously admits various patterns of interaction between them. For instance, failure to screen out task-irrelevant stimuli (interference control) allows too much information to seep into working memory, thereby overloading its limited storage capacity, resulting in reduced working memory performance (see **Illustration 1**). The pattern of interaction between the two systems is the mechanism that individuates ADHD *as* an HPC-kind. The ensuing suite of cognitive deficits – poor working memory, poor regulation of affect and so on – constitute the property cluster sustained by the mechanism. Consistent with the HPC-theory, it need not be case that every individual with ADHD instantiates every possible pattern or every possible cognitive deficit. However, if ADHD is an HPC-kind, then we should expect to find that most individuals diagnosed with the disorder implement a good deal of the structural and functional relationships specified by the EF-model.

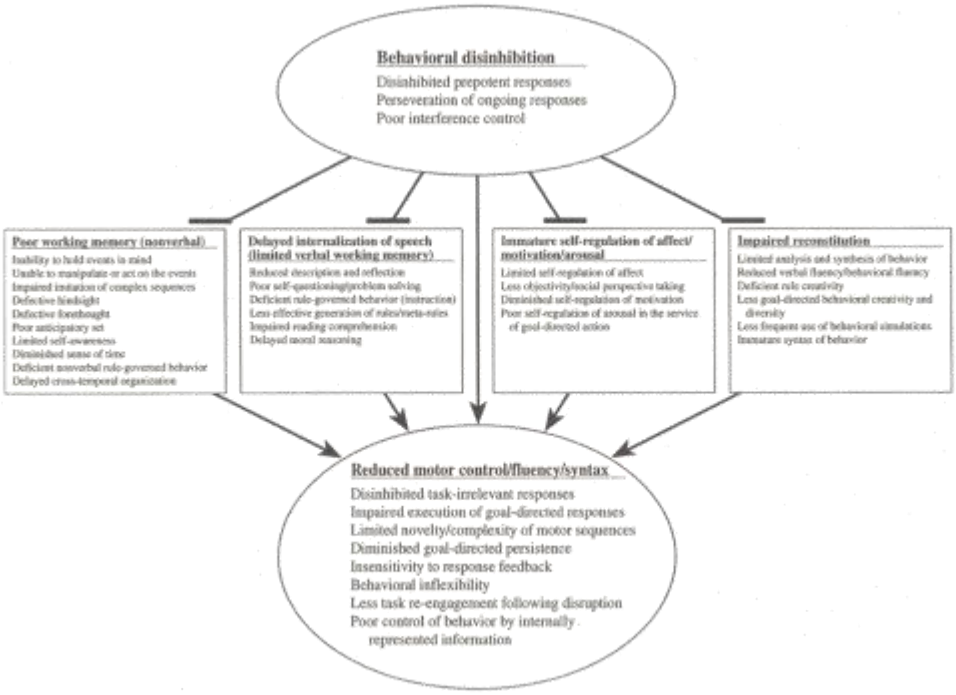

Illustration 1: Barkley's EF-Model of ADHD. From *ADHD and the Nature of Self-Control*, by Russell A. Barkley, (New York: Guilford Press, 1997). Reprinted with permission from the Guildford Press.

4.5: Evaluating the EF-Model as a Mechanism Sketch

The EF-model constrains the space of possible mechanisms in ADHD by bridging together information about the disorder from several different levels of organization. Barkley is quite explicit that any respectable model of ADHD *must* do this. After reviewing various lines of evidence – genetic, biological, and neuropsychological factors detected in ADHD individual – Barkley reflects on what a model of ADHD requires. He writes, "To meet these *constraints*, a model of ADHD must bridge the research findings on ADHD, neuropsychology (as it pertains to the

functions of the prefrontal lobes and related structures—the executive functions), and developmental psychology (as it pertains to the normal development of the behavioral or psychological traits mediated by these brain structures," (Barkley 1997, 45; emphasis mine).

Thus, the EF-model specifies a mechanism sketch for the underlying cause of ADHD behaviors. It provides the following features: (a) the computational processes likely to be responsible for the production of ADHD type behavior (i.e. the interaction between inhibitory and executive systems), (b) the regions of the brain implementing that mechanism (i.e. prefrontal-striatal network), and (c) a developmental processes responsible for the mechanism's operation (i.e. maturational delay in the behavioral inhibition system). The advantage of interpreting the EF-model as a mechanism sketch is that we can evaluate it empirically, just as we would any other mechanism sketch.

Although the EF-model of ADHD draws on a general theory of EF, our argument for ADHD realism need not be concerned with whether, strictly speaking, that particular general model of EF is precisely right. For, EF is widely recognized as a well-founded neuropsychological system. A number of alternative models of EF have been proposed, but all make use of similar types of components and processes (e.g. working memory). What we want to know about EF-models *qua* ADHD realism is

whether all (or most) of the individuals diagnosed with ADHD exemplify the same, or similar, type of cognitive neurobiological mechanism. With that in mind, I now turn to consider evidence in favor of Barkley's and similar models.

Neuroimaging studies consistently implicate the right hemisphere, including many components of the prefrontal system, in ADHD individuals. ADHD children show about a 3-5% reduction in right hemispheric volume (Hynd, et al. 1990, Castellanos, et al. 1994, 1996, and 2001; Filipek, et al. 1997; Seidman, Valera, and Makris 2005) and about a 12% volumetric reduction in four regions – the prefrontal cortices, basal ganglia/striatum, cerebellum, and corpus callosum – strongly associated with complex, goal-driven behavior (Nigg and Nikolas 2008, 311). On the assumption that structural differences imply functional differences, these findings are broadly favorable to the EF-theory of ADHD.[23]

In addition to these gross structural/volumetric differences, more

[23] Evidence from structural neuroimaging research strongly suggests that such structural differences *do* imply functional differences. The most robust evidences derives from sensory discrimination tasks. The size of the olfactory bulb, for example, positively co-varies with the ability to discriminate between smells. And although not as well understood, this positive correlation holds for other higher-order cognitive capacities as well, including memory and executive functioning. This is of course not to say that "bigger is better" tout court. These findings only hold within a particular range of volumetric values. Once a particular section of neural tissue exceeds those values, function begins to decline (Bigler 2013, 191-192).

specific support for the EF-model derives from neuropsychiatry. As it stands, there is well-replicated evidence for three neuropsychological deficits posited by the EF-model: inhibitory control, interference control, and working memory. Much of this data derives from task-based functional magnetic resonance imaging (fMRI), which comprises the largest pool of imaging research on ADHD (Cortese and Castellanos 2015, 47). Task-based fMRI experiments are a type of top-down, activation experiment, which we discussed in the previous chapter. In these studies, a subject performs a task believed to tap the relevant EF while situated in an fMRI machine. For the sake of illustration, I'll mention one example. The go/no-go test is a widely accepted measure of the capacity to inhibit a prepotent response. While there are several different experimental designs of the go/no-go test, one common variant presents the subject with a randomly alternating sequence of two stimuli, X and Y, with instructions to press a button ('go') when X appears on the screen and to withhold that response ('no-go') when presented with Y. Typically, researchers first prime the subject with a long sequence of X's (go-stimuli) in order to establish a prepotent response. Inhibitory capacity is calculated by the number of erroneous responses to no-go stimuli; that is, the number of instances when the subject presses the button in response to Y.

On average, ADHD individuals perform more poorly on this and another neuropsychological measure of interference control (e.g. Nigg 2001; Lansbergen, Kenemans, and van Engeland 2007); and the fMRI data consistently implicates both abnormal patterns of activation (e.g. hypoactivation) and structural abnormalities in the prefrontal and associated regions (Hard, Radua, and Nakao 2013). Moreover, hypofunctioning is likely the causative factor for failed inhibitory control and these abnormalities appear to be largely genetic in origin.[24] As for working memory, ADHD individuals consistently test as impaired for all three components, with the largest deficits residing in the central executive, a smaller deficit in the visuospatial system, and a still smaller deficit in the phonological component (Martinussen, et al. 2005; Alderson et al. 2007; Rapport, et al. 2008).

Lastly, recent work supports the claim that ADHD involves a maturational delay, and that this delay is strongly associated with impairment. A seven-year longitudinal study by Philip Shaw tracked

[24] Evidence for this observation derives from task-based studies examining adolescence with ADHD, their unaffected siblings, and typically developing controls. In these studies, ADHD kids and their siblings show similar patterns of under-activation in the frontal-striatal and frontal-parietal regions relative to controls. But, in instances of no-go responses – where behavioral inhibition is tapped - unaffected siblings do better than their affected siblings and show intermediate levels of hypoactivation between the ADHD group and controls (e.g. van Rooij, et al. 2015).

cortical development in ADHD children and non-ADHD children. In typically developing controls, Shaw and colleagues found that the brain reaches peak cortical thickness at age 7, and 50% peak cortical surface area at age 12. By comparison, the study showed that children with ADHD do not hit these developmental benchmarks until 10 and 14, respectively and that the most prominent site of delay resides in the lateral prefrontal cortex (2012). The same team examined cortical development in adults diagnosed with ADHD as children. They found that the number of ADHD symptoms present in adults co-varied with cortical thickness. Specifically, the developmental trajectory of adults whose symptoms had largely remitted converged with that of typically developing controls, while adults with persistent symptoms showed fixed, non-progressive deficits in cortical thickness (Shaw et al. 2013).

4.6**: ADHD: Multiple Deficits and Multiple Kinds**

Let us take stock. So far, we've considered evidence broadly sympathetic to the EF-model. At the neurobiological level, ADHD is robustly correlated with structural-functional abnormalities in the prefrontal region of the brain. These areas are all associated with inhibitory and interference control, and working memory. ADHD individuals consistently do worse on neuropsychological tests designed to measure these processes. Moreover, there is very good evidence for a

maturational lag in cortical development of these regions and that this lag co-varies with symptom severity.

Minimally, it is reasonable to conclude the DSM classification of ADHD picks out at least one HPC and so, one natural kind. To reiterate, I define an HPC as a cognitive neurobiological mechanism and a mental disorder is an HPC-kind insofar as the symptoms associated with it are sustained by a particular *kind* of mechanism. In terms of the HPC-theory, the mechanism is composed of the maturational delay and the dysregulated interaction of the inhibitory and executive systems, and the suite of cognitive deficits are the property clusters sustained by the causal interaction between the two. These conspire to cause the various behaviors indicative of ADHD: inattentiveness, hyperactivity, and impulsiveness. Thus, the DSM classification of ADHD corresponds to *at least* one HPC-kind. However, it is further question whether the DSM classification picks out *only* individuals with executive deficits. On this score, the evidence strongly suggests it does not.

The first set of considerations derives from an important 2005 meta-analysis conducted by Willcutt and colleagues, which examined over 3,734 ADHD children across 13 measures of executive functioning (Willcutt et al. 2005). On a group level, the study found a significant association between ADHD and decreased performance on EF tasks.

However, as Willcutt et al. further note, this association is weaker than what one would expect if the EF-model, qua single deficit causal model, were true. In the ADHD group, effects size for executive functioning were in the moderate range (d= 0.60–0.80), which is equivalent to 50% overlap with the control group. Another paper by the same group examined pooled results of EF measures from three ADHD research centers (267 ADHD individuals vs. 600 controls). They found that while 80% of ADHD children were impaired in at least one domain of EF, only 53% show impairment in two EF domains, and only 31% in three (Nigg, et al. 2005).

These findings indicate that only a minority individuals diagnosed with ADHD are executively impaired. But all, or most, of these individuals presumably exemplify the symptoms associated with ADHD, so the further question is whether there is evidence for *alternative* mechanisms sustaining these behaviors? If so, then that would constitute positive evidence *for* the presence of distinct HPC-kinds within the DSM classification of ADHD. I now argue there is such evidence for distinct, non-EF related HPC-kinds within the ADHD population.

First, ADHD individuals consistently show deficits in relatively independent, non-EF neuropsychological domains. These include: arousal and activation (Fair, et al. 2012), temporal processing, (Toblak, Dockstadera, and Tannock 2006), speech and language (Tomblin and

Mueller 2012), auditory processing (Cheung and Siu 2009) and fine motor control (Fliers, et al. 2009). Second, neuroimaging data shows functional and structural deviations that clearly exceed the EF-model's characterization of ADHD as a prefrontal syndrome. For a useful summary of this literature, including regions associated with both EF and non-EF, see **Illustration 2**.

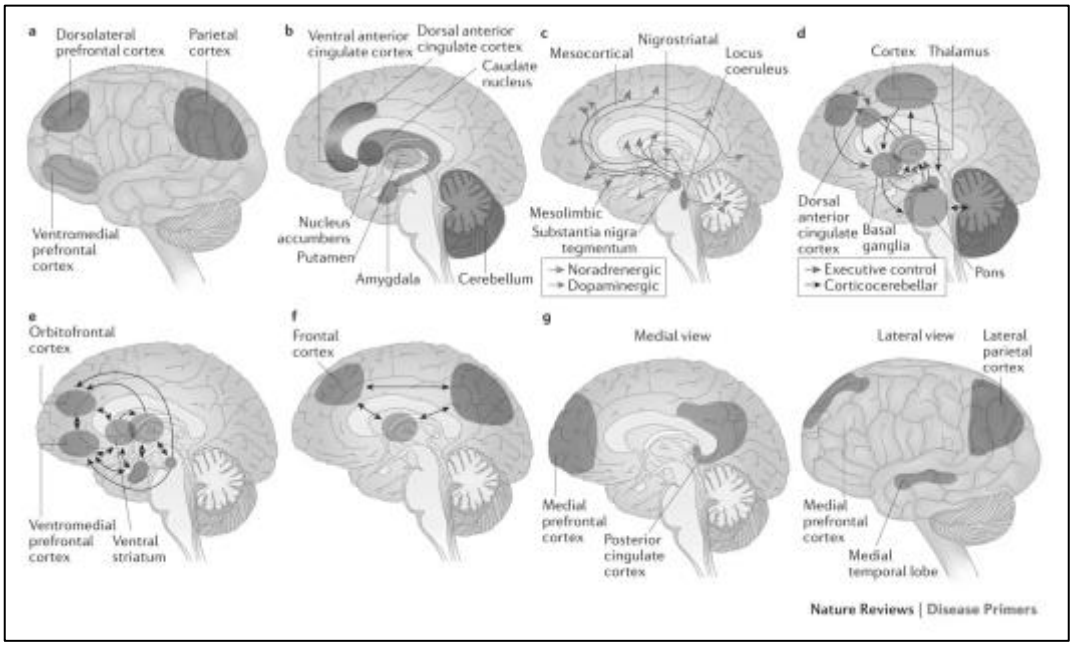

Illustration 2: Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Disease Primers (Faraone, et al. 2015), © 2015.[25]

[25] A) Cortical regions, linked to working memory (dorsolateral prefrontal cortex), planning and complex decision making (ventromedial prefrontal cortex), and orientation of attention (parietal cortex). B) Affective components of EF, C) Planning, initiating motor response (dopaminergic system) and modulation of arousal (noradrenergic system), D) Working memory, cognitive control, flexible adaptation, inhibition. E) Reward network. F) Alerting network; supports attentional functioning. G) Default mode network. Activates during "mind-wandering"; deactivates

Specific evidence for a distinct HPC-kind derives from research into the motivational systems of ADHD individuals. Individuals with ADHD commonly exhibit a preference pattern for smaller, immediate rewards over larger, delayed rewards. This has led many researchers to posit a motivational impairment involved in the production of ADHD-type behaviors.[26] Edmund Sonuga-Barke argues that this preference pattern is a behavioral manifestation of a motivational style called "delay aversion" (DAv) (2005). Sonuga-Barke's model is very sophisticated, but the gist of it is that DAv is the product of two interacting processes. The first is a biologically based deficit in reward signaling, specifically a higher than average temporal discounting of rewards delivered after delay. Neurologically, this deficit implicates the mesolimbic reward pathway. The second is environmental. Children with this impairment are more likely to come into conflict – censure, reproach, etc. - with their caregivers, since delaying gratification is a skill that most children are expected to acquire as they grow older. The negative input from the

---

when (D) activates in typically developing controls. ADHD individuals show distinctive patterns of desynchronization between the posterior cingulate cortex and medial prefrontal cortex.

[26] In the 1990's, motivational models were pitched as competitor accounts to the EF-model. For a comprehensive review of these, see: Luman et al. (2005).

environment conspires with the biologically based impairment in temporal signaling to create DAv. Sonuga-Barke reasons that many of the symptoms associated with ADHD are a behavioral expression of the DAv. Of course, the most obvious manifestation is just a consistent preference for short-term, smaller gains over long-term, larger, rewards, but Sonuga-Barke considers other possibilities. For instance, in circumstances when delay is unavoidable (e.g. a classroom), a child might engage in fidgety behavior as an effort to create non-temporal stimulation so as to reduce the experience of delay (Sonuga-Barke 2005).

In broad outline, several lines of research support the DAv-model. Preference for immediate rewards can reliably differentiate individuals with ADHD from typically developing controls on two measures of delay tasks, the Maudsley Index of Delay Aversion and The Choice Delay Task (Sonuga-Barke, et al. 2008). Delay aversion and executive dysfunction have both been found individuals with ADHD, but appear to be unrelated dimensions; that is, they affect different subsets of ADHD individuals (Solanto, et al. 2001) Attentional bias toward delay-related cues in the environment (Sonuga-Barke, et al. 2004) and higher levels of frustration when unexpected delays are imposed (Wilbertz, et al. 2013) have both been frequently documented in ADHD individuals. Imaging studies consistently implicate hypofunctioning in the ventral-striatal reward

system in response to delayed reward in ADHD individuals (Ströhle, et al. 2008, Plichta, et al. 2009). Worthy of mention here is a landmark study by Nora Volkow and colleagues, which imaged the mesolimbic dopamine pathway in a sample of 53 medication-naïve ADHD adults and 44 healthy controls over the course of eight years. Compared to controls, ADHD individuals show a significant reduction in the number of dopamine receptors in the midbrain and accumbens[27] (Volkow, et al. 2009). Given dopamine's well-known role in the production and maintenance of motivational salience (Bromberg-Martin, Matsumoto, and Hikosaka 2010), this is compelling physiological evidence, if not specifically for DAv, then at least for a non-EF mechanism underwriting ADHD. That is, the evidence indicates spatially *distinct* mechanisms (i.e. located in other regions besides the prefrontal cortex) that, in turn, implement computationally distinct *cognitive processes* (i.e. non-EF). Thus, this evidence supports the existence of at least two distinct natural kinds within the DSM category. One broadly involves an executive deficit, implicating the prefrontal region of the brain; the other involves a motivational deficit, which involves a functionally distinct neurological circuit in the midbrain and accumbens.

[27] AKA the mesolimbic dopamine pathway

4.7**: Conclusion**

It is of course possible that a single process will eventually explain all of the various impairments, and structural and functional abnormalities associated with ADHD. But most researchers deem it unlikely. Sonuga-Barke succinctly summarizes the present consensus in the field:

> Given the complexity of the picture and obvious causal heterogeneity in ADHD, researchers need to study the relation among cognitive, energetic, motivational, and executive processes in ADHD. The most likely scenario is *that there is not one single core deficit that explains the condition – ADHD is not a single pathophysiological entity*. More likely, ADHD is made up of different groups of patients with their own distinctive etiologies and pathophysiologies.
>
> [Sonuga-Barke 2013, 561; emphasis mine]

Over the course of the last fifteen years or so, research has shifted away from models emphasizing a single core deficit that explains all of the symptoms associated with ADHD, to multiple-deficit models. Sonuga-Barke's work was an early example of this trend. His "dual-pathway" model conjoined his DAv model with an executive function model, and proposed that each represented two dissociable deficits, either of which could result in an ADHD diagnosis[28]; other researchers carve up the ADHD population differently.

[28] Sonuga-Barke has since abandoned the dual-pathway model for a tri-pathway model, which adds a temporal processing deficit to the mix (2010).

It is impossible to say in advance what the final verdict of these investigations will be. But, the evidence to date is sufficiently robust that we can answer the primary question of this study: is ADHD a natural kind? The answer is no. But that provides no succor to the irrealist. ADHD is not like phlogiston; its extension isn't empty. It fails to be a natural kind in the sense of being a *single* HPC, but the totality of evidence strongly indicates it is *several* HPCs, with at least one involving the inhibitory-executive systems and another involving the motivational-reward systems. According to the realist strategy pursued here, ADHD thus fairly earns its keep as a *real* mental disorder.

Chapter 5: Summary of Conclusions and Further Reflections

Summary

In chapter 2 I examined the controversies surrounding the "reality" of mental illness. We cashed out those arguments in terms of mind-independence, and we cashed out mind-independence in terms of act-of-classification (AC) independence. Framed this way, realists hold that there is a positive body of fact about mental disorders (aside from our value judgments) that can be discovered and studied. In other words, the category of *mental disorder* is not exhausted merely by our value judgments. We then examined the primary way in which realists have sought to establish AC-independence about mental illness: the definitional debate.

Realists in the definitional debate do not deny the role of value judgments in guiding our decisions to medicalize certain conditions. However, they assert that a necessary condition for any category to count *as* a legitimate medical condition is that it must be produced by an objectively defined dysfunction. Thus, in the context of the definitional debate, establishing AC-independence qua mental disorders reduces to establishing AC-independence qua a theory of function and dysfunction. I argued that this burden is far stronger than anything AC-independence

requires, which in turn opens up alternative avenues to the realist for pursuing their case. One such way, which is my preferred approach, is establishing that mental disorders are natural kinds.

In chapter 3, we examined two alternative approaches to natural kinds: essentialism, and Richard Boyd's homeostatic property cluster (HPC) account. I argued that essentialism is a non-starter, since it results in a taxonomy radically at odds with the natural sciences, broadly construed. I discussed why the HPC-account is better, although its prospects as a *general* theory of kind-hood remain to be seen. I then proposed to interpret ADHD as an HPC-kind. I further precisified the view by consulting the philosophy of neuroscience literature on mechanisms. I argued that the DSM classification of ADHD is an HPC-kind provided that it corresponded to a common cognitive neurobiological mechanism within all individuals that have ADHD.

I pursued this approach in chapter 4, primarily through an explication of Russell Barkley's Executive Function Model (EF) of ADHD. Barkley's model argues that the symptoms of ADHD are caused by the dysregulated interaction between the brain's inhibitory and executive systems, which is ultimately due to a prefrontally mediated developmental delay in the former. I surveyed the evidence from the cognitive neuropsychological literature and found that it supports many of

the EF-model's claims. Specifically, ADHD individuals: a) show widespread structural and functional abnormalities in the prefrontal regions, b) consistently show deficits in inhibitory control, interference control, and verbal and non-verbal working memory, and c) ADHD individuals show a maturational lag in cortical thickness the right prefrontal regions. Furthermore, these developmental trajectories co-varied with symptom expression. I argued that this evidence constituted a reasonable basis to conclude that the DSM classification of ADHD corresponded to at least one HPC-kind. I further considered whether the EF-model accounts for all (or nearly all) of individual cases of ADHD, and argued that it does not. There is compelling evidence for the presence of alternative cognitive neurobiological mechanisms in some individuals with the condition. And, on the HPC-account of natural kinds, differences in mechanism imply a difference in kind. Therefore, the evidence shows that there are alternative HPC-*kinds* of ADHD disorders. Thus, I concluded that ADHD is a *real* disorder -- indeed it is several real disorders -- in that there are AC-independent facts about it, facts that are independent of anyone's value judgments.

Reflections

As Rick Mayes and his colleagues aptly remarked, "Attention deficit/hyperactivity disorder (ADHD) holds the distinction of being the

most extensively studied pediatric mental disorder and one of the most controversial…" (Mayes et al. 2009, 1). Like other scientific objects (e.g. climate change, vaccines) most of the controversy about its reality is largely located outside of disciplines tasked with studying it. Among the vast majority of medical professionals, ADHD's reality – at least in broad outline – is a settled question. There does persist a vocal and, to be frank, slightly blinkered minority that refuses to engage in any meaningful way with the vast body of evidence underwriting this consensus. Nonetheless, like global warming skeptics, ADHD skeptics have a knack for keeping the controversy alive in the public mind. In one way or another, these arguments try to undermine the scientific understanding of ADHD, primarily as a means to abolish it *as a medical category*. To these critics, I suggest that an honest reckoning with the scientific evidence would actually suit, rather than hinder, their purposes. For, even after the reality question is settled – and to be clear, for ADHD I believe it *is* settled – it remains an open issue as to whether we ought to medicalize it. I think there is a reasonable case for doing so. Individuals with ADHD are at risk for a whole range of undesirable life outcomes, but it is unclear whether these are sufficient to justify its medicalization. Thus, unlike questions about its reality, questions about its medicalization are not yet foreclosed. So a constructive dialogue is still possible, indeed mandatory, about how

we ought to care for individuals with ADHD. These people are worth caring about to be sure, but whether that care should be primarily medical is still a reasonable question to debate.

The primary conclusion of this dissertation sits somewhat ambiguously within the larger debates about the DSM in the philosophy of psychiatry. In one sense, it vindicates the standard criticism that the DSM's approach to classification is likely to yield heterogeneous categories. On the other hand, it is also true that our understanding of the disorder has grown exponentially over the last 30 years, which seems to cut against some of the more extreme forms of skepticism about the scientific prospects of the DSM categories. After all, nearly everything we now know about ADHD is in one way or another filtered through the DSM. This is not to say that the DSM classification of ADHD hasn't hindered research in some ways. It surely has. But it does suggest caution about endorsing global assessments of the DSM's suitability as a useful tool for research.

In making the case for realism about ADHD, I have consulted only a subset of the research literature about the disorder. Barkley's EF-model and Sonuga-Barke's dual-pathway model belong to the larger class of neuropsychological models of ADHD, including Sergeant's Cognitive-Energetic model (Sergeant, 2000), Tripp and Wickens' (2008) dopamine

transfer deficit model, and Sagvolden et al.'s Dynamic Developmental Theory (2005). These models may be thought of as the "first-wave" of causal theorizing about ADHD, as they were developed in tandem with the revolution in brain imaging technology that occurred during the 1990s. These first-wave models attempted to parsimoniously explain the suite of deficits and behaviors observed in ADHD individuals in terms of a set of primary deficits (Faraone and Biederman 2014). It is unclear whether this level of analysis will ultimately prove viable as our understanding of the etiology of ADHD increases.

At the neurological level, the emerging picture is that ADHD individuals exhibit widespread patterns of atypical structural and functional connectivity all over the brain. In individuals without ADHD, the general trajectory of brain development proceeds towards more focalized patterns of functional connectivity over time. ADHD individuals by contrast, tend to exemplify much more diffuse patterns of connectivity. For example, in tests of executive functioning, ADHD individuals activate a much more distributed set of structures than their typically developing counterparts, which produces a much more inefficient solution to the kinds of problems posed by these instruments (Faraone and Biederman 2014). The upshot is that many researchers are beginning to conceptualize ADHD – and more specifically, the global structural and functional abnormalities

associated with the condition – as the final product of an extended process of atypical neurological development. Once these developmental mechanisms are better understood, the resultant taxonomy might be different than a taxonomy constructed solely on the basis of the synchronic state of the nervous system, which is essentially what these first-wave models do. This doesn't render the argument offered in chapter 4 erroneous per se; ADHD individuals *do* embody, to a rough approximation, the deficits described in the EF and dual-pathway models. It only shows that further investigation might lead us to revise the first-wave models in such a way as to better account for the experimental evidence. At any rate, if there is future philosophical work to be done on ADHD along the lines of this dissertation, a good place to start is with trying to get a better understanding of the implications of this emerging *developmental* understanding of the disorder.

References

Aristotle. *Metaphysics*, translated by C.D.C. Reeve. Indianapolis, Indiana: Hackett, 2016.

Aristotle. *Aristotle: Categories and De Interpretationae*, translate by J.L. Ackrill. Oxford: Clarendon Press, 1963.

Alderson, Matt, Mark Rapport, and Michael Kofler. "Attention-Deficit/Hyperactivity Disorder and Behavioral Inhibition: A Meta-Analytic Review of the Stop-signal Paradigm." *Journal of Abnormal Child Psychology* 35, no. 5 (2007): 745–758.

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edition. Washington D.C.: American Psychiatric Association, 1994.

—. *Diagnostic and Statistical Manual of Mental Disorders*. 5th edition. Washington D.C.: American Psychiatric Association, 2013.

Armstrong, Thomas. "ADD as a Social Inventions." *Education Week*, October 18, 1995.

Aucouturier, Valérie and Steeves Demazeux. "The Concept of 'Mental Disorder.'" In *Health, Illness and Disease: Philosophical Essays*, edited by Havi Carel and Rachel Cooper, 77-93. New York: Taylor and Francis, 2014.

Audi, Robert. "The Applications of Conceptual Analysis." *Metaphilosophy* 14, no. 2 (1983): 87-106.

Baddeley, Alan. "Working Memory: Theories, Models, and Controversies." *The Annual Review of Psychology* 63 (2012): 1-29.

Barkley, Russell A. "Behavioral Inhibition, Sustained Attention, and Executive Functions: Constructing a Unifying Theory of ADHD." *Psychological Bulletin* 121, no. 1 (1997): 65-94.

Baughman, Fred and Craig Hovey. *The ADHD Fraud, How Psychiatry Makes "Patients" of Normal Children*. Victoria, British Columbia: Trafford Publishing, 2006,

Bayer, Ronald. *Homosexuality and American Psychiatry: The Politics of Diagnosis*. Princeton, N.J.: Princeton University Press, 1981).

Becker, Stephen P., Stephen A. Marshall, and Keith McBurnett. "Sluggish Cognitive Tempo in Abnormal Child Psychology: An Historical Overview and Introduction to the Special Section." *Journal of Abnormal Child Psychology*, January 2014: 1-16.

Bechtel, William, and Robert Richardson. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge: MIT Press, 2010.

Beebee, Helen, and Nigel Sabbarton-Levy. "Are Psychiatric Kinds Real? " *European Journal of Analytic Philosophy* 6, no. 1, (2010): 11-27.

Bigler, Erin. "Structural Neuroimaging." In *Neuropsychology: Science and Practice*, edited by Sandra Koffler, Joel Morgan, Ida Sue Baron and Manfred Greiffenstein, 189-207. Oxford: Oxford University Press, 2013.

Boone, Worth and Gualtiero Piccinini. "The Cognitive Neuroscience Revolution." *Synthese* 193, no. 5, (2016): 1509–1534.

Boorse, Christopher. "Health as a theoretical concept." *Philosophy of Science* 44, no. 4 (1977): 542-573.

—. "On the Distinction Between Health and Disease." *Philosophy and Public Affairs* 5, no. 1 (1975): 49-68.

—. "What a Theory of Mental Health Should Be." *Journal for the Theory of Social Behaviour* 6, no.1 (1976): 61–84.

—. "A Second Rebuttal on Health." *Journal of Medicine and Philosophy* 39, no. 6 (2014): 683-724.

Boyd, Richard. "Homeostasis, Species, and Higher Taxa." In *Species: New Interdisciplinary Essays*, edited by R.A. Wilson, 141-185. Cambridge: MIT Press, 1999.

—. "How to be a Moral Realist." In *Essays on Moral Realism*, edited by G. Sayre-McCord, 181-228. Ithaca, New York: Cornel University Press, 1988.

—. "Kinds as the 'Workmanship of Men': Realism, Constructivism, and Natural Kinds." In *Rationality, Realism, Revision,* edited by Julian Nida-Rümelin, 52-89. Berlin: Walter de Gruyter, 2000: 52-89.

—. "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 61, no. 1/2 (1991): 127-148.

—. "Realism, Natural Kinds, and Philosophical Methods." In *The Semantics and Metaphysics of Natural Kinds*, edited by Helen Beebee and Nigel Sabbarton-Leary, 212-234. New York: Routledge, 2010.

—. "What Realism Implies and What it Does Not." *Dialectica* 43, no. 1-2 (1989): 5-29.

Boyd, Richard, Keller Roberto, and Quentin Wheeler. "The Illogical Basis of Phylogenetic Nomenclature." *The Botanical Review* 69, no. 1 (2003): 93-110.

Bromberg-Martin, Ethan, Masayuki Matsumoto, and Okihide Hikosaka. "Dopamine in Motivational Control: Rewarding, Aversive, and Alerting." *Neuron* 68, no. 5 (2010): 815-834.

Buckner, Cameron. "Functional Kinds: A Skeptical Look." *Synthese* 192, no. 12, 2015: 3915–3942

Carlson, Stephanie F., Philip David Zelazo, and Susan Faja. "Executive Function." Vol. 1, in *Oxford Handbook of Developmental Psychology*, edited by Philip David Zelazo, 706-703. Oxford: Oxford University Press, 2013.

Carel, Havi and Rachel Cooper. *Introduction*. In *Health, Illness, and Disease: Philosophical Essays*, 1-22. New York: Routledge, 2014.

Castellanos, FX, et al. "Quantitative Brain Magnetic Resonance Imaging in Attention-Deficit Hyperactivity Disorder." *Archives of General Psychiatry* 53, no. 7 (1996): 607-616.

Castellanos, FX, et al. "Quantitative Brain Magnetic Resonance Imaging in Girls with Attention-Deficit/Hyperactivity Disorder." *Archives of General Psychiatry* 58, no. 3 (2001): 289-295.

Castellanos, FX, et al. "Quantitative Morphology of the Caudate Nucleus in Attention Deficit Hyperactivity Disorder." *The American Journal of Psychiatry* 151, no. 12 (1994): 1791-1796.

Cheung, PP, and AM Siu . "A Comparison of Patterns of Sensory Processing in Children with and without Developmental Disabilities." *Research in Developmental Disabilities* 30 (2009): 1468–1480.

Churchland, Paul. "Conceptual Progress and World-World Relations: In Search of the Essence of Natural Kinds." In *A Neurocomputational Perspective: The Nature of the Mind and the Structure of Science*, 281-296. Cambridge, MA: MIT Press, 1992.

Conrad, Peter. "The Discovery of Hyperkinesis: Notes on the Medicalization of Deviant Behavior." *Social Problems* 23 (1975): 12-21.

Conrad, Peter, and Joseph Schneider. *Deviance and Medicalization: From Badness to Sickness.* 2nd Edition. Philadelphia: Temple University Press, 1980/1992.

Cooper, Rachel. *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders.* Dordrecht, Netherlands: Springer, 2005.

Cortese, Samuele, and F. Xavier Castellanos. "Attention Deficit/Hyperactivity Disorder." In *Neurobiology of Brain Disorders: Biological Basis of Neurological and Psychiatric Disorders*, edited by Michael Zigmond , Lewis Roland and Joseph Coyle, 42-58. London: Elsevier Inc., 2015.

Crastley, Kelso, and Richard Samuels. "Cognitive Science and Explanations of Psychopathology." In *The Oxford Handbook of Philosophy and Psychiatry*, edited by K.W.M Fulford et al., 385-412. Oxford: Oxford University Press, 2013.

Craver, Carl. "Dissociable Realization and Kind Splitting." Philosophy of Science 71, No. 5, (2002): 940-971.

—. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience.* Oxford: Oxford University Press, 2007.

—. "Levels of Mechanisms: a Field Guide to the Hierarchical Structure of the World." In *The Routledge Companion to the Philosophy of Psychology*, edited by John Symons and Paco Calvo, 387-399. New York: Routledge, 2009.

Craver, Carl and James Tabery. "Mechanisms in Science." *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), edited by Edward N. Zalta, URL = <https://plato.stanford.edu/archives/spr2017/entries/science-mechanisms/>.

Craver, Carl and Lindley Darden. *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago, University of Chicago Press, 2013.

Craver, Carl, and William Bechtel. "Top-Down Causation without Top-Down Causes." *Biology and Philosophy* 20, (2007): 715-734.

Decker, Hannah. *The Making of DSM-III: A Diagnostic Manual's Conquest of American Psychiatry.* Oxford: Oxford University Press, 2013.

Denckla, MB, and RG Rudel. "Anomalies of motor development in hyperactive boys." *Annals of Neuroscience* 3, no. 3 (1978): 231-233.

Douglas, Heather. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press, 2009.

DuPaul, George and Gary Stoner. *ADHD in the Schools: Assessment and Intervention Strategies.* 3rd Edition. New York: Guilford Press, 2014.

Dupré, John. "Natural Kinds." In *A Companion to the Philosophy of Science*, edited by W. H. Newton-Smith, 311-319. Malden, MA: Blackwell Publishing, 2000.

Ellis, Brian. *Scientific Essentialism.* Cambridge University Press, 2001.

—. *The Philosophy of Nature: A Guide to the New Essentialism.* Chesham, U.K.: Acumen Publishing, 2002.

Engelhardt, H., Tristram. "The Concepts of Health and Disease." In *Evaluation and Explanation in the Medical Sciences*, edited by H. Tristram Engelhardt and Stuart Spicker, 125-142. Dordrechet: Springer, 1975.

Ereshefsky, Marc. *The Poverty of Linnean Heirarchy: A Philosophical Study in Biological Taxonomy.* Cambridge: University of Cambridge Press, 2004.

Fair, Damien, Deepti Bathula, Molly Nikolas, and Joel Nigg. "Distinct Neuropsychological Subgroups in Typically Developing Youth Inform Heterogeneity in Children with ADHD." *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 17 (2012): 6769–6774.

Faraone, Stephen, et al. "Attention-Deficit/Hyperactivity Disorder." *Nature Reviews Disease Primers* 1, no. 15020 (2015): 1-23.

Faraone, Stephen, et al. "Molecular Genetics of Attention-Deficit/Hyperactivity Disorder." *Biological Psychiatry* 57, (2005): 1313–1323.

Faraone, Stephen, and Joseph Biederman. "Neurobiology of Attention Deficit/Hyperactivity Disorder." In *Neurobiology of Mental Illness* (4th edition), edited by Dennis Charney, Joseph Buxbaum, Pamela Sklar, and Eric Nestler, (52763-53422): Oxford: Oxford University Press, (2014), Kindle edition.

Filipek, PA, M Semrud-Clikeman, RJ Steingard, PF Renshaw, DN Kennedy, and J Biederman. "Volumetric MRI analysis Comparing Subjects Having Attention-Deficit Hyperactivity Disorder with Normal Controls." *Neurology* 48, no. 3 (1997): 589-601.

Fliers, Ellen, et al. "Undertreatment of Motor Problems in Children with ADHD." *Child and Adolescent Mental Health* 15, no. 2 (2009): 85–90.

Fukuyama, Francis. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York: Farrar Straus & Giroux, 2002.

Goffman, Erving. "Insanity of Place." In *Relations in Public: Microstudies in Public Order*, by Erving Goffman, 345-390. New York: Basic Book Publications, 1971.

Graham, George. "Being a Mental Disorder." In *Classifying Psychopathology: Mental Kinds and Natural Kinds*, edited by Harold Kincaid and Jacqueline Sullivan, 123-143. Cambridge, Massachusetts: MIT Press, 2014.

Griffiths, Paul. *Squaring the Circle: Natural Kinds with Historical Essences*. edited by R.A. Wilson, 209-228. Cambridge: MIT Press, 1999.

Grob, Gerald. "The Attack of Psychiatric Legitimacy in the 1960s: Rhetoric and Reality." *Journal of the History of the Behavioral Sciences* 47, no. 4 (2011): 398–416.

Hale, Nathan. The Rise and Crisis of Psychoanalysis in the United States. Oxford: Oxford University Press. 1998.

Hard, Heledd, Joaquim Radua, and Tomohiro Nakao. "Meta-analysis of Functional Magnetic Resonance Imaging Studies of Inhibition and Attention in Attention-Deficit/Hyperactivity Disorder: Exploring Task-Specific, Stimulant Medication, and Age Effects." *JAMA Psychiatry* 70, no. 2 (2013): 185-198.

Hardcastle, Valerie, Gray. *The Myth of Pain*. Cambridge, MA: MIT Press, 1999.

Hawthorne, Susan. *Accidental Intolerance.* Oxford: Oxford University Press, 2014.

Hesslow, Germund. "Do We Need a Concept of Disease?" *Theoretical Medicine* 14, no. 1 (March 1993): 1-14.

Hinshaw, Stephen and Richard Scheffler. *The ADHD Explosion: Myths, Medication, Money and Today's Push for Performance*. Oxford: Oxford University Press, 2014.

Holt, John. "Quackery." *New York Review of Books*, August 13, 1970.

Hynd, GW, M. Semrud-Clikeman, AR Lorys, ES Novey, and D Eliopulos. "Brain Morphology in Developmental Dyslexia and Attention Deficit Disorder/Hyperactivity." *Archives of Neurology* 47, no. 8 (1990): 919-926.

Kendler, Kenneth, Peter Zachar, and Carl Craver. "What Kinds of Things are Psychiatric Disorders?" *Psychological Medicine* 41, no. 6 (2011): 1143-50.

Khalidi, Muhammad Ali. *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences.* New York: Cambridge University Press, 2013.

Kim, Jaegwon. "The Supervenience Argument: Motivated, Clarified, and Defended." In *Physicalism or Something Near Enough*, Jaegwon Kim, 32-69. Princeton: Princeton University Press, 2005.

Kincaid, Harold. "DSM Applications to Young Children: Are There Really Bipolar and Depressed Two-Year Olds? " In *Extraordinary Science and Psychiatry: Responses to the Crisis in Mental Health Research*, edited by Jeffrey Poland and Serfie Tekin, 267-292. Cambridge: MIT Press, 2017.

Kingma, Elselijn. "Health and Disease: Social Constructivism as a Combination of Naturalism and Normativism." In *Health, Illness and Disease: Philosophical Essays* edited by Havi Carel and Rachel Cooper, 37-56. New York: Taylor and Francis, 2014.

—. "Naturalist Accounts of Mental Disorder." In *The Oxford Handbook of Philosophy and Psychiatry*, edited by K.W.M Fulford et al., 363-384. Oxford: Oxford University Press, 2013.

—. "What is it to be Healthy?" *Analysis* 67, no. 2 (2007): 133-140.

Kripke, Saul. *Naming and Necessity.* Oxford: Basil Blackwell, 1980.

Lansbergen, Marieke, J. Leon Kenemans, and Herman van Engeland. "Stroop Interference and Attention-Deficit/Hyperactivity Disorder: A Review and Meta-Analysis." *Neuropsychology* 21, no. 2 (2007): 251–262.

Lemoine, Maël. "Defining Disease Beyond Conceptual Analysis." *Theoretical Medicine and Bioethics* 34, (2013): 309-325.

Lewis, David. "Extrinsic Properties." *Philosophical Studies* 44, no. 2 (1983): 197-200.

Locke, John. *An Essay Concerning Human Understanding*, edited by Kenneth P. Winkler. Indianapolis, Indiana: Hackett Publishing, 1996.

Loux, Michael. *Metaphysics: A Contemporary Introduction*. London: Routledge, 1998.

López-Muñoz, Francisco , Cecilio Alamo, Francisco Javier Quintero-Gutiérrez, and Pilar García-García. "A Bibliometric Study of International Scientific Productivity in Attention-Deficit Hyperactivity Disorder covering the period 1980–2005." *European Child & Adolescent Psychiatry* 17, no. 6 (2008): 381–391.

Luman, Margolein, Jaap Oosterlaan, and Joseph A. Sergeant. "The Impact of Reinforcement Contingencies on ADHD: A Review and Theoretical Appraisal." *Clinical Psychology Review* 25 (2005): 183-213.

Mackie, J.L. "Locke's Anticipation of Kripke." *Analysis* 34 (1974): 177-180.

Magnus, P.D. "NK≠HPC." *Philosophical Quarterly* 64, no. 256 (2014): 471-477.

—. *Scientific Enquiry and Natural Kinds: From Planets to Mallards.* New York: Palgrave Macmillan, 2012.

Martinussen, Rhonda, Jill Hayden, Sheilah Hogg-Johnson, and Rosemary Tannock. "A Meta-Analysis of Working Memory Impairments in Children with Attention-Deficit/Hyperactivity Disorder." *Journal of the American Academy of Child and Adolescent Psychiatry* 44, no. 4 (2005): 377-384.

Mayes, Rick, Catherine Bagwell, and Jennifer Erkulwater. *Medicating Children: ADHD and Pediatric Mental Health.* Cambridge, Massachusetts: Harvard University Press, 2009.

Murphy, Dominic. *Psychiatry in the Scientific Image.* Cambridge: MIT Press, 2006.

Murphy, Dominic, and Robert L. Woolfolk. "The Harmful Dysfunction Analysis of Mental Disorder." *Philosophy, Psychiatry, & Psychology* 7, no. 4 (2000): 241-252.

Nigg, Joel. "Is ADHD a Dishibitory Disorder?" *Psychological Bulletin* 127, no. 5 (2001): 571-598.

—. *What Causes ADHD: Understanding What Goes Wrong and Why.* New York: Guilford Press, 2006.

Nigg, Joel, and Molly Nikolas. "Attention-Deficit/Hyperactivity Disorder." In *Child and Adolescent Psychopathology*, edited by Theodore Beauchaine and Stephen Hinshaw, 301-334. Hoboken: John Wiley & Sons, 2008.

Nigg, Joel, Erik Willcut, Alysa Doyle, and Edmund Songua-Barke. "Causal Heterogeneity in Attention-Deficit/ Hyperactivity Disorder: Do We Need Neuropsychologically Impaired Subtypes?" *Biological Psychiatry* 57, no. 11 (2005): 1224 –1230.

Oderberg, David. *Real Essentialism*. New York: Routledge, 2007.

Page, Sam. "Mind-Independence Disambiguated: Separating the Meat from the Straw in the Realism/Anti-Realism Debate." *Ratio* 19, no. 3 (2006): 321-335.

Pasnau, Robert. *Metaphysical Themes: 1274-1671*. Oxford: Oxford University Press, 2011.

Plichta, Michael, et al. "Neural Hyporesponsiveness and Hyperresponsiveness During Immediate and Delayed Reward Processing in Adult Attention-Deficit/Hyperactivity Disorder." *Biological Psychiatry* 65, no. 1 (2009): 7-14.

Poland, Jeffrey, and Barbara Van Eckardt, "Mapping the Domain of Mental Illness." In *The Oxford Handbook of Philosophy and Psychiatry*, edited by K.W.M. Fulford et al., 735-752. Oxford: Oxford University Press, 2013.

Putnam, Hilary. The Meaning of 'Meaning. Vol. 2, in *Mind, Language and Reality: Philosophical Papers*, by Hilary Putnam, 215-271. Cambridge: Cambridge University Press, 1975.

Ramsey, J.L. "Molecular Shape, Reduction, Explanation, and Approximate Concepts." *Synthese* 111 (1997): 233-251.

Rapport, M.D., R. M. Alderson, M.J. Kofler, D.E. Sarver, J. Bolden, and V. Sims. "Working Memory Deficit in Boys with Attention Deficit Disorder: the Contribution of the Central Executive and Subsystem Processes." *Journal of Abnormal Child Psychology* 36 (2008): 825-837.

Roberts, Walter, Richard Milich, and Russell Barkley . "Primary Symptoms, Diagnostic Criteria, Subtyping, and Prevalence." In *Attention-Deficit Hyperactivity Disorder*, edited by Russell Barkley, 51-80. New York: Guilford Press, 2015.

Ross, Dorthea, and Sheila Ross. *Hyperactivity: Research, Theory, and Action.* New York: John Wiley and Sons, 1976.

Samuels, Richard. "Delusion as a Natural Kind." In *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, edited by Matthew R. Broome and Lisa Bortolotti, 49-79. New York: Oxford University Press, 2009.

Sagvolden, Terje, Espen Borga Johansen, Heidi Aase, and Vivienne Ann Russell. "A Dynamic Developmental Theory of Attention-Deficit/Hyperactivity Disorder (ADHD): Predominantly Hyperactive/Impulsive and Combined Subtypes." *Behavioral and Brain Sciences* 28, no. 3, (2005): 397-419.

Sergeant, Joseph. "The Cognitive-Energetic Model: an Empirical Approach to Attention-Deficit Hyperactivity Disorder." *Neuroscience and Biobehavioral Reviews* no. 1, (2000): 7-12

Saul, Richard. *ADHD Does Not Exist: The Truth About Attention Deficit and Hyperactivity Disorder.* New York: Harper, 2013.

Scheff, Thomas J. "Schizophrenia as Ideology." *Schizophrenia Bulletin* (Institute of Mental Health) 1, no. 2 (Fall 1970): 15-19.

Sedgwick, Peter. "Goffman's Anti-Psychiatry." *Salmagundi*, no. 26 (1974): 26-51.

—. "Mental Illness Is Illness." *Salmagundi*, no. 20 (1972): 196-224.

—. *Spiked*. September 21, 2012. http://www.spiked-online.com/newsite/article/12897 (accessed August 5, 2015).

Seidman, Larry, Eve Valera, and Nikos Makris. "Structural Brain Imaging of Attention-Deficit/ Hyperactivity Disorder." *Biological Psychiatry* 57 (2005): 1263-1272.

Schrag, Peter and Diane Divoky. *The Myth of the Hyperactive Child: And Other Means of Child Control*. New York: Pantheon Books, 1975.

Scully, J.H., et al. "Psychiatric Education after World War II." In *American Psychiatry After World War II*, edited by R.W. Menninger and J.C. Nemiah, 124-151. Washington D.C.: American Psychiatric Press, 2000.

Shaw, Philip, Meaghan Malek, Bethany Watson, Wendy Sharp, Alan Evans, and Deanna Greenstein. "Development of Cortical Surface Area and Gyrification in Attention-Deficit/Hyperactivity Disorder." *Biological Psychiatry* 72, no. 3 (2012): 191-197.

Shaw, Philip, et al. "Trajectories of Cerebral Cortical Development in Childhood and Adolescence and Adult Attention-Deficit/Hyperactivity Disorder." *Biological Psychiatry* 74 (2013): 599–606.

Shorter, Edward. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac.* New York: John Wiley and Sons, 1997.

Skoglund, Pontus, Erik Ersmark, Eleftheria Palkopoulou, and Love Dalén. "Ancient Wolf Genome Reveals an Early Divergence of Domestic Dog Ancestors and Admixture into High-Latitude Breeds." *Current Biology* 25, no. 11 (2015): 1515–1519.

Slater, Matthew. "Natural Kindness." *British Journal for the Philosophy of Science* 66, no. 2 (2015): 375-411.

Solanto, Mary, et al. "The Ecological Validity of Delay Aversion and Response Inhibition as Measures of Impulsivity in AD/HD: A Supplement to the NIMH Multimodal Treatment Study of AD/HD." *Journal of Abnormal Psychology* 29, no. 3 (2001): 215–228.

Sonuga-Barke, Edmund J.S. *Attention Deficit/Hyperactivity Disorder: Towards a Developmental Synthesis* . Vol. II, in *The Oxford Handbook of Developmental Psychology: Self and Other*, 549-578. Oxford: Oxford University Press, 2013.

—. "Causal Models of Attention-Deficit/Hyperactivity Disorder: From Common Simple Deficits to Multiple Developmental Pathways." *Biological Psychiatry* 57, no. 11 (2005): 1231–1238.

Sonuga-Barke, Edmund J.S., Jan De Houwer, Karen De Ruiter, Michal Ajzenstzen, and Sarah Holland. "AD/HD and the capture of attention by briefly exposed delay-related cues: evidence from a conditioning paradigm." *Journal of Child Psychology and Psychiatry* 45, no. 2 (2004): 274-283.

Sonuga-Barke, Edmund, Joseph Sergeant, Joel Nigg, and Erik Willcutt. "Executive Dysfunction and Delay Aversion in Attention Deficit Hyperactivity Disorder: Nosologic and Diagnostic Implications." *Child and Adolescent Psychiatric Clinics of North America* 17, no. 2 (2008): 367-384.

Sonuga-Barke, Edmund J.S., Paraskevi Bitsakou, and Margaret Thompson. "Beyond the Dual Pathway Model: Evidence for the Dissociation of Timing, Inhibitory, and Delay-Related Impairments in Attention-Deficit/Hyperactivity Disorder." *Journal of the American Academy of Child and Adolescent Psychiatry* 49, issue 4 (April 2010): 345–355.

Sowell, Thomas. "Drugging Children." *Townhall.com*, August 23, 2001. https://townhall.com/columnists/thomassowell/2001/08/23/drugging-children-n864940, accessed April 1st, 2017.

Spitzer, Robert and Donald Klein. *Critical Issues in Psychiatric Diagnosis*. New York: Raven Press, 1978.

Spitzer, Robert and Jean Endicott. "Medical and Mental Disorder: Proposed Definition and Criteria." In *Critical Issues in Psychiatric Diagnosis*, edited by Robert Spitzer and Donald Klein, 15-40. New York: Raven Press, 1978.

Squire, Larry R. "The Legacy of Patient H.M. for Neuroscience." *Neuron* 61, no. 1, (2009): 6-9.

Ströhle, A., et al. "Reward Anticipation and Outcomes in Adult Males with Attention-Deficit/Hyperactivity Disorder." *Neuroimage* 39, no. 3 (2008): 966-972.

Sullivan, Jacqueline. "Experimentation in Cognitive Neuroscience and Cognitive Neurobiology." In *Handbook of Neuroethics*, edited by Jens Clausen and Neil Levy, 31-47. Dordrecht: Springer, 2015.

Szasz, Thomas. *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct*. New York: Harper and Row, 1962.

Tabb, Kathryn. "Psychiatric Progress and the Assumption of Diagnostic Discrimination." *Philosophy of Science* 82, (2015): 1047–1058.

Taylor, Eric, and Edmund Sonuga-Barke. "Disorders of Attention and Activity." In *Rutter's Child and Adolescent Psychiatry*, edited by Michael Rutter, Dorothy Bishop, Daniel Pine, Jim Stevenson, Eric Taylor and Anita Tharpar, 521-542. Oxford: Blackwell Publishing, 2008.

Timimi, Sami. "ADHD is Best Understood as a Cultural Construct." *British Journal of Psychiatry* 184, (2004): 8-9.

Toblak, Maggie, Colleen Dockstadera, and Rosemary Tannock. "Temporal Information Processing in ADHD: Findings to Date and New Methods." *Journal of Neuroscience Methods*, 2006: 15–29.

Tomblin, J. Bruce, and Kathyrn Mueller. "How Can Comorbidity With Attention-Deficit/Hyperactivity Disorder Aid Understanding of Language and Speech Disorders?" *Topics in Language Disorders* 32, no. 3 (2012): 198–206.

Tripp, Gail, and Jeff Wickens. "Research review: dopamine transfer deficit: a neurobiological theory of altered reinforcement mechanisms in ADHD." *Journal of Child Psychology and Psychiatry* 49, (2008): 691–704.

Tsou, Jonathan. "Depression and Suicide are Natural Kinds: Implications for Physician-Assisted Suicide." *International Journal of Law and Psychiatry* 36, 5-6, (2010): 461-470.

van Rooij, Daan, et al. "Distinguishing Adolescents With ADHD From Their Unaffected Siblings and Healthy Comparison Subjects by Neural Activation Patterns During Response Inhibition." *American Journal of Psychiatry* 172, no. 7 (2015): 674-683.

Vande Wall, Holly. "Why Water Is Not H2O, and Other Critiques of Essentialist Ontology from the Philosophy of Chemistry." *Philosophy of Science* 74, no. 5 (2007): 906-919.

Venn, John. "The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability with an Especial Reference to its

Logical Bearings and its Application to Moral and Social Science, and to Statistics." London: Macmillan, 1866.

Volkow, Nora D., et al. "Evaluating Dopamine Reward Pathway in ADHD Clinical Implications." *Journal of the American Medical Association* 302, no. 10 (2009): 1084-1091.

Wakefield, Jerome C. "Evolutionary versus prototype analyses of the concept of disorder." *Journal of Abnormal Psychology* 108, no. 3 (1999): 374-299.

—. "Taking Disorder Seriously A Critique of Psychiatric Criteria for Mental Disorders from the Harmful-Dysfunction Perspective." In *Contemporary Directions in Psychopathology: Scientific Foundations of the DSM-V and ICD-11*, edited by Theodore Millon, Robert F. Krueger and Erik Simonsen, 275-300. New York: Guilford Press, 2010.

—."The Concept of a Mental Disorder: On the Boundary Between Biological Facts and Social Values." *American Psychologist*, March 1992: 373-388.

—. "The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis." *World Psychiatry* 6, no. 3 (2007): 149-156.

Weiskopf, Daniel. "Models and Mechanisms in Psychological Explanation." *Synthese* 183, no. 3 (2011): 313-338

Wilbertz, Gregor, Amalie Trueg, Edmund J.S. Sonuga-Barke, Jens Blechert, Alexandra Philipsen , and Ludger van Elst. "Neural and Psychophysiological Markers of Delay Aversion in Attention-Deficit Hyperactivity Disorder." *Journal of Abnormal Psychology* 122, no. 2 (2013): 566 –572.

Wilkerson, T.E. "Natural Kinds." *Royal Institute of Philosophy* 63, no. 203 (January 1988): 29-42.

—. *Natural Kinds.* Aldershot: Avebury, 1995.

Will, George. "Boys Will Be Boys." *Washington Post*, December 2, 1999.

Willcutt, Erik, Alysa Doyle, Joel Nigg, Stephen Faraone, and Bruce Pennington. "Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review." *Biological Psychiatry* 57 (2005): 1336–1346.

Willcutt, Erik, et al. "Validity of DSM-IV Attention Deficit/Hyperactivity Disorder Symptom Dimensions and Subtypes." *Journal of Abnormal Psychology*, 2012: 991-1010.

Wilson, Robert A., Matthew Barker, and Ingo Brigandt. "When Traditional Essentialism Fails." *Philosophical Topics* 35 (2007): 189-215.

Wimsatt, William. "Aggregativity: Reductive Heuristics for Finding Emergence." *Philosophy of Science* 64, (1997): S372-S384