Challenging the Sovereignty of Science in Classifying Mental Disorders

By

Stephen Rowe

Submitted to the Department of Philosophy, Central European University

In partial fulfilment of the requirements for the degree of Master of Art

Supervisor: Maria Kronfeldner

Vienna, Austria

2024

Acknowledgements

I am extremely grateful to my wonderful partner Brylea Hollinshead with whom many very passionate discussions on the topic of this thesis motivated me to keep trying to produce a more persuasive argument.

I would also like to express my thanks to my supervisor Maria Kronfeldner whose astonishing philosophical prowess and eye for precision vastly improved the quality of this thesis. Working with you has been a pleasure.

To the 2-year MA cohort I shared my time with at CEU (particularly Darren and Levan for their friendship in and outside of the classroom), thank you for all the animated discussions that have contributed to this thesis in incalculable ways.

Lastly, I would also like to express my extreme gratitude to my family. Were it not for the support you have provided me or the grit, courage, workaholism, and can-do attitude you have raised me to have I would not have been able to complete this thesis.

Copyright Notice

Part of this thesis appeared in a course essay titled *Can We Be Realist About Mental Disorders Like We Can Be About Universals?* for Tim Crane's class "Metaphysics of Mind" at Central European University in 2024.

Table of Contents

Acknowledgements i

Introduction 1

Section 1: The Current Approach to Classification and Three Alternatives 6

1.1. The Current Approach 6
1.2. Naturalistic Approaches 8
1.2.1. A Brief Note on Natural Kinds 9
1.2.2. A Naturalistic Account of Mental Disorders 11
1.3. Hybrid Approaches 15
1.3.1 Wakefield's Hybrid Approach 15
1.3.2. Tsou's Hybrid Approach 17
1.4. Pragmatic Approaches 17
1.4.1. Kukla's Pragmatic Approach 19
1.5. Section Summary 25

Section 2: Pragmatic Approaches are Preferable 26

2.1. Naturalist Response 30
2.1.1. Naturalism Protects Against Overmedicalisation, Pragmatism Promotes it 31
2.2. Hybrid Response 39
2.2.1. Wakefield's Response 40
2.2.2. Tsou's Response 42
2.2.2. Summary of Section 44

Conclusion 45

References 47

Introduction

Classification is a ubiquitous human activity. We do it by organising things into groups on the basis of their shared properties for various reasons. For instance, I classify my emails into important and junk so I can spend my time productively; reviewer two's comments on my paper into idiotic and sensible so I know which comments to spend time responding to; friends from enemies so I know who to kill first during a purge; and $C_{25}H_5OH$ from H_2O for bad days and good days.

Classification also occurs on a much larger scale than what you or I might classify for our own interests — groups of people with similar ends, like scientists who want to know how the world is, get together and decide collectively on what is the right (truth-conductive) way to classify different things. For instance, animals like horses and cows are classified according to evolutionary heritage into domain, kingdom, family, and species; plants are similarly classified; elements like gold and helium are classified according to their atomic structure and put into the periodic table; doctors classify people into healthy and unhealthy based on biomarkers of diseases (e.g., blood sugar levels for diabetes); and 970 million people are classified as having a mental disorder on the basis of their having a syndrome (cluster of symptoms) that is harmful to their psychological well-being (WHO, 2022). As there are approximately 8 billion people currently alive, this means that almost one in every eight people are classified as having a mental disorder.

Such a large estimate has led to many people claiming we are in a mental health crisis. But not everyone agrees we are in a crisis. Some argue that it only appears we are in a mental health crisis because we are unjustifiably classifying normal behaviours as disorders. For instance,

Horwitz and Wakefield (2007) argue we are wrongly labelling grief after losing a loved one as a disorder, and Charland (2006) argues we are wrongfully treating personality traits that we normatively disapprove of as medical conditions. Such criticisms have been so widely received that it is not uncommon to see headlines in popular press like this one: "The definition of mental health has been widened so much that it's now almost meaningless" (Gill, 2023).[1]

Criticisms about how mental disorders are classified have pushed many clinicians and philosophers into thinking that the whole reputation of psychiatry is on the line. For instance, Wakefield (2007) claims that the "credibility and even the coherence of psychiatry as a medical discipline" (p.149) rests on there being a single true answer as to what counts as a disorder: what psychiatry needs to defend its credibility, he claims, is a rigorous concept of mental disorder (i.e., one with one set of necessary and sufficient conditions that can be used in all contexts) that is based on empirical research in the natural sciences (e.g., biology, chemistry), like people assume medicine has for physiological diseases.

My aim here is to add to the debate on how we should decide whether a phenomenon is a mental disorder by questioning the spirit of our time — scientism, or the belief that science has the answer to everything. It is in the idea that disorder classification does not make adequate use of research in the natural sciences that I see people taking issue with psychiatric classification. What I want to know is whether the natural sciences should be sovereign in disorder classification – i.e., can the natural sciences provide us with the appropriate necessary and sufficient conditions for deciding whether something is a mental disorder? If not, can

[1] While the title says 'mental health' rather than 'mental disorder' the opinion piece is about how the definition of mental disorder is too inclusive.

science give us a necessary condition? What about a sufficient condition? Or can it neither provide a necessary nor sufficient condition?

Asking whether the natural sciences can provide us with necessary and or sufficient conditions is ambiguous. The debate about the relevancy of science will not get very far until the question is made more precise. One seemingly plausible option to specify the research question is in terms of causality — science is, in large part, the business of discovering causal relations in the world. Perhaps the question could then be: "is it a necessary condition for x to be a mental disorder that the causes of its symptoms are, at least in principle, something that the method of science can discover". This, unfortunately, is not very helpful: no one, even those who reject that mental disorders exist in any analogous sense to physiological diseases like tuberculosis, deny that symptoms like low mood and distress exist and that their causes are discoverable by the scientific method. If there is to be a debate for and against the sovereignty of science, the positions need to be stated in a way that gives each side something to bite onto — a point of controversy.

A more fruitful way of specifying the question is thus to look where the controversy exists. Two points of controversy stick out to me. The first is in whether one must be scientifically justified in grouping together the set of properties that constitute a mental disorder. Put differently, is it the case that depression needs to be a natural kind like horse, tuberculosis, and lithium are? I will aim to defend the claim that no, a mental disorder need not be a natural kind as I understand the term (see section 1.2.1).

The other point of controversy is whether the notion of biological dysfunction is necessary to classify mental disorders. The use of the term biological dysfunction in definitions of mental

disorders was at one point so pervasive that stating that a disorder is a biological dysfunction was an unhelpful tautology — one, as Wakefield (1992. p.235) claimed, we learn nothing from. In response to this, several different attempts were made to clarify the notion. Some (e.g., Boorse, 1977) followed Cummings (1975) and argued that we should have a causal notion of biological dysfunction, others like Wakefield (1992), followed Neander (1991) and argued that we should have an evolutionary notion of biological dysfunction, and others (e.g., Tsou, 2021, Kukla, 2022) have argued that we do not need the notion at all. I will defend the argument that not all mental disorders are biological dysfunctions in either of these two senses.

The plan is as follows: in section 1 I will first outline how mental disorders are currently classified and then I will describe three approaches that look to challenge the way we currently classify them: first, a naturalistic one which makes science sovereign in classification — i.e., it claims that the natural sciences alone are necessary and sufficient for disorder classification. The second approach is a hybrid one, which argues that the natural sciences are necessary, but not sufficient — a further normative condition being required. The final approach is a pragmatic one which argues that the conditions change depending on the contexts. In section 2 I will argue for the pragmatic approach over the other two by extending an argument from Kukla (2022). I will then defend my extended version of Kukla's argument from two objections: (a) one from the naturalistic approach which argues it is preferable because it prevents overmedicalisation whereas pragmatism promotes it, and (b) an argument from the hybrid approaches that Kukla fails to show that there can be no one set of necessary and sufficient conditions that work for all contexts. While these are strong objections, they can be responded to. Ultimately, I conclude that there is good reason to think that Kukla is right — depending on the context, the conditions for classifying something as a mental disorder change.

Disclaimers

In an effort to avoid confusion down the line, I want to state what it is I am not doing. I am not investigating how we should demarcate one disorder from another — how to distinguish between anxiety and depression, for instance. Nor am I investigating the specific method or criteria we should use to identify a particular condition — how to decide whether a phenomenon is a case of depression, for instance. What I am interested in is how we should decide whether phenomena like hysteria, schizophrenia, depression, and homosexuality should be classified as mental disorders. Deciding whether a phenomenon is a disorder is distinct from deciding how to identify a phenomenon: you can have a perfectly clear criteria for deciding whether a phenomenon is hysteria or depression but still wonder whether this phenomenon is a disorder.

I am also agnostic about whether the arguments I make about classifying certain phenomena as mental disorders (such as Major Depressive Disorder (MDD) and Speech Pragmatic Communication Disorder (SPCD), two examples that I will use) as mental disorders apply to classifying physiological diseases. While in everyday use disease and disorder are often used interchangeably, there is debate about whether they are synonyms. Murphy (2020) gives two reasons for why someone might distinguish between physiological diseases and mental disorders. The first is that mental disorders are typically understood as a cluster of symptoms. Schizophrenia, for instance, is understood as a combination of positive symptoms, like hallucinations and delusions, and negative symptoms, like disconnectedness from one's feelings and emotions. In contrast, physiological diseases are not typically understood as clusters of symptoms, but rather as destructive processes realised in bodily tissue. Secondly, it is commonly assumed that diseases, but not disorders, can be asymptomatic: we assume that

someone can have covid even if they do not have a sore throat or any other symptoms, but we do not assume that someone has depression if they do not have a low mood or any other symptoms. As such, I will not be treating diseases and disorders as synonyms. Nor is it within the scope of this project to discuss to what extent the arguments apply to classifying diseases.

Section 1: The Current Approach to Classification and Three Alternatives

In this section I first outline the current 'atheoretical' approach to classifying mental disorders and then discuss three alternative approaches. There are, of course, many more approaches than the ones I discuss here. For instance, eliminativism which argues that we should get rid of the concept of mental disorder (Szasz, 1960), and radical social constructivism which argues that mental disorders are necessarily whatever medicine labels as a disorder (Engelhardt, 1974). The approaches I have chosen to discuss were picked because the represent a mixture of approaches that dominate the discussion – Boorse's (1977) naturalism and Wakefield's (1992) hybrid approach – and recent innovations – Tsou's (2021) variation on the hybrid approach and Kukla's (2022) radical pragmatism.

1.1. The Current Approach

The current most influential approach to classifying mental disorders comes from the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5). The current approach was developed in response to a huge problem psychiatry faced in the 1970s. In a review of this problem, Aboraya et al (2006) note that before WW2 psychoanalysis was by far the dominant approach to psychiatry, and its view was that a diagnosis was largely irrelevant to therapy and treatment decisions. However, they show that as this began to change — as the number of diagnoses greatly expanded when the DSM-2 published in 1968, and as the

popularity of giving a diagnosis also increased – diagnoses become highly unreliable. The likelihood of one patient receiving the same diagnosis for their condition by two different clinicians was only around 50%. Aboraya et al (2006) argue that many factors influenced the lack of reliability, but prominent among them was the lack of consensus on what mental disorders are (as opposed to a mental phenomenon that are not). This problem, and what the authors of the DSM-3 did in response to it, led to the current 'atheoretical' (i.e., non-etiological) approach to classifying mental disorders.

In short, what Spitzer — the chief architect of the DSM-3 — did in response to this problem was to remove any particular theory about the etiology (origin) of mental disorders — like the Oedipal complex that is used in Freudian psychoanalysis to characterise disorders like depression — and approach the question of what should count as a mental disorder in an atheoretical way. He reasoned that, with the etiology of mental disorders disputed, the best way to decide what counts as a disorder would be to stay neutral with respect to any etiological assumptions (Spitzer & Williams, 1983). With no etiological theory to guide a conceptualisation of mental disorders, the approach Spitzer took to derive a definition of disorder was conceptual analysis, in the form of the so-called method of cases. Wakefield (1992) describes Spitzer's approach as one where a proposed definition of mental disorder is tested against paradigm cases of what falls under the concept, and against paradigm cases that should not fall under the concept. As typical in philosophical method (also used in this thesis), the idea was to develop a definition of mental disorder that includes as disorders what we paradigmatically think are disorders, and excludes as disorders what we paradigmatically think are not disorders. After revising many potential definitions, what he arrived at was a definition of mental disorder that has remained largely unchanged. The current definition of mental disorder in the DSM-5 states that:

> A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above (APA, 2013. p.20).

The DSM's definition of mental disorder has been the target of many criticisms. People have argued that, while the DSM has increased reliability, its groupings lack validity (Hyman, 2010). People have also launched concerns about the irrelevance of etiology (Thyer, 2015), its western bias (Murphy, 2020), its pathologisation of normal behaviour (Horwitz & Wakefield, 2007), and its focus on dysfunction (Tsou, 2021).

I am not going to add to this list of complaints or assess these arguments. Rather, I want to look at the proposed alternatives to what the DSM's definition states is (or is not) a mental disorder. In the following, I will describe three approaches: first, a naturalistic one which claims that the natural sciences alone are necessary and sufficient for disorder classification. The second approach is a hybrid one, which argues that the natural sciences are necessary, but not sufficient — a further normative condition being required. The final approach is a pragmatic one which argues that there is not one set of necessary and sufficient conditions that can be used in all contexts. After describing the three approaches, I will argue that the pragmatic one is preferable over the other two.

1.2. Naturalistic Approaches

The naturalist argues that the natural sciences can provide a way to classify mental disorders. They argue the world is constructed in such a way that certain properties are naturally grouped

together — there are "natural kinds" – and that these natural groupings provide us with a way to separate disorders from non-disorders. Before moving on to the specifics of this approach, I will explain what a natural kind is.

1.2.1. A Brief Note on Natural Kinds

Classifying is a human activity. There is no question that it is people who group things together. So, in some sense, all classifications are a product of human activity. However, it is a very popular thought that some classifications better reflect the way the world is structured — they 'carve nature at its joints'. Such classifications are called natural kinds. I find the analogy to cutting a chicken useful (adapted from Zachar (2014)): There are many ways one can carve up a whole, uncooked chicken. Some of these are much more difficult than others. If one is skilful and precise enough, a knife can disjoint the legs from the torso with minimal effort. The reason for this is that there is something about the chicken that makes carving it up in such a way easy — namely, that the chicken is structured in such a way that it has certain joints which are easy places to cut.

To those who believe in natural kinds, the world is much like the chicken — it has certain 'joints' which allow us to separate its parts in certain ways. Obviously, the world does not literally have joints, but it does have structure — some properties have strong causal relations to other properties, resulting in sets of properties that naturally occur together in things, which can therefore be grouped together according to these sets. This kind of structure, they argue, is what allows us to discover natural groupings of individual things in the world (i.e., natural kinds).

The contrast to natural kinds is non-natural kinds. Non-natural kinds are groups that do not accurately reflect the causal structure of the world, rather, they are categories we construct

based on certain interests. For example, ‘books I like’ or the items I group together on my to-do list. Considering that classifications is a human activity, the distinction between natural and non-natural kinds is best thought of as a scale, rather than one with sharp boundaries. At one end of the scale there are groups of individuals with properties that have an extremely strong connection — the properties of one member will (almost) always be there in the other members. The groups of properties at this end of the scale are paradigmatic natural kinds. For instance, take the category gold. Gold things have the property of having 79 protons, melting at 1,064 degrees Celsius, and conducting electricity. These three properties occur together all the time in each member in the category, and not because of some accident, rather, an individual having the property of 79 protons explains why it melts at 1,064 degrees Celsius and conducts electricity. Hence gold is a paradigmatic natural kind. At the other end of the scale there are kinds in which the properties shared by members have a very weak relation – the properties of one member are unlikely to be shared by any other member. For instance, take the set of things that constitute my to do list. The members of this non-natural kind are constantly changing and they do not have much similarity to each other beyond the fact that I need to do them. Knowing the properties of one member in the list will be of no help in predicting the properties of other members in the list. Because one can make inferences about further properties of a gold thing based on knowing that it is a member of gold (e.g., that is has 79 protons, will melt at 1,056 degrees Celsius, etc.) but cannot make inferences about the properties of something that is on my to do list (beyond the fact that I need to do it) based on knowing that it is a member of my to do list, science has a keen interest in figuring out which categories are likely to be natural one's and which ones are not.

Because scientists are so interested in discovering natural kinds, a great deal of work has gone into figuring out how to tell whether a category is a natural one or a non-natural one. One very

influential proposal by Boyd (1999), that is also popular in psychiatry (see e.g., Kendler et al., (2011) and Tsou (2016;2021)), argues that natural kinds, but not non-natural ones, are Homeostatic Property Clusters (HPC). HPC are clusters of properties that reliably co-occur in space and time due to homeostatic mechanisms that cause and sustain them. The presence of underlying mechanisms that cause the properties to cluster together further ensures that HPC are fit for use in scientific explanation — they can be used to make inductive inferences. Boyd (1999) uses the categories developed in biological taxonomy as his main example. Take the kind 'horse'. The members of the kind horse reliably share many properties, so, by knowing that something is a member of horse I can know many of its other properties. Further, the best explanation for why the members of horse share many properties is because the kind is picking up on a structure or mechanism in the world that causes the properties of one member to reliably co-occur in other members — namely, shared evolutionary heritage. This view of natural kinds as HPC has gained popularity within psychiatry because it fits well with the idea that mental disorders are syndromes (clusters of symptoms).

1.2.2. A Naturalistic Account of Mental Disorders

Applying this to mental disorders then, the thought is that, in order to accurately classify mental disorders, we need to identify groups of properties that reliably occur together in phenomena because of some mechanism that maintains them. Further, the natural sciences are all we need to do this — it can provide the necessary and sufficient conditions for classifying mental disorders. A classic and still discussed theory of mental disorders that makes use of such an idea is Christopher Boorse's (1977; 2014) theory of mental disorders as statistically abnormal losses of efficiency in biological function.

Boorse's (1977) theory of mental disorders consists of three parts that build on each other: (1) a reference class (i.e., a "natural class of organisms" (p.555) or a natural kind (Tsou, 2021.

p.20)), (2) a causal definition of biological function, and (3) a definition of health made up from 1 and 2. Taking these three parts together, you get the definition of mental disorder as the absence of psychological health. I will explain 2 first as it helps explain 1, which will help to see how you get 3.

Talk of biological function is popular in biology. Popular examples are that the function of the heart is to pump blood, the eye is to see, and the peacock's tail is to attract a mate. What determines the biological function of something (or what makes it the case that the function of X is to Y rather than to Z) is a matter of serious debate. There are two main contenders for how to define biological function in the literature – one is a *causal role account* (Cummins, 1975), the other is a *selected effects account* (Neander, 1991). A causal role account determines the biological function of something by its causal contribution to a larger system, whereas a selected effects account determines it by what it was selected by evolution for. Boorse takes a causal approach to defining function, whereas Wakefield (1992) takes an evolutionary approach (see section 1.3.1). Boorse takes the larger system to be the organism (e.g., a human being), and contribution is understood with reference to the organisms ongoing survival. Crucially, as Tsou (2021) notes, this is not an evolutionary notion – the role that an internal part or process has historically played in survival and reproduction is irrelevant to determining its function now. What matters is the role it currently plays in the organism's survival. Thus, for Boorse, the function of X (say, a peacock's tail) is Y (to attract a mate) rather than Z (make whistling sounds in the wind) because the statistically typical way that X (the tail) still contributes to survival and reproduction is through its effect Y (attracting mates) rather than Z (whistling in the wind). Similarly, the function of a human's heart is to pump blood rather than make sound because it is the effect of pumping blood that still contributes to the human's survival. Mental functions are understood the same way — for example, the function of the

mental mechanisms underlying a fear response is to deter us from danger because it reliably has this survival conducive effect. Function is thus understood as the role a psychological process currently reliably plays in reproduction and survival (broadly understood). What makes a biological function normal or 'natural' is its reference to a reference class, or a "natural class of organisms" (Boorse, 1977. p.555).

Boorse's notion of mental disorder rests on the simple premise that "the normal is the natural" (p.554), where normal is understood in statistical terms. The intuition being that abnormal or disordered psychology is what is a statistically significant deviation in functional performance from normal psychology. However, one must be careful with this. Humans, pigeons, and rats can all have mental processes that aid in sleep and let's assume they all perform the same function — to make sure that the organism rests and so does not die from exhaustion. Humans sleep less than pigeons, and rats sleep even more than pigeons. It does not make sense, however, to say that the mental processes that regulate sleep in humans are dysfunctional just because they cause humans to sleep significantly less than rats do. Rather, the performance of a function in an organism must be understood in relation to what is typical for members of its natural kind (i.e., reference class). For instance, let's assume that adult humans make up a reference class, or constitute a natural kind in the HPC sense. Further, let's assume that this reference class has a psychological process that typically causes them to sleep for 8 hours a night. Such a process functions normally in adults to the extent that it causes them to sleep for 8 hours a night and is dysfunctional to the extent that it does not.

Part 3, the definition of health, is made up of 1 and 2. Mental health just is the readiness of a part or process to perform its psychological function with at least efficiency typical of its reference class. The proviso of 'at least typical efficiency' is important because, while

statistical deviation from normal functioning is a necessary condition, it is not sufficient. Not all deviations are a loss in performance — a person with far greater working memory is not a deviation anyone would take to be constitutive of a disorder. Rather, the deviation must represent a loss of efficiency (Boorse, 1977. p.559). With the definition of mental health in place, its contrast of mental disorder as the absence of mental health is clear: a mental disorder is a phenomenon which impairs a psychological function below the typical efficiency of the reference class. To continue the earlier example, an adult human with a condition that impairs the psychological process underlying sleep, causing them to sleep for less than the statistically typical 8 hours, has a mental disorder.

In summary, naturalistic approaches argue that the natural sciences can provide us with the necessary and sufficient criteria to classify mental disorders. The world is structured in such a way that some properties naturally group together in a reference class — i.e., there are natural kinds. By carefully observing the world we can discover these kinds and adapt our categories to them, i.e., make use of them to develop a procedure for identifying mental disorders. Boorse (1977) is one of the main proponents of this approach. According to him, we can draw on what is a typical functional performance of a natural class of organisms to devise a set of necessary and sufficient criteria for identifying mental disorders. In his view, someone has a mental disorder when a psychological process fails to perform its psychological function with the efficiency typical of members of their natural kind. Importantly, the natural sciences give us all we need to develop this procedure for identifying mental disorders. As Boorse (1977) puts it "the judgment that something is a disease [or disorder] is a theoretical judgment that neither entails nor is entailed by any therapeutic judgment about people's need for medical treatment" (p. 544).

1.3. Hybrid Approaches

The basic idea of hybrid approaches is that there are two necessary conditions that are jointly sufficient for classifying something as a mental disorder. One necessary condition is taken from the naturalists who follow Boorse in some sense — for instance, mental disorders are necessarily biological dysfunctions (Wakefield, 1992) or natural kinds (Tsou, 2021). The other necessary condition is normative — mental disorders cause the individual harm. The most popular hybrid approach is Wakefield's (1992), and the latest attempt to argue for one comes from Tsou (2021). I will discuss each in turn.

1.3.1 Wakefield's Hybrid Approach

Wakefield (1992; 1992b) argues that mental disorders are harmful dysfunctions. For Wakefield, for a phenomenon to be classified as a mental disorder it must (1) be a biological dysfunction, and (2) cause significant harm to the person. Wakefield's notion of dysfunction is different from Boorse's. Wakefield clearly uses a selected effects account, whereby a biological function is the beneficial (or adaptive) effect of an internal part or process that it was selected for by natural selection. Wakefield (1992b) argues this is "the most viable approach" (p.236) to understanding biological function as natural selection provides the only known causal explanation we can provide for the existence of the effects of a biological mechanism. The key difference between Boorse's and Wakefield's notion of function is thus that for Wakefield function is understood in relation to the beneficial effects a part or process has to the survival of an organism across its evolutionary history, whereas for Boorse what matters is the effect a part or process has to the survival of the organism currently.[2]

[2] This difference has been a point of contention for Wakefield's view, with the objection that many mental process (e.g., language) are likely to be exaptations (or traits that are now adaptive to our survival but were not selected for their current effects) rather than adaptations. The point being that dysfunctions in adaptations do not account for everything we think is a mental disorder (see e.g. Murphy & Woolfolk, 2000 for this line of argument). I will return to this line of criticism in section 2.2.1.

Like Boorse, however, Wakefield (1992b) also thinks that there are natural functions of mental mechanisms:

> [M]ental mechanisms, such as cognitive, linguistic, perceptual, affective, and motivational mechanisms, have such strikingly beneficial effects and depend on such complex and harmonious interactions that the effects cannot be entirely accidental. Thus, functional explanations of mental mechanisms are sometimes justified by what we know about how people manage to survive and reproduce. For example, one function of linguistic mechanisms is to provide a capacity for communication, one function of the fear response is to help a person to avoid danger, and one function of tiredness is to bring about rest and sleep (p.383).

Functional explanations of mental mechanisms, for Wakefield, justify ascriptions of dysfunctions when a mechanism fails to perform its function. For instance, insomnia is a failure of a mental mechanism to regulate sleep. Wakefield acknowledges that it can be difficult to identify the natural functions of mental mechanisms, but he argues that it is an empirical matter for the natural sciences to decide.

The normative component of Wakefield's (1992b) approach states that dysfunction alone is not sufficient to call a phenomenon a mental disorder. Wakefield argues that phenomenon that are dysfunctions but not harmful should not be labelled as mental disorders because mental health professionals are "interested in the functions that people care about and need within the current social environment, not those that are interesting merely on evolutionary theoretical grounds" (p. 384). For instance, assuming that persistent auditory hallucinations in the form of hearing a foreign voice talk to you is the result of a dysfunctional mental mechanism, then if the person benefited from it (perhaps they make detailed wooden sculptures based on its suggestions) they would not count as disordered. The inclusion of a naturalistic aspect in the form of biological dysfunction, but also including a normative component of harmfulness, is why accounts like Wakefield's are termed hybrid accounts.

1.3.2. Tsou's Hybrid Approach

Like Wakefield, Tsou (2016;2021) has two necessary conditions that are jointly sufficient for calling something a mental disorder. The first necessary condition is that a disorder must have an identifiable biological mechanism that causes — in part or whole — its symptoms to reliably cluster (Tsou, 2016. p.416). In other words, that it is a HPC. Tsou argues that the underlying mechanisms that cause the symptoms of a disorder to reliably co-occur must be neurobiological. He backs this up with evidence that paradigmatic mental disorders have such mechanisms whereas pseudo-disorders have social mechanisms. For instance, there is evidence that the paradigmatic disorder schizophrenia is maintained by excessive dopamine in the mesolimbic pathway (Abi-dargham, 2004). In contrast, Tsou argues that pseudo-disorders like hysteria and multiple personality disorder are maintained by social causes, like the repression of women, rather than biological ones. This condition, however, is only half the story — while necessary, it alone is not sufficient. Like with Wakefield, to count as a mental disorder it must also cause harm to the individual.

1.4. Pragmatic Approaches

Pragmatists start from the commonly held position that psychiatry has multiple ends. Take a look at any psychiatry teaching programme or association of psychiatrists and you will find goals, missions, aims etc. ranging from patient advocacy, to ethical conduct, to furthering psychiatric research. For example, the longest standing medical specialty society, the American Psychiatric Association, lists such things as patient-focused treatment decisions, compassion for patients, and evidence-based principles of treatment among its missions and aims (Riba et al., 2005).

Zachar (2014), a pragmatist, organises the different ends of psychiatry under several different

categories. For instance, there are scientific or theoretical ends which encompass aims like finding out which psychopathological conditions are fit for scientific research and can be used to establish scientifically informed principles of treatment and intervention; professional ends like establishing the best standards of professional conduct, which can involve not treating conditions that will not easily remit on their own and distinguishing good from bad therapy responders; ethical ends like providing treatment to those who deserve it and respecting patient autonomy by avoiding invasive, non-beneficial 'treatment'; and sociopolitical ends like reducing stigma and advocating for patient welfare in various non-medical institutions.

The crux of the pragmatic approach comes in their argument that each of the different ends of psychiatry justify 'carving up' the world into disordered and non-disordered in different ways. In contrast to what the naturalist approach argues, they claim there is no one right way to carve out the concept of mental disorder from the world. Rather, the different ends of psychiatry provide lots of different, equally justified (or what Kukla (2022) calls legitimate and strategic) ways to choose whether something should count as a mental disorder. An example adapted from Kendler et al (2011. p.1145) helps demonstrate the pragmatic approach to classification: You have 1000 library books in need of classification. There is no one a-priori right way to approach classifying them. Rather, the way to approach classification is going to depend on your ends. If the goal is to have your library look pretty, utilising certain principles like 'order by colour and height' and 'throw out worn out books' would be better than doing it randomly. Similarly, if you want to make the library easy to navigate for readers, then principles or procedures for classification that help organise the books by topic and author makes sense.

Deciding how to classify something *as a mental disorder* is different from deciding how to organise the phenomena classified as mental disorders within a classification system. I am

talking about the later, but the analogy still illustrates the pragmatic approach: on the pragmatic view there are multiple different ways to decide whether a phenomenon counts as a mental disorder, there is no one a-priori right way to determine whether a phenomenon is a disorder. However, pragmatism is not an 'anything goes' approach – one is not completely free to determine how to classify something as a mental disorder. Each pragmatic approach has a different view on what restrictions there are to what can count as a mental disorder. In the next section, I discuss the most recent pragmatic approach to classification by Kukla (2022).

1.4.1. Kukla's Pragmatic Approach

Kukla's (2022) pragmatism is a radical departure from naturalist and hybrid approaches. While the naturalist argues that the natural sciences can provide the necessary and sufficient conditions for classification, and the hybridist argues that the insights from natural sciences are necessary but cannot be sufficient for that task since it cannot sufficiently specify the relevant harm involved, Kukla's pragmatism argues that the conditions for whether something counts as a disorder change depending on the contexts. Her central claim is that:

> Whether something is best classified as a disease [disorder] is a contingent, historically dependent, perhaps temporary, socially embedded fact that can vary from context to context. We can only settle whether it makes sense to call something a disease [disorder] by asking what our purpose is in using the concept in a given strategic context, and whether it will further our ends to classify it in this way. There just isn't, I will try to convince you, any neater story to be told. (p.132)

Two things about her claim are worthy of mentioning right off the bat. The first is that while Kukla speaks of disease she takes her view to apply to disorders too: the intent of her view is "to capture both the discourse of disease and the discourse of disorder" (p.132). The second, and most important to understanding her view, is fleshing out the important role the plurality and context-dependency of legitimate strategic ends or goals of classification. For Kukla, there is no one legitimate strategic goal of disorder classification that can suit the diverse contexts in

which classifying something as a disorder plays an important role. In contrast to naturalists like Boorse who argue that science is sovereign and so what can be classified as a disorder must suit the ends of scientific research, Kukla does not prize one end over another. She states that "the difference between him [Boorse] and me is that he thinks that the "theoretical" uses of the concept are somehow privileged, and show us its real meaning or essence, which I deny" (p.135). And, unlike hybrid approaches who provide a definition of mental disorder that works in all contexts, Kukla makes no such attempt, claiming it is unachievable.

The reason Kukla argues against searching for a definition of mental disorder that works in all contexts is because she argues that disorder classification is tied to many different institutions with vastly different ends, which leads to different criteria for whether it is strategic to mobilise the label of disorder. For Kukla, disorder is a pluralist institutional kind, which is to say that "the category of disease [disorder] is constitutively dependent upon and embedded within a social institution" (p.143). Again, several clarificatory comments are warranted. Kukla is not saying that specific disorders, like schizophrenia, are not natural kinds — Kukla is talking about the category of disorder, not about specific disorders. Kukla also does not tie the category of disorder to any single institution, but rather to a plurality of institutions (e.g., clinical institutions like hospitals, political institutions like the criminal justice system, and economic institutions like insurance companies). Kukla makes the analogy to work, which she also argues is a pluralist institutional kind:

> [W]e understand what counts as work through the lens of an economic system of differentiated wage labor. This is the institutional background against which the concept of work gets a grip. But we don't simply equate work with wage labor. Indeed, the question of what counts as work depends heavily on who is asking and for what purpose. The concept of work has life in many different domains, and what counts as work shifts depending on the domain and the context, although it always remains true that by calling something work we are drawing on the resources of the wage labor economy to understand it and perhaps to manage it (p.149).

According to Kukla, what counts as a disorder or work depends on the question being asked and the goal one is trying to achieve. For instance, she argues that if one is asking whether a person should pay taxes, then unpaid domestic labour is not classified as work, but, if one is settling an argument with one's partner on who has done more work lately and is deserving of a night off then it is. To expand her example to disorders, she would say that if a scientist is asking whether a phenomenon like narcissistic personality disorder is 'real' then it might not make sense to classify it as a disorder, but, if someone is outside of the workforce because of it and we are asking whether insurance companies should cover them then it might make sense to classify it as one.

Classification, on Kukla's view, is thus like classifying one's library. Deciding how a particular book is to be classified depends on the goal classification serves in the context. But, the same disanalogies mentioned in section 1.4 also apply. Kukla is not talking about how to organise books within one's library, but whether something should be included in the library (i.e., whether it counts as a book). Nor can one choose to classify a phenomenon as a mental disorder however one wants, unlike how one is free to choose whether something should be in their library. Rather, Kukla is claiming that when we are trying to decide whether something counts as work or a disorder, we must decide with reference to the wage labour system in the case of work, and medical institutions in the case of disorders. But, in doing so we do not just equate work with wage labour or disorders with whatever medical institutions count as disorders. Rather, these institutions serve as a reference point against which we can assess whether the end is legitimate and whether it is strategic (conducive to the legitimate end) to do so.

It is in this way that Kukla sets herself apart from radical social constructivist views (e.g., Englehart, 1974) which argue that a disorder is, by definition, whatever the institution of

medicine recognises as a disorder. She disagrees with these views because they cannot support the intuition that medical institutions can get classification wrong (a point I will return to in section 2.1.1). The main difference between her view and radical social constructivist views is that, on her view, classifying something as a disorder must serve as a legitimate goal of an institution. When it is not legitimate, then on her view (contra radical social constructivism) it is not a disorder, rather, the institution has made a mistake. Her view classifies things as disorders "not in terms of what we do medicalize, but in terms of what it is strategically a good idea to medicalize, given a legitimate set of goals, whether or not we do" (p.145). In agreement with the radical constructivist, however, her view does lead to the result that, had there been no institutions where labelling something as a disorder plays an important role, there would be no disorders: she states that "just like a locked room that is materially identical to a prison is not a prison if it is not located within a social system in the right way, pancreatic cancer, though a natural kind in its own right, is not a disease if it doesn't bear the right relationship to medical institutions" (p.144).

What, then, are the legitimate goals of classifying something as a disorder? True to her staunchly anti-unified anti-systematic approach, Kukla does not give a systematic or principled way to determine what is a legitimate goal and what is not. She does, however, outline that a "legitimate goal" of psychiatric classification is (i) "one that is socially acceptable according to our loosely shared ethical and political norms" and (ii) "within the epistemology and metaphysics of medicine, the condition or cluster can qualify as pathological" (p.136). Kukla does not specify either of these conditions and argues that, like with disorder, any attempt at a systematic approach to when medicalisation — i.e., making something medical (more on this is section 2.1.1) — will be acceptable according to our shared ethical and political norms will fail. So too, will any systematic approach to when something qualifies as pathological. Giving

a general philosophical definition of pathology, she writes, will "be as fruitless as trying to define disease" (p.137). They do, at least, provide several different examples of institutions where the reasons why it is legitimate and strategic to classify something as a disorder are different. They are:

1. From a medical institution, it is legitimate to classify a phenomenon as a disorder if it is a productive, unified object of scientific medical study. Under this end, we may legitimately decide to classify a phenomenon as a disorder on the basis of whether it is robust enough to be used in inductive inference or a variable in statistical analyses by medical researchers. In agreement with Boorse, this end serves as a sufficient reason to classify something as a disorder independent of any desire to treat it. However, she differs from Boorse in that this reason to label something as a disorder does not work in all contexts. For instance, a phenomenon might not be robust enough to be used as a variable in statistical analysis, but in the context of insurance companies deciding whether people outside of work because of the phenomenon should receive compensation we might still decide to label it as a disorder (see point 3).

2. From a clinical institution, it is legitimate to classify a phenomenon that is a productive, unified target of treatment as a disorder. Under this end, we may legitimately decide to classify a phenomenon as a disorder on the basis of whether it lends itself to being a unified target of treatment, regardless of whether we have any lead on its etiology. Kukla gives the example of depression as fitting under this end. She states that various cases of depression can respond well to similar treatment even if their underlying causes are different. So, in this context, it is strategic to label all cases of a phenomenon like

depression as a disorder even though it may not be strategic to do so if one is an epidemiologist interested in studying disorders.

3. An insurance institution may decide to classify a phenomenon that should be funded or covered by insurance as a disorder. Kukla claims this case is independent of the previous two, and thus leads to different considerations on whether it is strategic to label something as a disorder, because we can have no debate over the empirical facts (e.g., whether it is a unified target of treatment or study) but still wonder whether insurance companies should cover it.

4. Socio-political institutions can decide to label something as a disorder on the basis that people with it deserve special considerations in a just state. For instance, whether they should receive workplace accommodations or should have guaranteed access to care. These institutions can also base classification on whether labelling something as a disorder would legitimise and clarify people's actions towards individuals with the phenomenon (or communities comprised of individuals with the phenomenon) in a positive way. Kukla gives the example of oppositional defiant disorder — a disorder centred around uncooperative, defiant, and hostile behaviour towards others in multiple contexts. Kukla argues that, on the basis of whether we consider this a disorder, it can either legitimise and clarify people's interactions with those who have the disorder or obscure it.

The different institutional reasons for labelling something a disorder Kukla identifies are very broad and far reaching. It is a feature of her view that one might question what, if any, relation is there between justifying classifying a phenomenon as a mental disorder on the basis that

insurance companies should fund treatment for it and justifying calling it a mental disorder on the basis of the phenomenon being a unified object of medical study. In response to this Kukla answers that there is no systematic relationship, making a unified account impossible. She writes that there is no reason to think that "the same conditions that are productive unified objects of study are the ones that we want to insure" (p.135) and that "the burden of proof seems to be on the philosopher who thinks that one univocal concept [of disorder] can serve all these very different masters simultaneously to say how this is so" (p.135). This results in two important features: (1) it means that no set of context-independent necessary or sufficient conditions can be generated that will appropriately label a phenomenon as a disorder in a situation where it is legitimate to do so, and (2) it can count phenomenon that are not dysfunctions or natural kinds (in the naturalistic sense described above) as mental disorders since that naturalistic criterion is not a necessary condition anymore. To many, both of these may seem like weaknesses of an account, or reasons to disfavour it, rather than reasons to accept it. I agree that both of these 'strengths' are unintuitive, but I will argue that there is good reason to accept them as strengths in the next section.

1.5. Section Summary

In this section I have explained the current approach to classifying mental disorders in the DSM and noted several problems with it. I then discussed three alternative approaches to classification — Naturalistic, Hybrid, and Pragmatic. While the specifics of each camp can differ, each camp has a distinct relation to the question of 'how relevant are the natural sciences to classifying mental disorders?'. Naturalistic approaches argue that the natural sciences provide us with the necessary and sufficient conditions for classifying mental disorders. Hybrid approaches argue that the natural sciences provide a necessary but not sufficient condition, and Pragmatic approaches argue that the conditions for classifying something as a disorder change depending on the context. In the next section, I will argue that the pragmatic approach is to be

preferred.

Section 2: Pragmatic Approaches are Preferable

The best argument I have seen for a pragmatic approach comes from Kukla (2022). Kukla's (2022) argument for a pragmatic approach starts from the pragmatist's observation of the many different ends psychiatry is committed to. Recall the example from the American Psychiatric Association which highlighted that the association was committed to scientific or theoretical ends, like finding out which syndromes are natural kinds and so fit for scientific research; economic and ethical ends like providing treatment to those who we perceive are in need of it but not treating conditions that will easily remit on their own; and sociopolitical ends like reducing stigma (Zachar, 2014). The second step is to note that these different ends can lead to contexts in which different institutions have different reasons to count something as a disorder. The examples Kukla gives to back up their pragmatism, however, focus on diseases not disorders as they take the two to be synonymous. Consider the following puzzle case:

> How at one and the same time can it be perfectly appropriate to resist classifying deafness as a disease in response to arguments by the deaf community that deafness is a natural variation on human capacities, and also perfectly appropriate to endorse classifying deafness as a disease for the purposes of securing funding from insurance companies for those who do want treatment for it? (p.131)

Kukla's intuition here that it is "perfectly appropriate" to resist classifying deafness as a disorder in the context of the deaf community, but that it is "perfectly appropriate" to classify deafness as a disease in the context of deciding whether companies should fund treatment for it, is one that I share. It is, however, just an intuition, but one that is necessary in order for the argument to get off the ground. I will not defend this intuition, but rather assume it for the sake of argument.

Kukla uses examples like this to show that in different contexts there are different reasons to consider for whether something should count as a disorder. Here is an adapted version of her puzzle case that more clearly shows how different institutional contexts can lead to different reasons to consider whether something is a disease on her view:

> How can it be perfectly appropriate to resist classifying deafness as a disease in response to arguments by the deaf community that deafness is a natural variation on human capacities, and also perfectly appropriate to endorse classifying deafness as a disease for the purposes of researching the effectiveness of certain interventions on deafness?

In section 1.4.1 I outlined several different legitimate reasons on Kukla's view to classify something as a disorder depending on the context. Two are teased apart here (in the sense that each only applies to one context): one is whether a condition is a productive, unified target of treatment (point 2), the other is whether labelling the condition as a disorder would legitimise and clarify people's actions towards individuals with the phenomenon (or communities comprised of individuals with the phenomenon) in a positive way (point 4). In this case there are two contexts and the legitimate reasons to classify something as a disease do not apply equally. On the one hand, there is deafness in the context of the deaf community, and on the other hand there is deafness in the context of medical research. In the context of medical research – a researcher wanting to create a variable to be used in an analysis on the effectiveness of an intervention for a disease (like hearing aids), for instance – considering whether grouping all cases of deafness as disease would clarify people's actions towards certain individuals or communities is not strategic – it would not lead to better groupings in the analysis. So, generating sub-groups in the analysis based on whether a deaf person would benefit (in the sense described in point 4) from being labelled as diseased is irrelevant to the medical

researchers classificationary needs.

However, the same reasoning would not be appropriate for contexts where a community of deaf individuals constitutes the institutional kind at issue. Given that the label of disease can be stigmatising, considering whether the label would clarify how to interact with them in a positive way is relevant. Further, whether deafness is a productive, unified target of treatment might not be relevant. As is assumed in the example, a deaf person living in a deaf community might not care whether a medical treatment for their deafness can be treated in a similar way to other cases of deafness. In this context, grouping all deaf people together as diseased because the same medical treatment is effective is not strategic – it would not help to decide whether labelling a person as deaf would clarify, in a positive way, how others interact with them.

The point of the case is to show that whether a reason to classify something as a disease is legitimate and strategic depends on the contexts. This means that any approach which aims at giving a single, unified definition of mental disorders, or one with one set of necessary and sufficient conditions for all contexts, will be inadequate — it will inappropriately label something as disorder when it is not legitimate and strategic to do so. Unlike Kukla, I am not assuming that diseases and disorders are synonymous. Nonetheless, I think that the framework of the case that Kukla gives can be applied to mental disorders. Consider the following case:

> How can it be perfectly appropriate to resist classifying a phenomenon as a mental disorder in response to arguments by advocacy groups that the symptoms are a non-harmful variation on human capacities that constitutes an important, meaningful part of their life, and also perfectly appropriate to endorse classifying the phenomenon as a disorder for the purposes of securing funding from insurance companies for those who do want treatment for it?

The case is deliberately underspecified. I have left it as open as possible so the mould can fit many different examples. Here is one way of adding some realism to the example from Washington (2018):

> Vincent lives alone on a remote piece of property in the Pacific Northwest. After college he began to remove himself from his social circles, and now spends most of his time in his home, making highly detailed wooden sculptures, based on the suggestions of 'other voices.' Vincent has no phone or Internet access, and interacts mostly with the owner of the general store in a nearby town. He makes a modest living by occasionally making a sale of a sculpture. His buyers find him to be distant and abrasive, and are often disturbed by his behavior, and lack of adherence to social norms. (p.238)

Contrast the case of Vincent, who seems to benefit rather than be harmed by this phenomenon, with a case based on Stuart from the Hearing Voices Movement (HVM) (Cox, 2018):

> Stuart is an office worker who started hearing an intelligent, manipulative, aggressive and bitter voice in his 20s. The voice keeps him awake at night. It reads through his thoughts and memories to manipulate him. It is critical of him, constantly pointing out errors that are not his fault. It torments him and endangers him when he is trying to cross the road. It is pulling him and his life apart. Stuart is distressed by the voice and wants help dealing with it.

Realistic cases like Vincent and Stuart's show something important about how to decide whether something should be counted as a disorder. Namely, that one reason to legitimately count something as a disorder may not apply in all contexts. If one focused on the reasons stemming from a medical institution, like Boorse, and gave a scientistic procedure for identifying a condition as a mental disorder, then both Vincent and Stuart would count as disordered. However, just like in the deafness case, I take it as intuitive that it is wrong to count them both as disordered. Stuart may well embrace a diagnosis of a mental disorder to secure

funding to treat a condition he finds highly distressing. But, one can quite readily imagine Vincent rightly refusing the label of mentally disordered because of the consequences this has for him (e.g., it may ostracize him and harm his sculpting business).

Trying to apply the reasons based on point 4 regardless of context is also no more appropriate. Just like in the deafness case, this would have bad consequences when it comes to medical sciences – splitting Vincent and Stuart into different groups (Stuart included in the disordered group in the analysis but not Vincent) could harm the analysis if one was looking for genetic factors that caused the phenomenon. Just like in the deafness case then, the different contexts generate different conditions for classifying something as a disorder.

Because it seems implausible to generate any one set of necessary and sufficient conditions for counting something as a mental disorder that works for all contexts, Kukla (2022) argues that we should give up on trying to generate any. Rather, we should be pluralists and accept that there are different things to consider when deciding whether something is a disorder based on the institutional contexts. I think Kukla is right, but there are several counter-arguments the Naturalist and Hybrid approaches can give that need to be dealt with. In the following, I will go through what I take to be the best counter-argument each approach can give and defend Kukla against it.

2.1. Naturalist Response

The first, and most pressing response, is that we should revise the institutions that disorder plays a role in, and thus restrict the legitimate reasons to count something as a disorder. Naturalist's, like Boorse, may stand their ground and state that "the judgment that something is a disease [or disorder] is a theoretical judgment that neither entails nor is entailed by any therapeutic judgment about people's need for medical treatment" (1977. p.544). This response

involves biting the bullet and labelling people like Vincent as disordered even though it may be actively harmful to his welfare. This bullet seems big enough to blow one's brains out, but there is a reason someone might give in support of going this route — that naturalism prevents 'overmedicalisation' whereas pragmatism promotes it. In the following, I will explain what overmedicalisation is, why one might think that naturalism prevents it and pragmatism promotes it, and show why this is wrong.

2.1.1. Naturalism Protects Against Overmedicalisation, Pragmatism Promotes it

Medicalisation denotes a process whereby previously non-medical problems become defined and treated as medical problems (Conrad & Slodden, 2013). For instance, it is likely that there have always been children who were hyperactive, restless, and impulsive, however, it was not until the late 1950s that such behaviour was viewed as evidence of a mental disorder, namely, attention deficit hyperactivity disorder (ADHD) (Conrad & Slodden, 2013).

Medicalisation, according to Conrad and Slodden (2013), is a bit like urbanisation or secularization — it describes a process that is not inherently good or bad but can become so depending on the context. As Kigma (2013) discusses, medicalisation, particularly in the psychiatric domain, has many consequences. Some are beneficial — it can lead to special access to treatment, better workplace accommodations, special treatment in law, and lifting of emotional and social burdens. However, it can also have severe negative consequences like stigmatisation, ostracization, and infringement of rights and loss of autonomy. Further, as Hacking (1995) discusses, a phenomenon being labelled as a mental disorder shapes the way people with the condition see themselves, how they behave, and how others see and interact with them. It is vital, then, to make sure that the right phenomena are medicalised, less someone with a phenomenon like Vincent be medicalised to their detriment.

When the contexts of medicalisation are bad, or when a condition is medicalised for the wrong reasons and has negative consequences, it is termed "overmedicalisation". There is no shortage of examples of overmedicalisation. A few pertinent examples are homosexuality, masturbation, hysteria, and drapetomania (a slave's desire for freedom). Many people think that medicalisation in psychiatry goes wrong when social and moral values get factored into what should count as a mental disorder (see, for instance, Charland, 2006). [3] Naturalistic approaches, which give no role to social and moral values in whether a condition should count as a disorder, seem to be in a good position to deal with cases like homosexuality and drapetomania. In contrast, it seems like pragmatic approaches, which make social and moral norms relevant, are ill-placed to deal with them. However, the situation is the reverse — as I will argue, pragmatic approaches are in a better position to prevent cases of overmedicalisation than naturalistic one's. I will first show how naturalistic approaches fail to prevent overmedicalisation and then show how pragmatic approaches can prevent it.

The naturalist attempts to deal with overmedicalisation by making social and moral values irrelevant to what counts as a mental disorder. Masturbation, hysteria, etc., were only considered disorders because they are socially disvalued or 'deviant' behaviours. By removing our desire to 'treat' these socially deviant conditions from discussions of what should count as a mental disorder, and classifying as disorders only conditions that impair a psychological function below the typical efficiency of the reference class, Boorse seems to be in a good position to deal with such cases. For instance, it is clear that, however the reference class is carved out, masturbation and a desire for freedom and autonomy is unlikely to count as a mental

[3] It is not that people with cluster B personality disorders do not deserve treatment or help, Charland argues, but the help they get needs to be moral rather than medical in kind. Medicalising moral conditions, he argues, is a poor pragmatic choice.

disorder on Boorse's view. However, while Boorse may be able to deal with some cases, the theory does not have the resources to adequately demarcate paradigmatic cases of disorders from non-disorders, or, in other words, to prevent medicalising conditions we commonly think are not disorders.

To see why Boorse fails, consider again the case of Vincent and Stuart, where, on Boorse's view, the justification for medicalising them is the same but where Stuart is a paradigmatic case of a mental disorder, but Vincent's status as disordered is questionable (at best) because, rather than harm him, it benefits his work. At least that is what I will assume here, for the sake of the argument. On Boorse's view, schizophrenia — characterised by the presence of delusions, hallucinations, or disorganised speech but an absence of manic or depressive episodes (APA, 2013. p.99) — is classified as a mental disorder because it is a condition which impairs the functional readiness of the psychological process underlying the minds capacity to distinguish whether the cause of a sensation is internal or external to the body, given the reference class of contemporary human beings. If the only institutional context where it is appropriate to label something as a disorder is the kind of medical research institution Boorse imagines, then both Vincent and Stuart would be medicalised (i.e., counted as disordered). However, medicalising Vincent is a clear case of overmedicalisation according to the characterisation of overmedicalisation laid out above – that is, a context where labelling someone as disordered has harmful consequences for the medicalised individual. By making the scientific institution Boorse imagines sovereign, then all other institutions that use the concept must follow suit. For instance, the HVM institution mentioned in section 2 must classify themselves as an institution comprised by disordered individuals, even though that is what they are staunchly against – the aim of the HVM is to reduce the stigma around hearing voices by arguing that it is not a disorder but rather a phenomenon that can be consistent or aid

in flourishing (Corstens et al., 2014).

Kukla's pragmatic approach, on the other hand, which is context sensitive, is well placed to deal with cases like Vincent and Stuart. Pragmatic (as well as hybrid) approaches, can argue that it is not appropriate to label every individual that persistently hallucinates because doing so would cause unnecessary stigmatisation (i.e., is not consistent with point 4). This type of response is also available for a host of other conditions that were wrongfully medicalised — drapetomania, for instance.

Perhaps one might argue that the pragmatic response does not fully respond to the worry about overmedicalisation. The worry about overmedicalisation is not just that it is not appropriate to label something as a disorder in a certain context, but that, regardless of context it is wrong. What a theory of mental disorders should be able to do is give a reason why it was always wrong, regardless of contexts, to medicalise things like homosexuality, drapetomania, and masturbation. It seems like Kukla's pragmatism is unable to provide a reason.

This is an understandable criticism. It is, in fact, a feature of Kukla's pragmatism that it does not, *a-priori,* prevent a phenomenon from being medicalised in some institutional context. However, *a-posteriori* it is a different story. Kukla's (2022) way to argue, *a-posteriori*, why it was never legitimate to label something a disorder comes from their second clause that "within the epistemology and metaphysics of medicine, the condition or cluster can qualify as pathological" (p.136). I will explain what this means shortly, but the implication of this is that, regardless of which institution is trying to decide whether something should count as a disorder, a phenomenon is not (nor ever was) a disorder if it cannot count as pathological.

Explaining how this is the case for Kukla is no easy task as they do not give a detailed explanation of what they mean by 'can count as pathological within the epistemology and metaphysics of medicine'. In terms of explaining what can count as pathological, Kukla states:

> That diseases are pathological from within the perspective of medicine is the thin sense in which they are definitionally bad. This is consistent with them playing a harmless or even a positive role in many people's lives, since they need not be pathological in some essential, cross-contextual sense. But it does seem to me that if medicine doesn't count a condition as pathological at all, then it is misleading and inappropriate to call that condition a disease (p.138)

So, something is pathological from the perspective of medicine if it is defined as bad by it, with no expectation that the phenomenon defined as bad harms an individual's welfare. Further, she argues that "since the perspective and tools of medicine are not unified or internally consistent" (p.137) there does not need to be consensus from within medicine whether something is pathological for it to count as pathological – to count as pathological, it must only count as bad from *a* perspective within medicine.

In terms of what they mean by 'epistemology and metaphysics of medicine', even less is said. They do, however, at least provide some examples of how this response can mean it was always wrong to label something as a disorder. According to Kukla, one feature of the metaphysics and epistemology of medicine is that it draws on scientific methods to discover natural kinds. One feature of discovering that something is a natural kind, according to Kukla (2022), is that it allows us to use a kind of "reverse causality" (p.147) whereby we can assert that some conditions have always existed (or at least existed long before we discovered them). For instance, take the example of ADHD. Assuming ADHD is a natural kind, we can assert that it has existed long before it was recognised as a disorder because, in virtue of it being a natural kind, it is a robust cluster with a long natural history — after all, "most diseases [disorders] do not just pop into existence (although some do, of course)" (Kukla, 2022. p.147). So, although ADHD was only recognised as a medical condition in the 1960s, on her view, Kukla has the

resources to say that it always could have been one.

A similar line of reasoning using reverse causality can show why it was always wrong to consider something a disorder. Another feature of medical epistemology that Kukla also draws on is falsification — a method whereby a cause of a disorder is posited and predictions (which must be able to be shown false) are made. If the predictions are false, then the cause is refuted. Kukla's concept of reverse causality also applies here — if it is false now, then we can confidently assert that it was always false. After all, disorders and their causes do not (typically) just pop out of existence. For instance, consider the case of the 'wandering uterus' which was posited as the cause of hysteria. Hippocrates famously speculated that the cause of hysteria in women was the uterus detaching itself from the lower abdomen and moving throughout the body. As Lippi et al (2020) discuss, despite the long-lasting popularity of this hypothesis it was falsified around the 3rd century BC by Herophilus of Chalcedon. So, using the concept of reverse causality, we can assert that if it is false now that a wandering uterus causes hysteria, then it was always false that wandering uterus' caused hysteria.

When it comes to cases of overmedicalisation like homosexuality, Kukla can also use the 'reverse causality' argument to state why it was wrong to have ever considered it a disorder. The history of why homosexuality was considered pathological by medicine is long and complex (see, for instance, Drescher (2015) for an overview of the different views). For simplicities sake, I will consider only what Zachar (2014. p.115) calls the "conventional argument" for the continued medicalisation of homosexuality in the 20th century. According to Zachar (2014), the argument for why homosexuality was pathological was because it causes attachment issues — it prevents stable, loving romantic relationships. Homosexuality was thus considered a medical condition in the 20th century for reasons akin to other disorders —

namely, because it is a phenomenon that causes sub-optimal functioning (Zachar & Kendler, 2012). Specifically, sub-optimal functioning in attachment. By the lights of modern medical epistemology, however, this is a mistake — loving, stable relations between members of the same sex are not only possible but common. Thus, Kukla can apply her reverse causality argument here in much the same way that it is applied to the wandering uterus: Homosexuality was considered a medical condition in virtue of it causing sub-optimal attachment, however, we now know this to be false. So, Kukla can say that if it is wrong now to pathologize homosexuality on the basis that it causes sub-optimal attachment, then it was always wrong to pathologize homosexuality on this basis.

While this is only applied to one argument for why homosexuality was considered pathological, the general steps of the argument are clear: (1) identify why medicine considers a phenomenon as pathological, (2) assess, using the epistemological and metaphysical tools of medicine (e.g., evidence-based research and assumptions about what it means for something to be a natural kind), whether the reason is valid, (3) on the basis of whether the reason is valid or invalid, conclude that the phenomenon could have always counted as pathological (like in the case of ADHD), or that it was wrong to have ever considered it pathological (like in the case of homosexuality).

Admittedly, this is not the most satisfying conclusion for two reasons. The first is that Kukla does not place any restrictions on what medicine can count as pathological. Her characterisation of what can count as pathological is radical social constructivist – something is pathological if medicine counts it as such. Meaning that, depending on the reason why something like homosexuality is considered pathological, the above argument might not work. Though, Kukla is certain that given the reasons medicine considered homosexuality pathological it “is a clear

example" of something that is not (p.151). The second is that, in trying to account for the intuition that in some cases it was always wrong to medicalise a condition, her view ends up looking a lot like a hybrid account. She makes it a necessary condition that the phenomenon can count as pathological from the perspective of medicine, which seems to amount to the condition that it is necessary for something to be a disorder that it is a natural kind (like on Tsou's view) or a biological dysfunction (like on Wakefield's view). Indeed, she uses this language herself when she states that "medicine itself frames and understands it [deafness] as a pathological dysfunction of the aural system" (p.137). However, her view is still preferrable to Wakefield and Tsou's because it does not restrict itself to one way that something can count as pathological. Her criterion for what counts as pathological does not assume that the phenomenon must be a natural kind or must be a biological dysfunction. Rather, it is disjunctive – something can count as pathological from a variety of different perspectives within medicine. As I will argue in the next section, Wakefield's restriction that a phenomenon must be a biological dysfunction to count as a disorder, and Tsou's restriction that a phenomenon must be a natural kind maintained by a biological mechanism, generate problems for their view that do not exist for Kukla.

Further, it is misleading to characterise Kukla's view as a hybrid account. What characterises a hybrid account is having one set of jointly necessary and sufficient conditions (one based on medical science, the other based on something normative like harmfulness) that can be used to label something as a disorder regardless of context. While Kukla has a necessary condition that must be met in all contexts, all other conditions can vary depending on the institutional context. So, her view is still best characterised as a pluralist institutional approach.

In conclusion, one might be tempted to respond to cases like Vincent and Stuarts and get some

cross-contextual consistency in whether a phenomenon counts as a mental disorder by revising the institutions we consider having legitimate reasons to count something as a disorder. The idea being that we can generate one set of necessary and sufficient conditions for what counts as a disorder in all contexts if we restrict the institutions that can legitimately decide how to classify something as a disorder to only one – namely, Boorse's imagined medical research institute. The main motivation for this is avoiding overmedicalisation. However, I have shown that an institute that provides reasons to label something as a disorder based on Boorse's naturalism is unable to avoid overmedicalisation, whereas Kukla's pragmatism (and hybrid approaches) can. With the main motivation for such a revision disputed, I turn to the next objection.

2.2. Hybrid Response

The other main objection I see is a denial that cases like the deafness one and Vincent and Stuart's support Kukla's conclusion. The purpose of the Vincent and Stuart case is to show that different institutions have different classificationary needs that lead to different results on whether someone like Vincent or Stuart count as disordered. The conclusion Kukla infers from this is that there can be no one set of necessary and sufficient conditions for classifying something as a disorder that works in all contexts. One way to deny this conclusion is to show that there is such a set that can give the same results as Kukla's pragmatism. Namely, that depending on the contexts, two people with the same phenomenon do not both count as disordered. It seems that Hybrid approaches, which involve both naturalist and normative aspects, can provide such a result. In the following, I discuss how Wakefield and Tsou can provide such a result and argue that, while it works in some contexts, it does not provide the desirable result in as many contexts as Kukla's.

2.2.1. Wakefield's Response

Wakefield (1992) argues that there are two necessary conditions that are jointly sufficient for classifying mental disorders. One is naturalistic — the phenomenon must be a result of a dysfunction – the other is normative — the dysfunction must be harmful. Looking at Kukla's original deafness case, this appears to give a simple way of giving the assumed 'perfectly appropriate' response: the person in the deaf community is not diseased because, although they have a dysfunction, it is not harmful to them in their context. However, the person seeking treatment is (presumably) harmed by it, thus they are diseased and so their treatment should be funded. Wakefield can also give an identical response in the Vincent and Stuart case.

However, I think that a new case can be generated that does create a problem for Wakefield because of his commitment that disorders are necessarily dysfunctions of an evolved mechanism. This commitment means that Wakefield fails to produce the assumed desirable response. To see how this is possible, consider first a case by Murphey and Woolfolk (2000):

> If you put a smoke detector too close to your stove, it will persistently emit noises that you will not welcome. Taking it apart and looking for what is broken will not help you since the problem does not lie in the device. The gadget itself may be working exactly as its designer intended it to. The trouble is environmental. Even though the smoke detector conforms to its design, it is in an environment that guarantees that it will persistently behave in ways which you do not value (p.244).

In this case, it is clear that there is a problem worthy of intervention or treatment even though everything is functioning exactly as intended. Wakefield's commitment to biological dysfunction means that the intuition that there is a problem here worthy of treatment cannot be validated. While this case is about gadgets, not disorders, constructing an analogous case that relates to mental disorders is not difficult.

I think it is rather easy to imagine several different cases that present a problem for Wakefield, but there is one in particular that stands out — Major Depressive Disorder (MDD). Consider the following syndrome made up of the following symptoms:

> Depressed mood most of the day, nearly every day.
>
> Markedly diminished interest or pleasure in all, or almost all, activities most of the day nearly every day.
>
> Feelings of worthlessness or excessive or inappropriate guilt nearly every day.
>
> Diminished ability to think or concentrate, or indecisiveness, nearly every day.
>
> Recurrent thoughts of death.

According to the Diagnostic and Statistics Manual-4 (DSM-4) the presence of these symptoms over a two-week period counts as a mental disorder (i.e., MDD) *unless the person is grieving the loss of a loved one who has died less than 2-months ago* (APA, 1994. p.327). The thought being that "a full depressive syndrome" is an expectable response to the death of a loved one (or a common stressor). The DSM and Wakefield (Horwitz & Wakefield, 2007) both argue that it does not indicate a dysfunction to have these symptoms when grieving. It is for this reason that the so-called "bereavement exclusion" exists. However, there were a group of people who advocated for the bereavement exclusion to be removed from the next addition of the DSM (DSM-5). They argued that it did not matter for what reason people developed the syndrome, those who had it and wanted help should have access to it. Zisook et al (2007) and Zisook and Kendler (2007) further argued that the prognosis of people who developed the symptoms in response to loss of a loved one was no different than the prognosis of people who developed it for other reasons, thus, including a time criterion for grief only does not make sense. So, on the one hand there is the medical scientifical institutional reason that argues against medicalising a 'normal' or non-dysfunctional part of life, and on the other hand there is the argument based on the insurance institutional reason that we should count it as disordered in order to secure

funding for treatment for a harmful phenomenon that will not easily remit without intervention.

Moulding the example of MDD to fit the above case structure we get:

> How can it be perfectly appropriate to resist classifying a phenomenon consisting in depressed mood, loss of interest or pleasure in activities, thoughts of death etc. as a mental disorder in response to arguments that the phenomenon does not indicate a dysfunction, and also perfectly appropriate to endorse classifying the phenomenon as a mental disorder for the purposes of securing access to treatment for those in need of it?

Such a case presents a problem for Wakefield's view but not Kukla's. Wakefield's response to the deafness and Vincent and Stuart case does not work here because, unlike with those cases, Wakefield argues there is no dysfunction (Horwitz & Wakefield, 2007). The insurance institutional reasons for classifying something as a disorder call for people who have MDD in response to loss of a loved one to count as disordered so they can secure access to treatment, even despite the fact that there is no malfunctioning evolved mechanism. Thus, in the MDD case, Wakefield fails. His commitment makes dysfunction a necessary condition, which prevents the assumed desirable response of labelling MDD caused by grief as mental disorder. In contrast, Kukla can argue that MDD in response to grief is a mental disorder on the grounds that there is at least a perspective from within medicine that counts it as pathological - i.e., the perspective of Zisook et al (2007) and Zisook and Kendler (2007) – and that it is legitimate and strategic to count it as a disorder from insurance institutions. In conclusion, while Wakefield's approach looks promising, it fails when it comes to mental disorders like MDD.

2.2.2. Tsou's Response

Tsou seems to be in a stronger position than Wakefield. Tsou's hybrid approach argues that

mental disorders are natural kinds maintained by a biological mechanism that has harmful effects. Tsou can thus provide a similar response to Wakefield in the deafness and Vincent and Stuart case. Further, Tsou is in a better position to deal with the MDD case. His view of mental disorders, based on the most recent scientific evidence, does away with the notion of dysfunction. So, even though developing MDD in response to the loss of a loved one may not be a dysfunction, it can still count as a mental disorder so long as it is a natural kind maintained by a biological mechanism, which he argues MDD is (Tsou, 2021/2016). However, while it can deal with the MDD case, other cases present problems for his view.

As discussed in section 1.3.2, a necessary condition for something to be a mental disorder for Tsou is that the phenomenon is a natural kind maintained by a biological mechanism. As Tsou (2021) points out, in many cases there is converging evidence that paradigmatic mental disorders, like depression and schizophrenia, are natural kinds maintained by such a mechanism. However, for many disorders currently in the DSM there is no such convergence. For instance, Gagné-Julien and Bérubé (2023) give the example of Social Pragmatic Communication Disorder (SPCD). SPCD is characterised by a difficulty with the social use of language (e.g., implicatures) and manifests in such behaviours as not following the conventional rules of storytelling or not changing the use of language to meet the needs of the listener or situation. SPCD shares many features with autism, but is differentiated by the absence of restricted, repetitive behaviours (e.g., repetitive motor movements or speech) which must be present for an autism diagnosis. The diagnosis of SPCD was created because clinicians saw that a group of people needed access to treatment who did not meet the diagnosis of other neurodevelopmental disorders (like autism). Being a new diagnosis created from a perceived need for treatment, however, there is little evidence on whether SPCD is a natural kind maintained by a biological mechanism. It seems like, in this context, counting this phenomenon

as a disorder is justified based on the insurance institutional reason alone (point 3). Tsou's commitment to mental disorders necessarily being natural kinds thus generates problems when it comes to cases like SPCD. Putting this case into the format that teases two contexts apart, we get:

> How can it be perfectly appropriate to resist classifying a phenomenon a mental disorder in response to arguments that the phenomenon is not a natural kind maintained by a biological mechanism and so not fit for medical scientific research, and also perfectly appropriate to endorse classifying the same phenomenon as a mental disorder for the purposes of securing access to treatment for those in need of it?

Thus, while Tsou can respond to the MDD case, the SPCD case still presents a problem for his view but not Kukla's. Because Tsou makes it a necessary condition that to count as a disorder a phenomenon must be a natural kind maintained by a biological mechanism, his view is ill-placed to deal with cases like SPCD where the insurance institutional reasons justify classifying it as a disorder despite the lack of evidence whether it is a natural kind maintained by a biological mechanism. However, Kukla can argue that, in the right context, it should count as a disorder because there is a perspective from within medicine that sees it as pathological – i.e., the perspective of the researcher clinicians who advocated for its inclusion in the DSM – and that insurance institutions provide a legitimate and strategic reason to count it as one.

2.2.2. Summary of Section

I have extended Kukla's examples for their pragmatic approach to mental disorders. The basic idea of Kukla's argument is that, for any approach to classifying mental disorders that includes any one set of necessary and sufficient conditions that apply to all contexts, cases can be generated to show that it fails to count a phenomenon as a disorder in some context where it is

'perfectly appropriate' to. One response to this argument from the naturalist is that we should restrict the institutions that have legitimate reasons to count something as a disorder to only one – an institution based on Boorse's naturalism. I have argued that this response fails as the motivation for it, preventing overmedicalisation, is unfounded. The best response Hybrid approaches can give is that there can be a theory of mental disorders with one set of necessary and sufficient conditions that apply in all contexts which produces the same 'perfectly appropriate' response as Kukla's view. While this seems promising, I have argued that problem cases can be generated for these approaches also. As such, I think Kukla is right – there is no one set of necessary and sufficient conditions that can appropriately be used to decide whether something is a mental disorder in all contexts

Conclusion

The current approach to classifying mental disorders gets a lot of attention in popular media. With statistics showing that one in nearly eight people have a mental disorder, some people think that psychiatry is overstepping its bounds and wrongfully medicalising normal behaviours (Horwitz & Wakefield, 2007) and traits that we normatively disapprove of (Charland, 2006). The spirit of our time — scientism — suggests that the right way to resolve this problem is to look to the natural sciences for a way to classify disorders. Three approaches within the literature suggest different ways in which the natural sciences might be relevant to classification: a naturalistic approach, like Boorse's (1977), argues the natural sciences can provide us with the necessary and sufficient conditions required; hybrid approaches, like Wakefield's (1992) and Tsou's (2021), argue it can provide a necessary condition but alone is not sufficient; and pragmatic approaches, like Kukla's (2022), argue that there is no one set of necessary and sufficient conditions that can be used appropriately in all contexts.

I have argued that the pragmatic approach is to be preferred on the grounds that no one set of necessary and sufficient conditions for all contexts, be they based on the natural sciences or elsewhere, suffices for the different institutional needs of disorder classification. I defended this argument against two objections. The first was that we should restrict the number of institutional contexts we consider. Doing so would not only allow us to generate necessary and sufficient conditions for what should count as a disorder but also prevent overmedicalisation. However, this argument is fallacious — restricting the number of institutions to only one (an institution based on Boorse's naturalism) does not help prevent overmedicalisation. The second objection was that cases like Vincent and Stuart's do not support pragmatism. Rather, they are consistent with a hybrid approach. However, while hybrid approaches can respond to this particular case, other cases can be generated that do pose problems for these views and support the pragmatic conclusion.

In conclusion, deciding what should count as a mental disorder is an important and difficult task. Mental disorders play an important role in the lives of individuals and institutions and it is vital that the approach to classifying disorders gets things right and prevents against overmedicalisation. While one might think that the natural sciences are best suited to this role, analysing various cases reveals this is not so. While the natural sciences have their role to play in deciding what counts as pathological from the perspective of medicine, they are not sovereign. In deciding whether a phenomenon counts as a disorder, one must not just consider whether it is appropriate from a medical science institutional perspective, but, depending on the context, whether it is appropriate from the perspective of other institutions like insurance institutions and socio-political institutions.

References

Abi-Dargham, A. (2004). Do we still believe in the dopamine hypothesis? New data bring new evidence. *International Journal of Neuropsychopharmacology*, 7(1), 1-5.

Aboraya, A., Rankin, E., France, C., El-Missiry, A., & John, C. (2006). The reliability of psychiatric diagnosis revisited: The clinician's guide to improve the reliability of psychiatric diagnosis. *Psychiatry (Edgmont)*, 3(1), 41-50.

American Psychiatric Association (APA). (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). American Psychiatric Publishing, Inc.

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Publishing, Inc.

American Psychiatric Association (APA), DSM-5 Task Force. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing, Inc.

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of science*, 44(4), 542-573.

Boorse, C. (2014). A second rebuttal on health. *Journal of medicine and philosophy*, 39(6), 683-724.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. Wilson (ed.), *Species: New Interdisciplinary Essays*. MIT Press, 141–186.

Cox, S. (2018, Nov 8). How I learned to control and silence my aggressive voice. *Hearing Voices Movement*.

Charland, L. C. (2006). Moral nature of the DSM-IV Cluster B personality disorders. *Journal of personality disorders*, *20*(2), 116-125.

Conrad, P., & Slodden, C. (2013). The medicalization of mental disorder. In Aneshensel, C., Phelan, J., & Bierman., A (2nd ed.)., *Handbook of the sociology of mental health*. Springer, 61-73.

Corstens, D., Longden, E., McCarthy-Jones, S., Waddingham, R., & Thomas, N. (2014). Emerging perspectives from the hearing voices movement: implications for research and practice. *Schizophrenia bulletin*, 40(4), 285-294

Cummins, Robert (1975). Functional analysis. *Journal of Philosophy*, 72(20), 741–65.

Drescher, J. (2015). Out of DSM: Depathologizing homosexuality. *Behavioral sciences*, 5(4), 565-575.

Engelhardt, H. T. (1974). The disease of masturbation: values and the concept of disease. *Bulletin of the History of Medicine*, 48(2), 234-248.

Gagné-Julien, A. M., & Bérubé, A. (2023). Should social pragmatic communication disorder be included in DSM-5? On uncertainties, pragmatic considerations, and the psychiatric kind debate. *European Journal for Philosophy of Science*, 13(3), 1-21.

Gill, M. (2023, May 14). The definition of mental health has been widened so much that it's now almost meaningless. *The Guardian.*

Hacking, I. (1995). The looping effects of human kinds. In Sperber, D., Premack, D., & Premack, A. J. (ed.), *Causal cognition: A multidisciplinary debate*. Oxford University Press, 351–394.

Horwitz, A. V., & Wakefield, J. C. (2007). *The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder*. Oxford University Press.

Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual review of clinical psychology*, 6(1), 155-179.

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders?. *Psychological medicine*, 41(6), 1143-1150.

Kingma, E. (2013). Naturalist accounts of mental disorder. In Fulford, K.W.M, Davies, M., Gipps, R. G. T., Graham, G., Sadler, J. Z., Stanghellini, G., & Thornton, T. (ed.), *The Oxford Handbook of Philosophy and Psychiatry*. Oxford University Press, 363–384.

Kukla, Q. R. (2022). What counts as a disease, and why does it matter?. *The Journal of Philosophy of Disability*, 2(1), 130-156

Lippi, D., Bianucci, R., & Donell, S. (2020). Gender medicine: its historical roots. *Postgraduate medical journal*, 96(1138), 480-486.

Murphy, D. (2020) Philosophy of psychiatry. In Zalta, E. N (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 ed.).

Murphy, D., & Woolfolk, R. L. (2000). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, *7*(4), 241-252.

Neander, K. (1991). Functions as selected effects: The conceptual analysts' defense. *Philosophy of Science*, 58(2), 168–84.

Riba, M. B., Sharfstein, S. S., & Tasman, A. (2005). The American psychiatric association. *International Psychiatry*, *2*(9), 18-20.

Spitzer, R. L., & Williams, J. B. (1983). The revision of DSM-III. *Psychiatric Annals*, 13(10), 808-811.

Szasz, T. S. (1960). The myth of mental illness. *American psychologist*, 15(2), 113-118.

Tsou, J. Y. (2021). *Philosophy of psychiatry*. Cambridge University Press.

Tsou, J. Y. (2016). Natural kinds, psychiatric classification and the history of the DSM. *History of Psychiatry*, *27*(4), 406-424.

Thyer, B. A. (2015). The DSM-5 definition of mental disorder: Critique and Alternatives. In Probst, B. (ed.). *Critical Thinking in Clinical Assessment and Diagnosis. Essential Clinical Social Work Series*. Springer, 45-68

Wakefield, J. C. (2007). The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World psychiatry*, 6(3), 149-156.

Wakefield, J. C. (1992). Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. *Psychological review*, 99(2), 232–247.

Wakefield, J. C. (1992b). The concept of mental disorder: on the boundary between biological facts and social values. *American psychologist*, 47(3), 373–388.

Washington, N. (2018). Contextualism as a solution to paternalism in psychiatric practice. *Philosophy, Psychiatry, & Psychology*, 25(4), 235-243.

World Health Organization (2022, June 8). Mental disorders. *World Health Organization*.

Zachar, P. (2014). *A metaphysics of psychopathology*. MIT press.

Zachar, P., & Kendler, K. S. (2012). The removal of Pluto from the class of planets and homosexuality from the class of psychiatric disorders: a comparison. *Philosophy, Ethics, and Humanities in Medicine*, 7(1), 1-7.

Zisook, S., & Kendler, K. S. (2007). Is bereavement-related depression different than non-bereavement-related depression. *Psychological Medicine*, 37(1), 779-794.

Zisook, S., Shear, K., & Kendler, K. S. (2007). Validity of the bereavement exclusion criterion for the diagnosis of major depressive episode. *World Psychiatry; Official Journal of the World Psychiatric Association*, 6(1), 102-107.