Conceptual Distress:

the constitutive problem for accounts of mental disorder

Elyse C. Oakley

Charlottesville, Virginia

B.A. *Philosophy*, Rhodes College, 2014

M.A. *Philosophy,* Northern Illinois University, 2017

A dissertation presented to the Graduate Faculty of the University of Virginia for the Degree of Doctor of Philosophy

Corcoran Department of Philosophy

University of Virginia

July 2023

This one's for me.

Acknowledgements

I could not have done this without Elizabeth Barnes. As my advisor, she was unfailingly patient, unwaveringly kind, and always met me exactly where I was. It was her initial graduate course in Philosophy of Health and Healthcare that inspired me to switch gears from modality to focus on metaphysics of mental health. Her teaching can and does change the trajectory of lives, as my own case shows. It is Elizabeth's model that I hope to follow in my own classes. Thank you for believing in my work and in my teaching. Thank you for being available to help me defeat raccoons in the marketplace of ideas and to navigate the general weirdness that comes with a long graduate career. And thank you for sharing your dogs with me—I'd be particularly remiss not to mention the instrumental role that Breccan played in my degree progress. It is the least I can do to immortalize him here as the Best Boy.

I also owe a great deal of my success in finishing this project to Zachary Irving, without whom I might have given up on this endeavor entirely. Zac never once failed to meet my ideas with an infectious enthusiasm that reignited my own passion for my projects. It means a great to deal to have been on the receiving end of such generosity of spirit, time, and effort.

Without my spouse, Jonathan Cavell, this dissertation would have never even started. His indefatigable patience, his grace and care when I was my most grotesquely Academic self, and his constant demonstration of love without conditions made all this possible. No marriage should have to endure a dissertation. During my own we also weathered years apart, his master's thesis, three moves, a global pandemic, and so many this-will-be-funny-in-the-future's. Now, we can finally sit down and finish Tears of the Kingdom together. Thank you for waiting.

I think this is supposed to be short, but it's my show so here's the rest: Thank you to Tyg for being the best cat a girl could have ever bought off a sketchy craiglist ad, and for always forcing me to keep things in perspective with aggressive cuddling. Howl and Faye, you both are great cats but let's be real, you love your Dad most anyways. I'd like to thank my Granddaddy for setting the example; he was one of the best teachers in the game. Thank you to my mother-in-law, Margaret Saunders, for housing us during the entirety of the pandemic and keeping us fed for the low price of painting her whole house. Best deal I've ever made. Thanks to my Grammie for the surprise boxes of food—you can't do good philosophy hungry, and I certainly do better philosophy when it is good food! And last but not least, thanks to my Mom and Dad, for telling me early and often that I could do anything I put my mind to. Bet you didn't think this would be it!

Index

Introduction

This project started, as most of mine do, with a bit of hubris. I was going to "fix the DSM." I had it all figured out. These psychiatric diagnostic categories were simply waiting for the under-formed contributions of an early-stage *philosophy* doctoral student. I was, of course, out of my depth. And working a little too far outside my discipline.

But hubris has a way of being somewhat motivating, and I spent many nights with my digital copy of the Diagnostic and Statistical Manual for Mental Disorders (DSM) pulling her apart at the seams. The goal was broad but the method was simple. You always start with an understanding of the material you aim to dismantle. Except, the more time I spent with the DSM, the more confused I became. We use this thing? Is this not alarming? I was only about a third of the way through the manual when I counted that over half the disorders I had examined listed "normal variance" as their primary differential diagnosis. Had no one raised any alarm bells for this?

That's when I shifted my research focus to modern critical writers like Allen Frances, many of whom had in fact raised this alarm and more. But I was unsatisfied with his critical analysis of the diagnostic manual—he had been a primary contributor and editor up until the most recent editions after all. Could he not see the hazards coming while he was driving the car? Moreover, it seemed like most modern criticisms were driven in large part by criticism of the way in which the diagnostic categories, and applications of them, are influenced by pharmacological or insurance-based interests. Many criticisms of the diagnostic manuals were missing critical examination that portioned off these concerns and looked solely at the conceptual problems borne out in our best diagnostic categories. And conceptual analysis was firmly in my discipline, and well within my wheelhouse.

There is a reason why philosophy is called the queen of the sciences—she informs us how to best ask and answer questions about the most basic elements of our disciplines. It was my view that

we needed a little conceptual clean-up in our existing critical analysis of these mental disorders. While many have devoted pages to the critical project, these projects ranged from attempts to provide models of disorder, to accounts of disorder, to focused critiques of the social elements of disorder. And it was my view that all of these accounts had so far failed to capture the phenomena of mental disorder in a convincing and complete way. We have not yet succeeded in giving an account of what, in fact, mental disorders *are*.

So my project was reborn—a newly remade sympathetic but skeptical project. I know that mental disorders are real. I know that the general sentiment about mental disorders has taken a long time to shift to this as the default position. So I worried that a skeptical project would do the same work that the anti-psychiatric movement did in dismantling the legitimacy of mental disorders. It was important to me that I start off small(er than my original hope). I thought it would be best to start with modern accounts of mental disorder that fail to capture mental disorders generally—and fail by their own metrics. So I started off with one account in particular, a causal mechanistic account of mental disorder, and examined it.

In my first chapter, I undertake this project. But it was clear from the outset that there were some unacknowledged assumptions that were at play. In my reviews of other accounts of mental disorder, at least one assumption became salient to me that was left widely unstated. This assumption was couched in the particular way that we engage with and interpret our diagnostic categories. My first chapter, therefore, begins my project with an argument that demonstrates this implicit commitment and how it is an unavoidable element in all our attempts to unpack the nature of mental disorder. Since we have to start with our diagnostic manuals, we are starting with a certain set of background commitments that inform our accounts of what mental disorders are. It was my view that these background commitments were being glossed over. The very manuals we use, like the DSM, come

with commitments built into their understanding of the diagnostic categories. Why, then, does it seem like we commonly set them aside?

Bringing these commitments to the forefront, I argue that we are left with only one way to interpret our diagnostic categories: literally. If we want to operate from the assumption that mental disorders, whatever they are, are *real,* then this interpretative position is the only one that fits with both our goals and the goals of our diagnostic manuals. This interpretative stance becomes a running theme throughout all my arguments. In fact, I eventually conclude in my last chapter that this interpretative position is also the root of the conceptual trouble we ran into throughout all my arguments. But in this first chapter, I merely conclude that it is the root of a particular problem for our diagnostic categories: the problem of reification. Then I turn to examine whether a modern account of mental disorder can, as it purports to do, validate mental disorders while avoiding the reification problem.

Ultimately, I dismiss one modern account of mental disorders that appeals *only* to biological mechanisms for validating mental disorders. I argue that, since social features and mechanisms cannot be cleanly disentangled from mental disorders, social mechanisms must be regarded as validating ones as well. The point is not a new one, I simply demonstrate that it applies here too as well. The problem that arises after the inclusion of these social mechanisms is that, even with both social and biological mechanisms at play, we cannot offer principled methods for validating mental disorders. And so, at the end of chapter one, I conclude that we cannot avoid the reification problem.

Since it seems like the problem of reification is unavoidable, I thought it might be time to turn attention toward accounts of mental disorder that do not seek to provide validating mechanisms. The project of validation is one that, should we be able to provide a successful account of the nature of mental disorder, should come as a consequence of this success. In other words, perhaps our conceptual issue was that we were starting with the problem of validation without any clear account

of *what* we were validating. And when we ask *what* we are validating when we talk about mental disorders, it seems that the received view of mental disorders answers: mental disorders as natural kinds.

In my second chapter, I argue that modern accounts of mental disorders as natural kinds fail. I argue this because it seems that much of what connects mental disorders to natural kinds talk is a pre-theoretical understanding of the theoretical structure of kinds. I suggest this diagnosis of the nature of mental kinds is misapplied. Part of what makes the importation of the kinds framework fail for mental disorders is that we must take our diagnostic categories literally. Included in these diagnostic categories is an explicit relationship between the symptoms that one must meet and something I call *the distress criterion*. Suffice to say, for now, that the requirement of the distress criterion is prohibitive of mental disorders *ever* meeting the minimum criteria of our most amenable account of natural kinds. And that is indeed what I conclude in chapter two, along with the further conclusion that, since this most amenable account is the only one that was suitable for mental disorders, mental disorders cannot be natural kinds. Again, we see the literal interpretation of our diagnostic manuals playing a pivotal role, albeit more quietly in this argument than the first.

So now, the question became: *can* we give an account of mental disorder? There are many candidates that could unpack the nature of disorder in lieu of natural kinds. In my third chapter, I took up an account that I thought used only the elements of mental disorder that were already present in our diagnostic manuals and imported as little theoretical framework into disorders as possible. The goal was to examine whether this most deflationary account of mental disorder could capture the phenomena and avoid the trouble that importing philosophical scaffolding raised for previous accounts.

But again, here the assumption of the literal interpretation of diagnostic categories posed a problem. The only way that an account like this one could tell the difference between cases of disorder and just normal variance was to include another element essential to disorder. Including this element would violate the initial assumption that there are, and literally must only be, two component parts of disorder: the symptoms and the distress criterion. So the commitment to the literal interpretation of our diagnostic manuals again proved prohibitive of another account of the nature of mental disorders. I conclude in chapter three that we must abandon this account rather than our commitment to the interpretative position. However, I also concede that the failure of this account, and likely any other account, will stem from our commitment to this interpretative position. Without running each and every possible account of mental disorder from beginning to ultimate end, though, it is hard to say for sure whether this interpretation of our diagnostic categories is the root of our conceptual problems. Though I strongly suspect it will be the case.

Even so, I suggest we abide by the commitment to literally interpret our diagnostic categories. We should remain committed to it since it is the only option that is commensurate with both the view that mental disorders are real and aligned with the goals stated by the diagnostic manuals themselves. It is the best we can do with what we have. And though the product of my project is far from my far-fetched starting point, I still count this conceptual tidying as a success. If I am correct about the relationship between the interpretative position and our accounts, then we should be happy with *any* conceptual clean-up we can get here. I hope that, in what follows, I will convince you of the same. Don't mind the mess.

I.

The status of mental health diagnostic categories remains contested among philosophers and mental health professionals. This dispute deepens with each edition of the primary diagnostic manual for mental illness. With the recent shift to precision medicine, the need to establish the status of diagnostic categories by giving a specific and principled method of determining which diagnostic categories are scientifically legitimate is growing.[1] I argue that this problem, exemplified in what we know as the reification problem, persists because we must opt for a specific interpretive method of our diagnostic manuals. This paper suggests that there is not currently a way to avoid or resolve this problem by showing that: (§2) attempts to interpret diagnostic manuals in a way that avoids this problem for the status of diagnostic categories undermines the core purpose of the diagnostic manuals; (§3-5) a modern alternative diagnostic manual merely pushes this problem back for determining scientifically legitimate diagnostic categories; (§7-11) an account of diagnostic categories that appeals to causal mechanisms also fails to provide a principled method for selecting out scientifically legitimate diagnostic categories because it ignores the causal power of social mechanisms; and that (§12) even amending such a causal mechanistic account to include such mechanisms still falls short of resolving the reification problem.

Before I get started, I should note that I am going to take it for granted, throughout, that we ought to regard mental and physical health as importantly distinct. Moreover, on the basis of this assumption, I will take it for granted that the collapse of mental health into physical health is to be avoided wherever the threat of reduction of mental health to physical health is a result of some philosophical theory. There are many reasons, both psychological and philosophical, that I want to avoid this reduction, but I will mention one practical reason here. As Strawbridge et al. puts it,

[1] Tabb (2017), pg. 5.

"Although psychiatry has a disease-related burden greater than any single other medical diagnostic category, a disparity of esteem is still apparent between physical and mental health across many domains including research funding and publication."[2] Such disparity of esteem is ultimately symptomatic of a tendency to regard mental health as, at worst, less real than physical health (which we might tie to Strawbridge et al.'s concern with esteem) or, and only slightly better, that the realness of mental health derives from its capability to be tied to physical health alone.[3] With this assumption in mind, we can move on to the task at hand.

1] The Diagnostic and Statistical Manual

The *Diagnostic and Statistical Manual* (henceforth, DSM) is a tool for mental health professionals. The DSM enables mental health professionals to give more uniform diagnoses and treatments to their patients by providing diagnostic criteria for each disorder that the patients must exhibit to be considered disordered or dysfunctional. We consider the DSM purely descriptive because it lists the criteria for membership in a particular diagnostic category; it describes the conditions that a patient must meet for diagnosis. Since its introductory edition in 1952, the DSM has undergone four major revisions. The most recent major edition, the DSM-V, was published in 2013.[4]

2] The DSM's Discontents: diagnostic interpretation and the problem of reification

Recently though, there has been much concern about the DSM and its purely descriptive state—even after these revisions. This concern is about the status of diagnostic categories, and it arises from interpreting the DSM's descriptive categories in a particular way.

---

[2] Strawbridge, R., et al., (2017).

[3] There are, to me, several compelling arguments for this position. I do not have the space to offer them here in full, but some of the considerations I offer in §10 can serve as support for this view.

[4] APA (1952), (1965), (1980), (2000), (2013).

2.1) The constitutive interpretation of the DSM: Consider depression. The DSM has a list of criteria that mental health professionals use to identify depression in their patients—anhedonia, changes in eating and sleeping habits, suicidal ideation, etc.[5] One interpretation of the diagnostic category, then, is that depression simply *is* the list of criteria. Since the list of criteria constitute the mental health diagnostic category, we can call this interpretation of the contents of the DSM the *constitutive position.*[6] The constitutive position "locks in our definitions" of mental disorders, such that we can consider the criteria literal and definitive.[7] This kind of position is sometimes referred to as *diagnostic literalism*.[8] But choosing the constitutive interpretation of the DSM leads us to the primary problem with the DSM: the reification problem.

2.2) The reification problem: The reification problem concerns the status of diagnostic categories. The status of diagnostic categories in question is whether these categories are *valid* in the psychiatric sense; whether such categories are "accurate in representing the true state of nature."[9] This notion of psychiatric validity is, more specifically, external validity. The problem with this psychiatric conception of validity is that it requires that these categories exist independently of those who develop and use the DSM. Whatever a diagnostic category is, for it to be valid it must be as it is regardless of our interaction with it. But, as Hyman points out, these diagnostic categories do not exist independently of those who use and create them.[10] Most importantly, our critical evaluation of these categories relies upon the assumption that these categories reflect the state of nature until proven psychiatrically invalid. To put it alternatively, the reification problem is that we consider these diagnostic categories

[5] APA (2013).
[6] Kendler (2017).
[7] Ibid, pg. 2059.
[8] Zachar (2014).
[9] Blacker and Endicott (2000), pg. 7.
[10] Hyman (2010).

real entities in the metaphysical sense and we have no principled method for determining whether they are or are not. Why?

We use the DSM's diagnostic categories in "the generation and presentation of research studies" which in turn obscures our research findings.[11] Such use also prevents discovery of underlying pathophysiological mechanisms which causally contribute to diagnostic categories.[12] How? We design research studies and select participants on the basis of the DSM criteria for categorization. So, our evaluation of these categories relies on their use. So, the worry is that "if the DSM's categories were not valid, it would be difficult to make headway in discovering the mechanisms causing psychopathology. Few causally relevant mechanisms would be universally shared by the heterogenous patients sharing a diagnosis."[13] Furthermore, when we generate experimental design using the DSM we strongly influence the aspects of the disorder that we study. Hyman notes this influence in the lack of empirical study on the cognitive deficits that schizophrenic patients experience; he argues that such was due to the fact that studies of schizophrenia were created to target diagnoses that used the DSM's criteria, and those criteria did not include cognitive deficits at the time.[14] In short, when we use the DSM's categories for experimental design, we not only assume that the category picks out something real, but we further reify that category as it is currently given by using experimental parameters that occlude the kind of discovery which motivates category revisions. Thus the problem of reification.

2.3) The second interpretation of the DSM: A mental health professional might respond to this concern by stating that they are not interested in whether or not the DSM categories are valid in that they reflect

[11] Tabb (2017).
[12] See Zachar and Kendler (2017) and Tabb (2015).
[13] Tabb (2017). See also Tabb (2015).
[14] Hyman (2010).

the true state of nature.[15] Instead they are only useful tools for picking out entities that are responsive to treatment. This is the second interpretation of the DSM's diagnostic categories.

On this alternative interpretation we instead assert that the lists of criteria for any diagnostic category merely "index or reflect the disorder," which is itself only a "hypothetical diagnostic construct."[16] Again, take depression. Depression *might* be diagnosed because the patient meets the criteria listed in the DSM, but perhaps not diagnosed *solely* on those criteria. The patient might have anhedonia and a change in appetite, but these symptoms are not the only features upon which the professional draws for diagnosis. On this interpretation, depression can be something more than just the criteria listed in the DSM. The diagnostic criteria listed in the diagnostic category is just a helpful guide, or heuristic, for diagnosis. Departing from established literature, I will call this interpretation of the DSM diagnostic categories, and their respective diagnostic criteria, the *heuristic position*.[17]

Certainly, the heuristic position avoids the reification problem—it leaves conceptual space for revision of diagnostic criteria or the denial of categories altogether in the future. But it is not obvious it avoids the "independent of use and development" concerns that Hyman thinks is part and parcel of the reification problem. Even if we bracket such concerns and concede that the heuristic interpretation avoids the reification problem, it does so at the cost of introducing further problems for the DSM's diagnostic categories.

---

[15] In fact, on several occasions when I have discussed these concerns with mental health professionals of various credentials, this was approximately the response I received. But I argue here shortly that these individuals are attempting to remain agnostic about the reality of these categories while necessarily committing to them. Arguments for a middle position, like that of van Loo and Romeijn (2015), also seem to necessarily commit to a realist view, despite claiming that their view is not fully so.

[16] Kendler (2017), pg. 2054.

[17] Kendler (2017). Kendler, from whom I draw this distinction, calls this interpretative position the *indexical position*. However, "indexical" is a loaded philosophical term with a history of associated terminological baggage that would complicate matters unnecessarily if I maintain his original use here.

2.4) Why the second interpretation fails: The cost of avoiding the reification problem is a steep one: internal validity. For the DSM internal validity is inter-rater reliability, which ensures that a patient presenting with the same set of symptoms would be diagnosed near uniformly among mental health professionals. To put it alternatively, internal validity ensures that individuals presenting with similar symptoms will be classified such that each are members of the same diagnostic category regardless of who diagnoses them. The DSM was compiled and revised explicitly for this purpose in the first place.[18] The easiest way to see how the heuristic interpretation of the DSM fails to uphold internal validity is to examine how the constitutive interpretation is well suited for the task. Why? It is evident that the constitutive and heuristic interpretations of the DSM are mutually exclusive. I cannot assert that the DSM gives definitive criteria for mental health diagnostic categories and acknowledge that diagnostic categories might not be fully captured in such criteria. If it is the way that the constitutive interpretation treats diagnostic categories and criteria that make it well suited to upholding internal validity, then of course the heuristic interpretation will fail to do so.

A mental health professional must use the tools she has to make diagnoses. When she uses the DSM, she determines whether her patient has a particular disorder if they meet the criteria listed in the diagnostic category. Whenever a mental health professional makes a diagnosis on the basis of a patient meeting a set list of criteria for a diagnostic category, she inherently interprets the DSM constitutively. Why? Because she is using the fact that the patient meets the diagnostic criteria as determinative of whether the patient has a disorder. Whether she agrees the diagnostic criteria *in fact* constitute the disorder is irrelevant here—her use of these criteria requires *at least* the momentary commitment that these criteria constitute the disorder. And so does each individual use of those

[18] Frances (2013).

criteria to diagnose that disorder thereafter. Since those criteria are the ones used by mental health professionals globally, it is this feature of the constitutive position that allows for internal validity.

If the motivation for the inception of the DSM is enough to support the constitutive over the heuristic position alone, then we need not worry about the presence a second possible interpretation of it. However, any decision as to whether the DSM is best interpreted constitutively must also consider reasons why the heuristic interpretation does not work for the DSM's aims. Recall that the heuristic position allows mental health professionals to consider features outside the diagnostic criteria listed in the diagnostic category for making diagnoses. Is there a way for mental health professionals to interpret the DSM heuristically *and* maintain internal validity? If the heuristic interpretation can maintain internal validity, then mental health professionals must agree on the features outside those listed in the DSM's diagnostic criteria that lead to diagnosis. But if such consensus obtains about these additional features that are also determinative of diagnosis, then it is not clear how these features would differ from those diagnostic criteria that the DSM lists. In fact, it seems the only difference is that some of the criteria used for diagnosis are listed in the handbook, and others are not. Moreover, the heuristic position requires that—whatever the diagnostic features of a category are and regardless of whether they are all capable of being listed among the diagnostic criteria for a disorder—the heuristics that point to the diagnostic category must be taken to reflect some aspect of it that tracks some intervention-responsive phenomena. Otherwise, these features would not be useful for diagnosis. Since uniformity in diagnosis is the aim of the DSM, as long as there is consensus in the heuristic position then it seems to, in principle, also fit the aims of the DSM. But it does not seem that there could be consensus and uniformity in diagnosis with the heuristic position. Why?

The heuristic position not only leaves open the possibility of, but also *justifies*, differences in what mental health professionals determine regarding diagnosis for the same patient presenting the

same symptoms. The heuristic position justifies diagnostic differences by allowing that a disorder can be determined by additional factors than those in the DSM. If there is no regulation on what those additional factors are, like the minimal regulation given by the descriptive list of criteria in the DSM, then those factors could differ between professionals. Whatever the failings of the DSM are that might allow for such differences in diagnosis on the basis of *just its listed criteria*, its aim is *prevent* such diagnostic disparity, not enable or even justify it. So, the heuristic interpretation of the DSM seems inapt for its goals. Because we have found the heuristic position incapable of motivating the internal validity that the DSM aims to provide, we must opt for the constitutive position and arrive back at the problem of reification.

3] The Traditional versus the Dimensional Psychiatric Model

But is taking the constitutive position really to blame for the persistence of this problem? Perhaps instead the reification problem arises from our treatment of disorders as diagnostic categories in a particular manner—we assume that there is a *qualitative* difference between normal and abnormal, and that there is a clear and objective boundary between them. We might call this assumption, following Zachar and Kendler, a premise of the "traditional psychiatric model" which is "composed of categories."[19] However, a recent shift in the mental health profession opposes viewing diagnostic categories as so clearly, objectively divided. Instead, the working assumption is that the difference between normal and abnormal is *quantitative* and, thereby, admits of degrees. That a clinical entity can admit of degrees leads to (potentially) fuzzy or vague diagnostic boundaries.[20] Such an assumption is an important premise of *the dimensional psychiatric model.*

It is not at all clear that the DSM is committed to the traditional psychiatric model over and above the dimensional one. In fact, the portions of the DSM-V which require professionals to rate

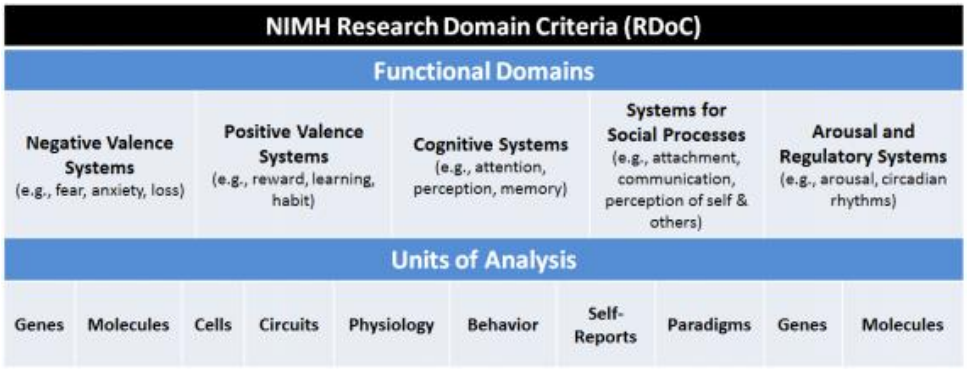
[19] Zachar and Kendler (2017), pg. 50.
[20] Zachar and Kendler (2017).

severity of *diagnostic criteria* on a sliding scale seems to obviously allow that differences between normal and abnormal admit of degrees. But it is fair to argue that, since this is the only dimension that allows for quantitative degrees of disorder, the DSM views *diagnostic categories* as qualitatively determined entities with otherwise clear boundaries. So, I will assume the dimensional psychiatric model remains orthogonal to the traditional model used in the DSM. So, despite both being constitutive accounts, the dimensional psychiatric model offers a direct challenge to the very assumption upon which the DSM lies. But does it avoid the reification problem? In other words, can it explain the validity of some diagnostic categories while ruling out others? The only way that the dimensional psychiatric model can replace the DSM is if it resolves the reification problem without raising further problems for diagnostic categories. To answer these questions, let's look at a specific version of a dimensional psychiatric model, the research domain of criteria.

3.1) The dimensional psychiatric model and the RDoC: The Research Domain Criteria (RDoC) project is an initiative started by the United States National Institute of Mental Health that culminated in a new research protocol for mental health phenomena.[21] The RDoC gets its name from the cluster of research domains which target different aspects or dimensions of mental function. One of these domains is called "cognitive systems." The domain of cognitive systems is broken down into "constructs" (e.g. attention, working memory, etc.).[22] These are measured, of course, quantitatively, through a various array of units of analysis, listed in the figure below.

[21] It should be noted that there is more than one articulation of the dimensional psychiatric model. Importantly, however, these articulations all rest on the same assumptions (more or less) and so it is likely that any of my arguments against the dimensional psychiatric model given in the RDoC project will also extend to the dimensional models used in other formulations.

[22] National Institute of Mental Health. (2020a).



Note that none of the domains are considered "diagnostic categories" like those reported in the DSM. Instead, these domains give lists of constructs, which are themselves so grouped for their relevance to that particular domain. Thus, the RDoC "changes the targets of validation from the DSM disorder to any sort of phenomenon that may be viewed either as an extreme on a spectrum of human variation or as a dysfunctional structure or process."[23] In other words, the object of validation shifts from a diagnostic category to individual constructs within a research dimension.[24]

3.2) The RDoC and moving the goalpost: So, does the RDoC solve the reification problem? That remains to be seen. By changing the target of validation, the RDoC may not provide a *solution* to the reification problem for the DSM so much as a hopeful circumvention. The RDoC may simply offer an alternative method—namely, the dimensional model—for categorization without settling the matter about whether such categories are externally valid.

4] Is the Dimensional Model of the RDoC a Solution to the Reification Problem? No.

While the DSM retains diagnostic categories as the objects of validation, the RDoC focuses on the underlying mechanisms for *constructs* which are the objects of validation. But the RDoC, as an exemplar

[23] Tabb (2017), pg. 5.

[24] Tabb (2017), pg. 6. In fact, Tabb writes that "the RDoC's proximate goal is to classify proper targets for psychiatric research, and its distal goal is to inform diagnostics." The idea is that if researchers can start grouping together biomarkers (i.e., the underlying mechanisms which causally contribute to an individual's quantitative placement on the continuum from statistically "normal" to "abnormal" within various dimensions), then researchers can use these subgroups to inform treatment plans based on demonstrably successful interventions relative to that biomarker. These treatment plans will then inform the construction of clinical handbooks for mental health diagnosis.

of a dimensional psychiatric model, does not avoid the reification problem just because it allows for fuzzy diagnostic boundaries, quantitative measurement, and changes the object of validation. In fact, I suggest that the reification problem stands for the RDoC's dimensional model as well. Why?

By shifting the object of validation, the RDoC asks whether the constructs are "accurate in representing the true state of nature."[25] The validation of the constructs comes in terms of their determination by statistical variation within a specific unit of analysis. Since these units of analysis are measurable quantitatively, it seems that researchers would happily commit to the "realness" of these mechanisms' effects in the ontologically heavy sense: they do reflect the way the world is. As a result, the RDoC project merely pushes back the reification problem onto the constructs via the mechanisms measured in the units of analysis.

5] The Reification Problem and the Dimensional Psychiatric Model

Take depression again. For some time, we thought that serotonin dysregulation in neurocognition was determinative of depression, and we had some clinical results that supported this view.[26] Now we have research that shows such dysregulation is not determinative of depression, and so we no longer regard serotonin as a biological mechanism, or biomarker, of depression.[27] In fact, the specific neurological and biochemical pathways involved with depression (if any) may not be as isolatable as previously thought; neural networks prohibit extreme isolation of any given neurotransmitter for truly independent analysis. So, any previous findings about the relationship between serotonin and depression likely involve mediating neurobiological chemicals which temper our conclusions from

[25] Blacker and Endicott (2000), pg. 7.
[26] See Young et. al (2016) and Lichtblau et. al (2013).
[27] See Boska (2013) and Cowen and Browning (2015).

experimental results.[28] Thus, "[such] evidence suggests that impair[ed] serotonin function can cause clinical depression in some circumstances, but [it] is neither necessary nor sufficient."[29]

If we think that a diagnostic category is externally valid whenever its symptomology is tied to dysregulation in biomarkers, then depression loses its status if it loses this tie. If we need to preserve depression as a valid diagnostic category—and it seems the general consensus that we do—then we have to preserve it by assuming *there must be some biomarker which so validates it.* Then the idea is that we just have yet to discover this underlying biological mechanism for depression which will confirm what we already know about its status as a diagnostic category. But such an assumption is to play into the reification problem.

I have so far put this concern in terms of diagnostic categories which is the object of validation in the DSM, but how is this a problem for the RDoC's object of validation: constructs? Take the construct of anxiety in the negative valence systems domain. While anxiety is a diagnostic category in the DSM, described in terms of diagnostic criteria, the construct of anxiety is a description of phenomena that is measured via units of analysis. Researchers examine the relationship between manipulation of variables within the various units of analysis under the construct of anxiety like cortisol levels, and determine which among them demonstrate determinative deviation from the statistical norm in participants. Then they can group those patients together who share the same kind of deviation along the same unit of analysis.

But there are few problems with this overall model. The first problem is that using statistical deviation for determining whether an individual is healthy has met with significant rebuke, none more

[28] Cowen (2008) and Cowen and Browning (2015).
[29] Cowen and Browning (2015).

so than Christopher Boorse's biostatistical theory of health.[30] For example, the biostatistical theory cannot determine whether athletes, who demonstrate incredible deviations from physical statistical norms, are healthy, despite our general inclination to think so. Mere statistical deviation alone is not enough to determine health status for accounts of physical health,[31] let alone the increasingly evasive accounts of mental health.

The second problem is that the degree of specificity required to group patients together to inform diagnostics would be too high; Kingma raises this kind of argument against Boorsean accounts of health.[32] To see how this problem plays out imagine the following. I might match another patient in terms of my deviation from the norm with regard to serotonin levels but not with regard to either self-report or my genetic predisposition. All of these are aspects of units of analysis within the RDoC matrix, so should we be grouped and receive the same treatment on the similarity of variation in only one unit of analysis? It is not clear across how many units of analysis I must converge with other patients to create a group for treatment.[33]

The last problem, and most important to my project here, is that the RDoC seems to have the same structural relationship between its constructs and units of analysis that the DSM has to its diagnostic categories. Consider again the construct of anxiety within the domain of negative valence systems. This construct captures the phenomena of

> Activation of a brain system in which harm may potentially occur but is distant, ambiguous, or low/uncertain in probability, characterized by a pattern of responses such as enhanced risk

[30] See Christopher Boorse (1987) and (1997) for more on the biostatistical theory of health. I do not have the space here to offer explanations of such rebukes, and see Kingma (2014), Schwartz (2007), for examples of them.

[31] See again Kingma (2014) and Schwartz (2007).

[32] Boorse (1997) for accounts of reference classes, and Kingma (2014) for a rebuke of them.

[33] This is a distal goal for the RDoC but it is an important one if it is to replace the DSM entirely. It is not as important if the RDoC is a mere supplement to the DSM, but that would raise new issues about how to integrate the two classification systems.

> assessment (vigilance). These responses to low imminence threats are qualitatively different than the high imminence threat behaviors that characterize fear.[34]

The RDoC is ready-made to admit that it may not be the case that one unit of analysis alone will validate the construct and the phenomena it captures. However, it is also ready-made to commit to the assumption that at least one aspect of at least one unit of analysis will.

If we think that a construct is externally valid whenever it's described phenomena is tied to determinative statistical variation in units of analysis, then the construct loses its status if it loses this tie. If we need to preserve anxiety as a valid construct—and it seems the general consensus that we do—then we have to preserve it by assuming *there must be some unit of analysis which so validates it.* Then the idea is that we just have yet to discover which aspect under a unit of analysis for anxiety which will confirm what we already know about its status as a construct. But, again, such an assumption is to play into the reification problem. And while the RDoC is made to be open for revision and rejection of constructs, the constructs were decided by consensus from previously established findings—artifacts of DSM driven research. To date, no revisions have culminated in the *removal* of any of these constructs.

Recall from §3 that the only way that a dimensional psychiatric model can replace the DSM is if it resolves the reification problem without raising further problems for diagnostic categories. I have shown that the RDoC, a paradigm of the dimensional psychiatric model, not only fails to resolve the reification problem, but also introduces further problems for diagnostic classifications. Thus, the RDoC can neither replace nor even supplement the DSM. But now we return, again, to the DSM and the persistent reification problem.

[34] National Institute of Mental Health. (2020b).

Maybe the reification problem resists our attempts to solve it not just because we take the constitutive position to ensure internal validity. Maybe the reification problem resists resolution because we have not considered one which focuses on what diagnostic categories *are* rather than how we interpret their listed criteria or model them. Perhaps the solution to the reification problem lies in providing an account of what diagnostic categories are that gives a principled method for ruling in externally valid diagnostic categories and ruling out those that are not.

6] Philosophical Problems for Valid Diagnostic Categories: natural and artificial kinds

Some articulations of the reification problem come as a comparison between diagnostic categories which are widely regarded as "genuine" or externally valid categories (e.g. depression, schizophrenia) and those that are not. Those categories that are not so regarded are considered unpalatably artificial or constructed (e.g. histrionic and narcissistic personality disorders).[35]

Those who make these comparisons argue that we can resolve the reification problem for the DSM by adding a supplementary principled method for ruling in valid diagnostic categories and others out. Recently, philosophers and mental health professionals alike attempt to solve the reification problem by appealing to and identifying underlying causal mechanisms for diagnostic categories. These causal mechanisms supposedly provide the principled method required to select out valid diagnostic categories. These causal mechanisms vindicate the diagnostic criteria in a given category but are not themselves necessarily determinative of the category. Thus, we preserve the constitutive view as criteria literally constituting the disorder but allow a principled method for ruling criteria in or out as partially determinative of disorder.

Such a solution relies upon assumptions about the sort of entity a diagnostic category is. I should note that diagnostic categories are, from the metaphysical standpoint, kinds. Diagnostic

[35] See Tsou (2016), Kendler (2012), and Murphy (2006) for examples of these.

categories are recognized as kinds because any individual who receives a diagnosis should meet some set of criteria for membership. So diagnostic categories are kinds. But what kind of kinds are they? To answer this question, we need to recall how mental health professionals talk about diagnostic categories.

I noted in §2.2 that Hyman argues that the reification problem stems directly from the mental health profession's definition of validity. Recall that a diagnostic category is valid if it accurately represents the "true state of nature."[36] Philosophically, this amounts to calling diagnostic categories *natural kinds*. As Tenkin puts it, "the scientific legitimacy of mental disorders has hinged on their status as natural kinds."[37] Natural kinds can admit of multiple definitions, philosophically, but there are two main views of natural kinds that are at work here.

6.1) Natural Kinds as Essences: The first view of natural kinds holds that kinds are determined by some essential feature which alone determines some individual's membership in that kind.[38] Essentialist readings of natural kinds are easy targets for modern scientific counterexamples,[39] since they offer necessary and sufficient conditions for membership.[40] But the diagnostic categories in the DSM arguably do not give necessary *and* sufficient conditions for membership. Further, it is not the case that many of the diagnostic criteria in a category are even considered individually necessary or alone sufficient.[41] The diagnostic criteria listed for a particular category *never* requires that a patient meet *all and only* the criteria listed, and that only patients with such diagnoses exhibit those symptoms given by

[36] Blacker and Endicott (2000), pg. 7.

[37] Tenkin (2016), pg. 148.

[38] See Ellis (2001), and Haslam (2002) for discussion of this position.

[39] For detailed accounts with such counterexamples, see Dupré (1981), Sober (1994), and Zachar (2000).

[40] See Tabb (2017) and Wilson (2005) for discussions of counterexamples to natural kind essentialism. See Kornblith (1993) for an account of essentialism about natural kinds explicitly formulated in terms of necessary and sufficient conditions.

[41] I can, for example, be diagnosed with depression without exhibiting suicidal ideation, or changes in eating habits—these two symptoms are listed among the diagnostic criteria for the category but are not *required* for a diagnosis of it. And I can have changes in my eating and sleeping habits and not be depressed!

the criteria. So, the view that natural kinds are established via essences seems to be off the mark, both philosophically and because the DSM's diagnostic method is incommensurate with it.[42]

6.2) Natural Kinds as Something Else: Another view of natural kinds simply suggests that kinds do not require such stringent conditions as essences.[43] Before I examine a popular candidate theory of non-essentialist natural kinds (in §7), it will be helpful to look at the things that these views have in common: a few desiderata that, if met, fulfill the *minimal necessary conditions* to support the assertion that some entity is a natural kind.

6.2.1) Projectability and Stability—indifferent versus interactive kinds: It has been said that any successful account of natural kinds will be able to explain why and how we can make inferences about members of that kind solely because of their membership in it.[44] In terms of mental health diagnostic categories, that means that I should be able to reasonably infer that someone has severely reduced interest in daily activities, that they are either not sleeping enough or sleeping too much, they experience suicidal ideation, etc., if they tell me that they are battling depression.[45] This desideratum for natural kinds is a feature sometimes called *projectability.*[46] A kind is PROJECTABLE when it exhibits a degree of regularity that allows us to make these reasonable inductions at a rate greater than chance. In terms of psychiatric conception of diagnostic categories, projectability seems to be tied to internal validity—that future instances of this kind will either share at least some of the symptoms the category describes or interact with interventions in the same way as previous instances of the kind did.

---

[42] It would be a different argument altogether to say that the DSM *ought* to give diagnostic criteria that are so stringent. But such an argument would be an entirely uphill battle that would have to ignore a multitude of research which suggests otherwise. As a result, I will not consider this here.

[43] See Dupré (1981), Haslam (2002), Hacking (1992) and (1995), Tsou (2017), and Zachar (2000), (2008), and (2014), for a few such examples in more detail.

[44] See Boyd (1985) and (2010).

[45] APA (2013).

[46] Tsou (2016).

Where does the kind get the regularity it requires for supporting these reasonable inductions? Some point to the *stability* of kind. A kind is STABLE whenever the presence of one of its features or properties allows the reasonable inductive inference that other features of the kind are also present.[47] It is taken as a hallmark of natural kinds that they exhibit STABILITY. Hacking famously articulates this desideratum for kinds as indifference to *looping effects*. Looping effects occur whenever the object of classification is causally affected by being classified. A STABLE kind is an indifferent kind. Alternatively, kinds with members that are causally affected by their classification are interactive kinds.[48]

Modern iterations of Hacking's interactive kinds require more than the mere presence of looping effects on a kind. The idea here is that looping effects are unavoidable in some classifications, but that their presence need not necessarily alter the object of classification. In our case, I might suggest that psychiatric classifications obviously *affect* their members, but this fact need not alter them. So, the presence of looping effects alone is not determinative of a kind's status as an interactive one.[49] I accept an account of interactive kinds offered by Jonathan Tsou, which states that for a kind to be interactive it must meet two conditions. The first is that a kind demonstrates the presence of looping effects and, the second, that these looping effects uniformly affect and interact with their objects of classification.[50] Interactive kinds lack the regularity necessary for STABLE because they are subject to looping effects that uniformly change the objects they classify; so the presence of one property of that kind cannot allow for reasonable inference to the other properties of that kind since classification of its members alters the *majority* of them. [51]

---

[47] I here draw from the clearest articulation of stability that is given without recourse to specific unifying or underlying mechanisms, structures, or principles, a full account of which can be found in Slater (2015).

[48] See Hacking (1995), (1999), and (2007).

[49] For a detailed argument against Hacking's interactive types as constituted by the mere presence of looping effects, see Tsou (2015). For a brief exposition of his general (2015) argument, see also (2016), pgs. 409-410.

[50] See Tsou (2016), pg. 410.

[51] Tsou writes that for a kind to be interactive it must uniformly alter a *sufficient* amount of its members (2016, pg. 410). I take "sufficient" here to require at least a significant majority. The way of parsing this out is not within the purview of this paper, but it seems reasonable to assume that sufficient should be, at the very least, greater than chance.

When I take this account of interactive kinds, I allow that some kinds are indifferent in the presence of looping effects *so long as these looping effects neither uniformly affected nor uniformly interact with their objects of classification.* This indifference allows for STABILITY, and so meets the desideratum for a non-essentialist account of natural kinds. While Hacking did not intend his distinction to vindicate natural kinds, it is not obviously a mistake to think that natural kinds are *at least* indifferent ones, if they are anything at all.

We might think of STABILITY as a philosophical correlate to the psychiatric concept of external validity. It is something about the kind that allows for our inference from the presence of at least one of its properties to the existence of others; in this case, a diagnostic category is externally valid because it allows these inferences. STABILITY does the individuating work of providing boundaries between diagnostic classifications, while PROJECTABILITY does the diagnostic work by allowing the inference that new diagnoses under established categories share symptoms.

7] Natural Kinds as Homeostatic or Mechanistic Property Clusters

Most modern accounts mental health diagnostic categories agree are variations of a view about non-essentialist natural kinds called cluster kinds.[52] Each of these accounts relies upon Boyd's foundational work on homeostatic property clusters. Let's look at this account first.

7.1) Homeostatic property clusters: Boyd first proposed the theory that homeostatic property clusters (henceforth, HPCs) are what constitute kinds. HPCs are determined by group of properties that are clustered in virtue of their nature. These properties are clustered because they are mutually promoted by a shared function. So HPCs meet the STABILITY desideratum from §6.2.1 because the presence of one property allows the reasonable inference to the potential presence of others for that kind. Membership in an HPC kind is established by having an adequate subset of these properties. No

[52] Tabb (2017).

member of the kind must have any individual one of these properties and no member of the kind must have all of these properties. Given these two allowances, it is entirely possible that no two members of a given kind will share *identical* properties. The important aspect of HPCs is that the properties of the kind must be stable enough to allow for predictions about the members of that kind that are greater than chance.[53] The function that supports the mutual promotion of the properties in the cluster is the feature of the cluster kind that makes the property expression stable in this way. So these functions enable HPCs to meet the PROJECTABILITY desideratum.

7.2) Mechanistic property clusters: Attempts to more explicitly meet these desiderata lead eventually to the adaptation of HPCs into the mechanistic property cluster (henceforth, MPC) account of natural kinds.[54] The MPC account consider kinds clusters that "are constituted by networks of *biological mechanisms* at multiple levels that interact to produce key features of the kind."[55] What are mechanisms? They are generally taken to be "entities and activities organized together such that they do something."[56] Since it is a feature of MPCs that these kinds have shared biological structures (organized entities) that support (doing something) the clustering of their associated properties, these structures account for the "stability of such kinds over time."[57] So MPCs meet the STABILITY desideratum, but exceed its requirements because they identify the mechanisms responsible for stability within property clusters. These clusters can arise out of single mechanisms or many mechanisms, the MPC theorist need not commit to a single mechanism motivating a clustering. Moreover, these mechanisms can be inter- or intra-level mechanisms, and causal or constitutive.[58]

[53] See Boyd (1985), (1991), (1999), and (2010).
[54] The MPC account of natural kinds originates in Craver (2009).
[55] Tsou (2016), pg. 412 (my emphasis).
[56] Craver (2009).
[57] Tsou (2016), pg. 412.
[58] Craver (2009).

Some have argued that some mental disorders are obvious examples of MPC kinds because they are purportedly founded upon a set of stable *biological* mechanisms.[59] Recall that in §6 I indicate that the primary goal of those arguments that appeal to causal mechanisms is to offer a principled method of ruling in valid (i.e. *natural)* diagnostic categories like depression, obsessive compulsive disorder, and schizophrenia,[60] while ruling out others (i.e., artificial ones) like histrionic personality disorder and narcissistic personality disorder. Giving this principled method would, so the argument goes, solve the problem of reification for the diagnostic categories in the DSM. The MPC account does not necessarily require the explicit commitment that the mechanisms involved in stabilizing the kind are causal ones. A causal mechanistic approach to diagnostic categories would. To explore whether an appeal to causal mechanisms in this way can resolve the reification problem for the DSM, let's first examine the approach.

8] A Causal Mechanistic Approach

What is a *causal mechanistic approach* to diagnostic categories? There are many ways to articulate a causal mechanistic approach (henceforth, CMA) to diagnostic categories,[61] but I will examine one offered by Jonathan Tsou.[62] He writes:

> Compared to purely descriptive approaches, the distinguishing features of the approach that I advocate is that diagnostic categories would *incorporate information about the causes of disorders*… —as an ideal—the operational symptoms that constitute the diagnostic criteria are caused—in whole or in part—by identifiable biological mechanisms. According to this ideal, in order to be included in the DSM, there needs to be evidence that there is a *stable and distinctive biological causal structure for classified disorders.*[63]

---

[59] Tsou (2016), Murphy (2009) and (2014), BeeBee and Sabbarton-Leary (2010), and Kendler et. al (2010) do so to varying degrees and ends.

[60] Many theorists list these among the obvious candidates for valid diagnostic categories. See BeeBee and Sabbarton-Leary (2010), Hyman (2010), Tsou (2016), Kendler et. al (2010) for just a few examples of the invocation of these disorders as valid diagnostic categories.

[61] See BeeBee and Sabbarton-Leary (2010) and Hyman (2010) for two such examples.

[62] I think, and this will perhaps become more evident in §9 and §10, that my arguments against Tsou's version of the CMA will apply to any philosopher or mental health professional who tries to avoid the reification problem by solely appealing to causal mechanisms and MPCs.

[63] Tsou (2016), pgs. 417-418.

Essentially, this CMA states that diagnostic categories are externally valid only if there are empirically verifiable, stable and distinctive biological, causal mechanisms that at least partly determine their symptoms. Consequently, this CMA rules out other diagnostic categories wherever their symptoms are not even partially so determined—or, at least whenever there is no evidence to suggest so.

So, can this CMA resolve the reification problem for diagnostic categories? To answer this question, I must examine whether this account offers new conditions or desiderata that conflict with the minimum necessary conditions for kindhood. I turn to this next.

9] More Desiderata for Diagnostic Kinds as Natural Kinds

Tsou's CMA requires that mental diagnostic categories have a biological basis, which requires that biological mechanisms causally determine these diagnostic categories' symptomology. Effectively, Tsou indirectly identifies valid (i.e. natural) diagnostic categories with the biological mechanisms which causally determine their symptoms. But it is not altogether clear that a CMA must select out diagnostic categories solely because they have symptoms underwritten by biological mechanisms. While Tsou's CMA equates natural with biological kinds, such identification is not obviously necessary for a principled method for determining valid diagnostic categories from artificial ones.

In what follows, I parse out desiderata that an account like Tsou's requires above the minimal necessary conditions for natural kindhood given in §6.2.1 without recourse to biological mechanisms. Then (§10) I suggest that at least one of these desiderata is undermined by sole appeal to biological mechanisms vindicating diagnostic categories and so an account that uses it cannot resolve the reification problem. Finally, (§11-12) I examine whether the remaining desideratum together with the minimal necessary conditions for natural kindhood can resolve the problem of reification.

9.1) The COMPELLING EVIDENCE desideratum: Given PROJECTABILITY, any satisfactory account of natural kinds must allow for reasonable inductive inference about the future members of kind. To

demonstrate that a kind meets PROJECTABILITY there must be a plausible, and empirically verifiable, explanation for why the kind allows these inferences. The sort of explanation that such kinds give for the projectability of the natural kind terms is, at least in part, etiological. So, natural kinds identify the presence of at least some underlying structure that causally explains the presentation of a particular symptoms. Whether such structure underpins some symptom is supported by the presence of *compelling evidence*[64]—the presence of multiple lines of independent research converging upon a common conclusion. If there is compelling evidence that disruption or dysregulation of a mechanism causally contributes to a symptom, then such an etiological theory meets what we will call the COMPELLING EVIDENCE desideratum. COMPELLING EVIDENCE is the third desideratum for natural kinds.

9.2) The DISTINCT desideratum: A CMA requires that a symptom in a diagnostic category must be etiologically determined by a distinct mechanism.[65] What does 'distinct' mean? There are at least two readings. The first is that distinct just means identifiable. But that cannot be the case, since distinctness would be built into the COMPELLING EVIDENCE desideratum. Why? When we establish COMPELLING EVIDENCE for a mechanism, we must obviously identify it. So, this first reading cannot be correct.

The second reading interprets 'distinct' more literally. The mechanism that causally contributes to a diagnostic category must be unique to it *and* distinct from that involved in others. Given that our first reading failed, it is this reading of DISTINCT that constitutes the final desideratum for valid diagnostic categories.

[64] This origin of this idea, presented as the consilience of induction, is due to William Whewell (1858).
[65] Tsou (2016), pg. 418.

10] DISTINCT and Biological Commitments Sinks CMA Solution to Reification Problem

I suggest that any CMA that conflates natural kinds with biological kinds of the basis of DISTINCT biological mechanisms is ill-conceived. Why? There are at least two reasons to think so.

10.1) The first problem—DISTINCT: First, consider that some patients can present with the relevant symptoms of a particular disorder, but without the underlying dysregulation or dysfunction of the associated distinct biological mechanism. More to the point, recall from §5 that evidence supporting hypotheses about dysregulation of serotonin causally determining depression has been undermined. Recall that the aim of CMAs is to stabilize diagnostic categories through showing that the biological mechanisms causally determine symptomology via COMPELLING EVIDENCE. COMPELLING EVIDENCE is met whenever multiple lines of independent research arrive at a common conclusion. Yet, the evidence that links serotonin to depression is *not* supported by multiple independent lines of research converging upon a common conclusion. Moreover, surely evidence that is "neither necessary nor sufficient" does not count as compelling. Without the evidence required to tie depression to some biological mechanism, depression seems to fall out as a valid diagnostic category.

But Tsou, and other cluster theorists who appeal to biological mechanisms, want to rule in depression, not rule it out. It could be that the simple failure of serotonin as the biomarker for depression is not convincing enough to abandon depression—we may think that there we just have not found the right one yet. Recall from §5 that if we take this position, then we play into the very problem that originally led us to the CMA: *the reification problem.* The CMA aims to resolve the reification problem but instead seems a vehicle for its reprise. If the CMA theorists wants to give a solution here, then she cannot wait until she finds the diagnostic criteria of a category. A CMA theorist like Tsou, who commits to such unique biomarkers underpinning a diagnostic category would have to give up a diagnostic category when a single biomarker failed. If a CMA theorist allows for more than a single

biomarker to be partially causally determinative of diagnostic categories, then the threat of falling into the reification problem may not loom so large.

Again, a causal theory has COMPELLING EVIDENCE when it draws its support multiple converging lines of independent research. By COMPELLING EVIDENCE, if the CMA theorist sticks with single biological mechanisms validating diagnostic categories, then they must dismiss either depression a valid diagnostic category or those patients with depression-like symptoms without serotonin dysregulation. Yet depression *does* feature as a valid diagnostic category in many CMA accounts—albeit pre-theoretically—and depressed patients *can and do* present without serotonin dysregulation. In fact, under 30% of depression patients respond to pharmacological interventions that target serotonin dysregulation and, of those, up to 60% are eventually treatment resistant.[66]

However, there is nothing obviously wrong with COMPELLING EVIDENCE as a desideratum. COMPELLING EVIDENCE merely requires that there be significant empirical support for claims about mechanisms that determine symptoms. So the problem, in this case, is that the DISTINCT biological mechanism involved with depression relies on a causal theory that does not meet COMPELLING EVIDENCE.

10.2) The second problem—DISTINCT: Second, I argue that using DISTINCT biological mechanisms to validate diagnostic categories is ill-conceived is because different disorders share similar biological pathways for their symptoms. For example, anxiety, depression, agoraphobia, and obsessive-compulsive disorder all (can) arise from dysregulation of the same neurobiological mechanism. The phenomenon of comorbidity exemplifies how such sharing occurs. The comorbidity, or cooccurrence, of one disorder with another *beyond the rate of chance* is necessary to establish a connection between

[66] Strawbridge, R., et al., (2017).

disorders.[67] Recall from §7.1 that the cooccurrence of properties in a cluster is feature of cluster kinds that appeals to such rates beyond chance. If two disorders are comorbid at rates beyond that of chance, then this feature suggests that these two disorders are actually part of one *larger* kind on a cluster account of kinds. Since CMA theorists are cluster theorists, any CMA theorist that appeals to single mechanisms causally determining diagnostic categories must concede a category collapse between two comorbid disorders. In other words, given what I consider in §9.2, since CMA theorists like Tsou commit to DISTINCT, the "sharing" of biological mechanisms cannot occur between different diagnostic categories.

There is an abundance of psychological data that supports comorbidity with greater than chance rates. Analyses of lifetime diagnoses of depression show that its comorbidity rate with anxiety is above 58%—a rate that is certainly greater than chance.[68] The CMA theorist thinks the only thing that can support these greater than chance occurrences are biological mechanisms. So, a CMA of diagnostic categories requires its proponents conclude that anxiety and depression collapse into one diagnostic category, since their differentiating symptoms have no DISTINCT biological mechanism that validates their difference.

The problem with collapsing these diagnostic categories is that many cluster theorists readily commit to these disorders counting among the valid ones. Although, a CMA theorist might simply accept that, while using the categories at the outset was a useful heuristic for analysis, it is a result of the theory that these categories must collapse. They could very well bite this bullet. Then they also accept that the result of the CMA is to reject any diagnostic category without sufficiently distinct biological mechanisms—even if it seemed otherwise at the start. Would it be an unfortunate consequence if the CMA requires that most (to perhaps all) of the diagnostic categories of the DSM

[67] Goodwin (2015), pg. 250.
[68] Ibid.

collapse into incredibly broad and generalized categories? Perhaps not. Would there be something wrong with biting that bullet? I think so.

The problem with collapsing diagnostic categories into broader ones is that doing so eliminates considerations of context and patient needs; while neurobiological mechanisms may be monolithic, patients are not. Practitioners must consider a multitude of factors outside mere biological mechanisms when diagnosing and setting treatment plans for individual patients. After all, recall from §1 that the DSM is a tool to help mental health professionals diagnose and treat patients. Moreover, as we saw in §6.2.1, the PROJECTABILITY desideratum explicitly relates to the diagnosis and treatment of patients. If we collapse diagnostic categories by shared biological mechanisms, then we undermine the purpose for which the DSM was designed by denying diagnoses to patients without neurobiological dysfunction. And denying diagnoses undermines the implementation of appropriate treatment. If one of the primary purposes for valid diagnostic categories is for diagnosis and treatment, then if the collapsed diagnostic categories prevent these aims, then they fail as valid diagnostic categories. Any account of diagnostic categories that requires a collapse, the result of which is a failure to diagnose and treat is not an account of *diagnostic* categories at all.

Another significant worry is that the requirement the biological mechanisms be distinct motivates more that the collapse of mental disorders into larger diagnostic categories. It might also motivate the collapse of mental and physical disorders into one larger diagnostic category. For instance, serotonin dysregulation is also associated with sleep dysregulation in Parkinson's patients.[69] If one shared biological mechanism suffices to motive collapsing two (or more) mental diagnostic categories into one larger diagnostic category, then one shared biological mechanism also collapses the distinction between mental and physical diagnostic categories. The CMA that we get from Tsou in §8 has no

[69] Wilson, et al. (2018).

apparent condition which prevents a "lumping"[70] of this kind. But surely *any* CMA theorist *must* acknowledge that the treatment, pharmacological or not, for physical disorders like Parkinson's is not suitable for patients with depression or anxiety, and vice versa.

One might object, at this point, that the method of argument I pursue here is flawed. There are patients with the same physical health condition, for example, who require distinct types of treatment. Distinct treatment requirements for different patients do not undermine the presence of a unifying, causal biological mechanism, the dysregulation of which presents in patients with varied symptoms.[71] I concede that there are obviously cases in which a unifying causally determinative biological mechanism presents with symptoms that differ between individual instances. So, the presence of a unified causally determinative biological mechanism is not undermined by requirements of distinct treatment methods.

But my point is that the DISTINCT desideratum requires that each disorder has one *and only one* biological mechanism that at least partially causally determines its symptomology. So, my point is that wherever there is symptom overlap *or* distinct symptoms between diagnostic categories, either of which rely on the *same* biological mechanism, the CMA *requires* a collapse between them. On top of that, when we so broaden diagnostic categories, we compound issues for establishing successful treatment strategies. Why? Well, successful treatment strategies are supposed to derive from the identification of the biological mechanism at work. I have shown how depressed patients are unlikely to respond to serotonergic pharmacological intervention. But serotonergic pharmacological intervention for Parkinson's sleep disturbance has been demonstrably effective.[72] The creation of diagnostic categories on the basis of a unique biological mechanism is supposed to allow for predictive

[70] This terminology is due, in part, to Craver (2009), where he discusses the motivation for isolating one mechanism over another.

[71] Many thanks to Paul Humphreys for raising this point.

[72] Yasue, I., et al. (2016).

inductive inference about the members of the category (PROJECTABILITY). If we collapse Parkinson's patients with depressed or anxious ones into a larger diagnostic category of which they are all members, then it is difficult to make successful inductive predictive inferences about how these members will respond to serotonergic pharmacological intervention. Since one of the DSM's primary aims is internal validity *for the explicit purpose of aiding diagnosis and treatment,* lumping together mental and physical diagnostic categories only confounds this already elusive goal.

Since the CMA is supposed to supplement the DSM to avoid the reification problem, thereby securing external validity for the DSM, it cannot resolve the problem at the cost of introducing another (§3,5). But because the CMA, as it stands, uses the DISTINCT desideratum, it does introduce further problems for the DSM. So, the CMA currently cannot serve as a supplement to the DSM to resolve the reification problem.

But what if the CMA theorist dropped DISTINCT? The CMA theorist would then be giving an account of diagnostic categories that are valid whenever they meet STABILITY, PROJECTABILITY, and COMPELLING EVIDENCE. Could such a CMA supplement the DSM and resolve the reification problem?

11] The problem with collapsing natural kinds into biological kinds

I previously suggested that these desiderata are not obviously corrosive to an account of diagnostic categories. So, I will assume that these desiderata are acceptable when characterized without reference to specific kinds of mechanisms. Instead, my remaining objection to the CMA of diagnostic categories on offer is that it relies upon an assumption that natural kinds are simply biological kinds.

It is hard to see how such a strong association between natural and biological does not, when applied to mental health categories result in the unfortunate collapse of mental health with physical health, as I examined in §9.2. If, for the sake of argument, depression simply is (at least partially)

causally determined by serotonin dysregulation, then in what sense is depression a mental health diagnostic category rather than a physical one? CMA theorists do not seem to have any recourse to answer this question. CMA theorists cannot appeal to the remaining symptoms of a mental health category that are not underwritten by biological mechanisms because they cannot be adequately stable. In other words, these symptoms cannot be "adequately stable" because they are not even partially caused by dysregulation in biological mechanisms—at least not in a way that meets COMPELLING EVIDENCE. But what makes CMA theorists think that a causal theoretical account of biological mechanisms meet STABILITY and PROJECTABILITY in a way that social mechanisms cannot?

There is overwhelming evidence (COMPELLING EVIDENCE) that social mechanisms can and do enable successful predictive inductive inference.[73] So, social mechanisms meet the minimum necessary conditions of STABILITY and PROJECTABILITY. Moreover, and perhaps more devastating to the default precedence granted to causal biological mechanisms, there is COMPELLING EVIDENCE that social mechanisms can and do causally underwrite changes in biological mechanisms. For instance, women with PTSD from childhood sexual abuse demonstrate distinctive changes in the functioning of conditioned fear responses to the extent that their brains literally have less blood flow in their prefrontal cortex when presented with imagery that triggers memories of the traumatic event.[74] Child abuse is certainly underwritten by social mechanisms, not biological ones. This means that the operational symptoms of PTSD in these women are caused, at least in part, by the dysregulation of biological mechanisms *that are themselves caused by identifiable social mechanisms*. And there is COMPELLING EVIDENCE that affirms this causal relationship between the two types of mechanism. So, placing STABILITY and PROJECTABILITY on biological mechanisms alone does not occlude the causal

[73] De Bellis and Pan (2003). I do not have the space here to elaborate, but there are many examples which demonstrate this general claim. Famous among them is the Milgram experiment and the Stanford experiment, though these are not without their significant faults in experimental design.

[74] De Bellis and Pan (2003), see particularly pg. 111.

contribution of social mechanisms, because there are cases in which the relationship between biological mechanisms and social mechanisms is causally determinative.

The dismissal of social kinds, and their mechanisms, might be due to their default treatment as interactive kinds. Interactive kinds, as we saw in §6.2.1, are those that do not allow for STABILITY and PROJECTABILITY and so are regarded as artificial because they are not externally valid. But since social mechanisms can causally determine symptoms with the degree of regularity required for predictive inductive inference, we cannot blanket identify social kinds with interactive kinds. Further, given that there is COMPELLING EVIDENCE that, like in the PTSD case, social mechanisms causally determine biological mechanisms, which in turn causally determine diagnostic symptomology, what is the kind-status of a diagnostic category so determined? I cannot imagine the CMA theorist has an easy answer. A CMA that does not rely solely on biological mechanisms to validate diagnostic categories could provide an answer by denying the conflation of artificial kinds with social kinds. A CMA of diagnostic categories that does rely on biological mechanisms alone would have to make an extra-theoretical appeal or an exception. Either choice seems unavoidably *ad hoc*.

12] Does a CMA of diagnostic categories solve the reification problem? No.

So, what if we give a CMA that does not rely solely on biological mechanisms to underwrite diagnostic categories? Can such a CMA resolve the reification problem for the DSM? All that we need now is to show that this new CMA has a principled method of ruling in valid diagnostic categories and ruling out others.

However, a CMA of this form seems either unprincipled or inchoate. This CMA would amount to nothing more than the general concern that empirical findings need support a range of causal hypotheses whenever we evaluate diagnostic categories for external validity. And it seems we already demonstrate this concern in both the DSM and the RDoC. Yet this is so general an account

that it is widely uncontroversial. Even if I attempted to adequately specify a CMA that uses both social and biological mechanisms, the failures of biological mechanisms for truly causally determining symptomology of diagnostic categories would likely be mirrored in social mechanisms as well. So, such a CMA cannot offer a solution to the reification problem because it would not give a principled method for ruling in valid categories and ruling out others.

Recall from §5 that the RDoC fails to resolve the reification problem and instead merely pushes it back. I suggest that appeals to any mechanism, social or biological, for validating diagnostic categories also merely pushes back the problem. How? Such appeals allow the same room for the assumption that at least one symptom of a diagnostic category must be at least partially causally determined by either biological or social mechanisms. It is worthwhile to note that the CMA's failure in providing a principled method for ruling in valid diagnostic categories and ruling out others is because the CMA must posit these theoretical connections between some mechanism and a symptom in the DSM. As we saw with the RDoC, it is this conceptual space that invites the reification problem for diagnostic categories. So, now I conclude that even the most inclusively formulated CMA of diagnostic categories cannot resolve the reification problem either.

13] Constitutive Position: Conclusions and Considerations

So, we remain at the constitutive position, without a solution to the reification problem. I have shown that all of the attempts to either avoid or resolve the reification problem either fail to do so or avoid the problem at the expense of established internal validity. At the very least, we know that the reification problem is a problem for any account of mental health diagnostic categories. But that the problem is ubiquitous among accounts does not reassure those worried about establishing external validity for diagnostic categories. Must we be satisfied with externally valid until proven otherwise?

Assuming this as a default position is obviously problematic. There is the unfortunate reality that some of diagnostic categories in previous editions of the DSM were hopelessly misguided and still others that were outright offensive. There is also the problem that assuming this default position, taking the constitutive interpretation, could result in individuals being disordered in virtue of simply meeting the minimum criteria despite not actually being so. But the constitutive position originates in the assumption that mental illnesses are *just as real* as physical ones; even if such an assumption is the root of the reification problem, it seems at least a well-meaning one. Moreover, my arguments in §11 further support this parity of status without collapsing mental and physical illness by showing that mental diagnostic categories cannot be validated by physical mechanisms alone. I conclude that, even without a resolution to the reification problem, the constitutive position remains the best we can do with what we have. It preserves internal validity and supports the reality of mental illness in a way that does not require reduction to physical illness.

II.

It is easy to say that mental disorders are like physical disorders, but it is not obviously accurate. This poses a problem when much of our arguments surrounding mental health rely on this type of assumption—either explicitly or implicitly. For example, much of the debate for parity between mental and physical health trades on an assumed similarity between mental and physical illnesses. Some of the arguments that point to robust interaction between mental and physical illness assume that these interactions require likeness between the two. Many arguments about the nature of mental disorders, what mental disorders *are*, similarly rely on this assumption. And all of these arguments persist despite independent arguments against direct similarities between mental disorders and physical disorders.

The most (in)famous of these independent arguments against direct comparisons between mental and physical disorders is given by Thomas Szaz.[75] Since Szaz's view is paradigmatically anti-psychiatry, it is not unreasonable for his concerns to have been set aside when we debate the above arguments. Today, the default view of mental disorders is that they are undeniably real[76] and so there *is* such a thing as mental illness, directly contra Szaz. So now the various debates proceed from the default view that there is such a thing as mental illness together with the assumption that mental and physical illness are alike. Since these debates proceed from this assumption, however, it is still worth critical attention. Thus, my project takes a sympathetic stance on the received view that mental disorders are real, but critically examines the way in which that assumption typically plays out in arguments about the nature of mental disorder.

---

[75] Szaz (1960), (2009).

[76] This is not to say that *every* disorder in our diagnostic manuals are genuine disorders. Instead, I simply defend the view that there are such things as mental disorders, even if we have not pinpointed all and only mental disorders in our manuals.

My argument starts from a simple idea, then. It seems like most accounts about the nature of mental disorder require that, if they are real, then they are real in the same way that physical disorders are. But these views require an additional implicit commitment to the assumption that mental disorders are somehow like physical disorders. Now, if mental disorders are real *and they are like physical disorders*, then they are real in the same way that physical disorders are. Now the question is, in what sense are physical disorders real, on the typical view? We tend to think that physical disorders are real because they are, or correspond appropriately to, natural kinds. Now our version of the basis for arguments about the nature of mental disorder states that if mental disorders are real and they are like physical disorders, then they are natural kinds. So, if mental disorders are natural kinds, then they must be captured by either of our two most successful accounts of natural kinds.

I argue that mental disorders cannot be captured by either of these two most successful accounts of natural kinds. I dismiss the first nearly outright (§2), since it is overwhelming uncontroversial that it is not commensurate with mental disorders as we currently understand them. Then I consider the second option (§3) and explain that if mental disorders can be captured by this second account they cannot violate its minimum conditions (§4). Then I show that mental disorders do violate these minimum requirements (§5-7) even under the most idealized conditions (§8-9). So I conclude (§10) that mental disorders are not *natural kinds*.

Before we begin, I want to reiterate, explicitly, that this conclusion does not commit me to the view that mental disorders are not natural, or that they are not real. I am committed to mental disorders being real and being natural. So this argument is one that proceeds from the assumption that, since mental disorders are very much real, they are phenomena *worthy of* and *important candidates for* critical examination. If anything, my conclusion should only bring to the forefront that mental disorders are real, and whatever they are, *really* conceptually messy.

1] Natural kinds accounts

Consider the following collection: my laptop charger, a marker, the number 5, and Dolly Parton. This is obviously not a natural grouping. There is nothing that unites this group aside from it being just a random collection I've thrown together off the top of my head. It is arbitrary. We can compare this non-natural grouping to another type of non-natural grouping: young adult fiction. This group is not arbitrary because it reflects human interests, which means that there are some features that the things in that group share. But it is obviously not a natural grouping. We can contrast these non-natural groupings with natural ones.

The view that things that exist in the world—trees, animals, grasses, chemicals, rocks—are grouped together naturally in genuine divisions in the world is called *naturalism*. Taking naturalism one step further, *natural kind realism* commits to the view that those naturally occurring groups are made up of genuinely distinct entities that exist in the world: natural kinds. When we opt to take the realist position about natural kinds, we are making a stronger claim about the ontological status of the grouped entities. There are two views which make up the bulk of the positive accounts of natural kind realism.[77] These views simply differ in what they think makes a natural kind a natural kind. In what follows, I will examine how each account fares when extended to mental disorders. Next, I will explain the first (§2), its two most persuasive objections (§2.1), and dismiss it (§2.2). Then I examine the second (§3).

2] Essential kinds

What makes gold the thing that it is? Take the smallest unit of gold we can, the gold atom. One way that we can answer that question is to say that it is essential to being a gold atom that it have 79 protons

[77] Technically there are three. But we are not entertaining Dupre's (1993) promiscuous realism: the view that every reasonable grouping is a natural kind when indexed appropriately to our interests, and so no kind should be privileged over others.

in its nucleus. If an atom had more or less than 79 atoms in its nucleus, then it can no longer be a gold atom. Now apply this kind of reasoning to things outside of chemical elements. We might claim that what makes a species a species is that it is reproductively isolated breeding population; to be a new species is essentially to be a reproductively isolated breeding population.[78] When we make claims like these, we are giving an account of these kinds as *essential kinds*. But it is when we move away from the neat cases presented by chemical elements into other kinds in the natural sciences, an essential kinds view begins to fail. For example, if we can find just one example where two similar groups of animals are not reproductively isolated but are in fact different species, then we must revise our essentialist account of what makes a species a species. If sympatric speciation is possible,[79] for example, then what makes a species a species cannot be reproductive isolation. Alternatively, we might have a species that cannot viably reproduce despite being a reproductively isolated breeding population. Or we might come across a species that is reproductively isolated but not limited to a single natural breeding population.[80] So we will need to find something else that is essential to being a species. Such revisions don't necessarily require we abandon essentialism about natural kinds, but it leaves us with two consequences that give us good reasons to think we should.

2.1) Problems with essential kinds: When faced with a counterexample that shows a species is not essentially a reproductively isolated breeding population, we might say we were just wrong about what is essential to being a species. But admitting we were mistaken about what was essential to species is not admitting that there is nothing essential to species. So we just have to find a new feature that is essential to species and, if we are wrong about that one, revise our account again. But this move

[78] Mayr (1969), Matthen (2009).

[79] This is somewhat scientifically contentious and has been for a while. Sympatric speciation is when a breeding population in the same reproductive geographic region spontaneously acquires some members with new traits, and those members begin to selectively reproduce with only those who also have those traits. The reproductive isolation is self-selected by trait matching in a population that would otherwise allow free breeding between members.

[80] The liger is a fertile hybrid that can mate with other ligers, or lions, or tigers (Colston-Nepali & Leigh, 2019).

assumes that *there must be* something essential to being a species—we simply must pinpoint what that feature is. This assumption is question-begging; it simply assumes that there is an essential feature that establishes species as kinds.[81] This question-begging quality of essential kinds accounts is the first and potentially worst problem for essentialism. If we simply can revise our account each time a suggested essential feature of species fails, at what point must we give up on species as the type of thing that is a natural kind at all? On this understanding, it seems we assume the essential feature is out there because we already take species to be the type of thing we consider a natural kind. So we would not reasonably give up on species as a natural kind. Instead we might just conclude that whatever is essential to its kindhood is not something to which we have epistemic access. But now we lack good reasons to call species natural kinds besides our own intuitions that they are, and we have no principled explanation for calling some seemingly natural grouping species a natural kind. Since kinds are supposed to explain *why* they are kinds,[82] this epistemic inaccessibility would make our assumption that species are natural kinds dubious.

But let's assume that species are neat and tidy things we all can agree are natural kinds anyway. If that is the case, then any single counterexample for a suggested essential feature will not do away with our assumption that species are natural kinds. Instead, these counterexamples will force us back to the drawing board, where we will see that an essentialist account of natural kinds is unable to account for the flexibility in feature expression that we see in species. Some species might be reproductively isolated breeding populations, and others might not. But both are still species, and we want to insist that species are still natural kinds. So we are forced to recognize that essentialism about species is either going to lead us to the explanatory dead-end of epistemic inaccessibility or to conclude

[81] See Mellor (1977), Dupre (1981, 1993). Salmon (1982) introduces the question-begging concern against Putnam and Kripke's essentialist accounts, but these are distinct from the default view we are taking. Putnam's semantic externalism and Kripke's essentialist account amount only to claims of necessary features for membership, not necessary and sufficient features for membership.

[82] Hawley and Bird (2011).

that essentialism cannot adequately capture natural kinds' in-group variation. This is the second problem for essentialist accounts of natural kinds—they are too restrictive to capture the types of things we think are reasonably natural kind candidates. So the combination of the explanatory insufficiency and this inflexibility is strong motivation to abandon essentialism about natural kinds.

2.2) Failure of mental disorders as essential kinds: So far we have seen some philosophical reasons why we want to abandon essentialist accounts of natural kinds generally. But the purpose of this project is to discuss whether mental disorders can be natural kinds. So we should also ask: why can't mental disorders be essential natural kinds? The most obvious answer requires only a cursory look at mental disorders. There is more than one way to be diagnosed with Attention Deficit and Hyperactivity Disorder (ADHD), for example. Let's look at all the disorder criteria for hyperactive-impulsivity type ADHD given in the Diagnostic and Statistical Manual-V-Text Revision (DSM-V-TR) and see if we can find an essential feature.

> Attention Deficit and Hyperactivity Disorder:
> A. A persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, as characterized by (1) and/or (2):
> …
>
> > 2. Hyperactivity and impulsivity: Six (or more) of the following symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level and that negatively impacts directly on social and academic/occupational activities: Note: The symptoms are not solely a manifestation of oppositional behavior, defiance, hostility, or a failure to understand tasks or instructions. For older adolescents and adults (age 17 and older), at least five symptoms are required.
> >
> > > a. Often fidgets with or taps hands or feet or squirms in seat.
> > > b. Often leaves seat in situations when remaining seated is expected (e.g., leaves his or her place in the classroom, in the office or other workplace, or in other situations that require remaining in place).
> > > c. Often runs about or climbs in situations where it is inappropriate. (Note: In adolescents or adults, may be limited to feeling restless.)
> > > d. Often unable to play or engage in leisure activities quietly.
> > > e. Is often "on the go," acting as if "driven by a motor" (e.g., is unable to be or uncomfortable being still for extended time, as in restaurants, meetings; may be experienced by others as being restless or difficult to keep up with).
> > > f. Often talks excessively.
> > > g. Often blurts out an answer before a question has been completed (e.g., completes people's sentences; cannot wait for turn in conversation).
> > > h. Often has difficulty waiting his or her turn (e.g., while waiting in line).

> i. Often interrupts or intrudes on others (e.g., butts into conversations, games, or activities; may start using other people's things without asking or receiving permission; for adolescents and adults, may intrude into or take over what others are doing).
>
> B. Several inattentive or hyperactive-impulsive symptoms were present prior to age 12 years.
>
> C. Several inattentive or hyperactive-impulsive symptoms are present in two or more settings (e.g., at home, school, or work; with friends or relatives; in other activities).
>
> D. There is clear evidence that the symptoms interfere with, or reduce the quality of, social, academic, or occupational functioning.
>
> E. The symptoms do not occur exclusively during the course of schizophrenia or another psychotic disorder and are not better explained by another mental disorder (e.g., mood disorder, anxiety disorder, dissociative disorder, personality disorder, substance intoxication or withdrawal).[83]

We can see that there is no single feature of ADHD that is like that of a gold atom having 79 protons in the nucleus; no individual symptom of ADHD is essential for having ADHD. For any mental disorder to be a natural kind on the essentialist account it must have some feature that is both necessary and sufficient for being that disorder. A feature that is both necessary and sufficient for being a member of a kind, like having 79 protons in the nucleus is necessary and sufficient for being a gold atom, is an essential feature. And most—if not all—mental disorders do not have any features that are both necessary and sufficient for being that mental disorder. I will not take up the project of proving this claim here. It is sufficient to state that the way that mental disorders are currently defined is explicitly prohibitive of an essentialist account of disorders. I will talk more about the way that mental disorders are defined in §4. But, for now, we can see how the argument against mental disorders as essentialist kinds would proceed *mutatis mutandis* from our ADHD example case to any other mental disorder.

3] Cluster kinds

Now let's consider a specific species: domesticated cats (*Felis catus*). What are the features of domesticated cats? We can give a truncated version of a more exhaustive list of cat features: cats are mammals, have fur, have whiskers, have four legs, can jump to heights up to six times their length

[83] APA (2017).

from a standing position, almost always land on their feet, can fit through any opening their head can fit through, are predators, are obligate carnivores, etc. Now consider a sphinx cat—these are domesticated cats that don't have fur but are still cats. Sphinx cats are still four-legged mammals with whiskers, can jump to heights up to six times their length from standing position, etc., and so are still cats. My fat domestic shorthair tabby has fur, is a predator (in a loose, lazy bug-hunting sense), has all the other listed features of cat, but could not actually fit his bulk through any space he could fit his head into. He is still a cat! And so is any cat that incidentally does not have four legs. And so on. We fill out the list of features that cats have so that we can reasonably allow for the flexibility of a kinds' in-group variation that was missing from essentialist accounts. So we see that the inability of a particular cat to meet a feature or some features from the 'list' does not disqualify it from membership in the kind. A cat simply must exhibit a sufficient number of those features to be a member of the kind and, like we saw with the furless sphinx cat or my fat cat, these features need not be identical to other members in the kind.

Call the features we listed of cats above properties, and we have something close to the typical account of natural kinds as clusters of properties. When we give accounts of natural kinds in this way, they are called cluster kinds. Like we saw in our case of domestic cats, cluster kinds do not require that every member of the kind have all the listed properties or have identical properties to every other member of the kind. In fact, members of the kind need only have a sufficient number of the properties in the cluster to be considered a member, and do not need to share any properties in common with another member. This means that no property is individually necessary or alone sufficient for membership in a kind. This is the assumption under the standard view to allow for the in-group variation that the essentialist account failed to accommodate. Additionally, the requirement that there be no individual necessary nor individually sufficient properties for membership implies that there are no individually necessary but alone insufficient properties for membership. And so on. Without these

conditions on the properties in the cluster, the second requirement for the cluster of properties is that *something* makes it the case that we can reasonably predict the properties a member of that kind might exhibit.[84] How?

We allow for such reasonable inductions through the presence of some mechanism or function that ensures the mutual promotion of the expression of other properties in the cluster kind list. So it is the mechanism that does the explanatory work for the cluster kind by making it the case that we can reasonably predict the properties a member might exhibit. This is another central feature required in any account of kinds: that the kind support predictive inferences about its members.[85] On the original account of cluster kinds given by Boyd (1999), these mechanisms were causal mechanisms that supported stable equilibrium, or *homeostatic causal mechanisms*. For our case of cats then, Boyd would point to the homeostatic causal mechanisms that result from breeding, which self-regulate the cluster of properties relevant to domesticated cats. It is these homeostatic causal mechanisms that ensure cats are likely to have whiskers, fur, be capable of jumping heights six times their length from standing, etc. Subsequent iterations of cluster kinds accounts maintain that the relevant causal mechanisms are not necessarily homeostatic.[86] However, nearly all retain the use of these determinative mechanisms, albeit unpacked differently, to explain the clustering of the properties in a kind.[87]

No matter how these mechanisms are unpacked, they allow for us to make reasonable predictions about the properties a cat might express because these mechanisms make the cat property cluster *stable*. The idea, according to the cluster theorist, is that the mechanisms that promote the expression of the properties in the cluster are such that they promote only properties that support the

[84] I am interested in the explanatory feature of a cluster kind that allows for predictive inferences, here, as partly a metaphysical fact about kinds, rather than any other epistemic goal for projectability of cluster kinds.

[85] Quine (1969), Mill (1884).

[86] Craver (2009) for example, argues that homeostasis is insufficient for delineating distinctions in the causal functions that determine unique property clusters; see especially his argument about lumping and splitting.

[87] For some examples of this in action see: Craver (2007), Wilson (1999), and Slater (2015).

presence of others. Whenever a mechanism gives rise to a property that is detrimental to the cluster, the expression of this property is suppressed or eliminated in members for the sake of the cluster's stability. As a result, we can close our eyes in a room full of domestic cats of all breeds with all kinds of incidental variations, spin around, point to any cat, and reasonably predict "that individual is an obligate carnivore." And we can do this because "being an obligate carnivore" is the result of a mechanism that promotes only properties that support the overall stability of the cluster.

Another feature of these causal mechanisms is that they are supposed to weed out other individuals from being members of the cat kind, even when that individual might have some similar properties to the cat. Consider the "bear cat" or binturong. Binturongs are four-legged mammals that have whiskers and fur. So the Binturong has properties that overlap with the domestic cat. But is the bear cat a cat? No. It has other features that the domestic cat does not; it has the property of naturally smelling like popcorn, has a prehensile tail, is an opportunistic feeder.[88] The binturong also lacks properties of domestic cats; it cannot jump to heights up to six times its length from standing and it is not an obligate carnivore. While it shares some properties with the domestic cat, the binturong is not a cat because it has a different cluster of properties. And, according to the cluster theory of kinds, it is the mechanism that enables the expression of these clusters of properties and differentiates between the two animals. So, according to the cluster theory of kinds, there are mechanisms that produce a cluster of properties relevant to the domestic cat and a cluster of properties that is relevant to the binturong. And both the cat and binturong are distinct species because the production of these clusters by these mechanisms is stable. So the causal mechanisms that promote the expression of properties in the cluster also establish the boundaries of the cluster kind.

[88] Davenport (2016).

3.1) Cluster kind clarification: Before we move on, we need to clarify a distinction that occurs in the literature surrounding cluster kinds. We might have in the world a group of observable properties going around together. What makes the difference between some properties simply going around together and those properties going around together naturally? In other words, what tells us why one group of properties is a natural kind and another is not? The hallmark response is that in the case of properties that just happen to go around together, like the properties involved in young adult fiction, these properties are not carving nature at its joints. Instead, these groups of observable properties might be hanging around together by convention, for example. Or we might think of these kinds of clusters as *useful* and so are real in this sense, like cuts of meat or America. There is no *natural* part of a cow called a filet but when I go to the butcher and request one they know what I am asking for, and which part of a cow to cut into to get it.[89] It is obvious that cuts of meat, America, and young adult fiction are not kinds like gold. Yet cuts of meat, America, and young adult fiction are kinds—these groupings are not arbitrary and they reflect divisions in the world that are relevant to human interests. So there is something that is doing the work of holding these clusters together. The difference between these and natural cluster kinds, roughly, is that the natural cluster kinds have mechanisms that participate in the determinative structure of the world.[90]

4] Mental Disorders as cluster kinds

Now let's return to our main project. What about mental disorders, are they natural cluster kinds? Before we begin, let's start with a few stipulations. For now, let's consider "meeting the requirements for diagnosis" to be equivalent to "exhibiting the properties required for membership." This is not uncontroversial, since (professional) diagnosis always requires an external individual's judgment of another person's membership. For example, if I was to seek a formal diagnosis for anxiety, I would

[89] Haslanger & Saul (2006), Dupre (1996), to name a few.
[90] Wilson (1999), (2005).

need an appropriately credentialed mental health professional to decide if the kinds of properties I express are those associated with anxiety. These decisions are often impacted by various external factors, such as time constraints on clinical examinations,[91] gendered and racial diagnostic biases,[92] and the limitations of self-report measures for evaluating the presence of symptoms.[93] So the adjudication aspect of diagnosis leaves obvious space for diagnosis and membership to come apart. And this space between diagnosis and membership persists even when we, as I will do below, leave out those diagnostic criteria that are arguably aimed *only* at aiding clinical decisions about diagnosis. But we will proceed as if it does not.[94]

Under the assumption that meeting requirements for diagnosis is equivalent to exhibiting the properties required for membership, are mental disorders cluster kinds? The existing literature suggests that, if mental disorders are natural kinds, then they are cluster kinds.[95] Why? Let's look at the popular example of major depressive disorder for our answer. Major depressive disorder (MDD) is a specific disorder, located in the DSM under the larger umbrella category of depressive disorders. The diagnostic criteria for MDD are

> Major Depressive Disorder:
> A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.
> *Note: Do not include symptoms that are clearly attributable to another medical condition.*
> (1) Depressed most of the day, nearly every day as indicated by subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful)
> (2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by subjective account or observation)
> (3) Significant weight loss when not dieting or weight gain (e.g., change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day
> (4) Insomnia or hypersomnia nearly every day

[91] For one example, see Tsiga *et al* (2013).

[92] Just a few among many examples are Garb (2021), Bjorklund (2006), and Ussher (2013).

[93] See, for example, Nevin (2009).

[94] I have argued elsewhere (Oakley, Chapter 1) that, as a matter of how we in fact use diagnostic categories, we must opt to view at least some of their diagnostic criteria as literally constitutive of the disorder. This contravenes how psychiatric practice thinks of these diagnostic categories. The idea is that despite their own views, the actual application conflicts with their reflective theory about the diagnostic manuals.

[95] See, for example, Tsou (2015, 2016), Tabb (2015, 2017), and Tekin (2016).

(5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)
(6) Fatigue or loss of energy nearly every day
(7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
(8) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
(9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The episode is not attributable to the physiological effects of a substance or to another medical condition.

Note: The above criteria represent a major depressive episode.

D. The occurrence of the major depressive episode is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.
E. There has never been a manic episode or a hypomanic episode.

Note: This exclusion does not apply if all of the manic-like or hypomanic-like episodes are substance induced or are attributable to the physiological effects of another medical condition.[96]

We can see from these diagnostic criteria for MDD that there is more than one way to be diagnosed with MDD. This method of defining the diagnostic criteria is what mental health professionals call *polythetic*. Polythetic is a term applied specifically to clusters within groups—this makes sense when you take into account that "depressive disorders" is a broad umbrella group of disorders of which MDD is a member. That MDD or any other mental disorder is polythetic is meant to indicate that members in a cluster share some, but not all, properties in common.[97] It is easy to see how we took the step from polythetic diagnostic categories to diagnostic categories as cluster kinds then. Reasonably, both have the requirement that individual members exhibit a sufficient number of the properties in the cluster, but they need not be the same properties or all the properties.

But this is where the similarity between polythetic clusters and cluster kinds seems to end. Why? Cluster kinds have further requirements that polythetic groupings do not; that there are no

[96] APA (2017).

[97] Mental health professionals further use polythetic to also indicate that the disorder criteria include symptoms that are presented *disjunctively*—such that individual symptoms can include "either/or" formats. This simply reinforces the typical use for the term, which implies that members will share some but not all properties in common.

individually necessary but alone insufficient property in the cluster, and that there is a mechanism that stabilizes the cluster by promoting the mutual expression of properties in the cluster. So if we move forward with mental disorders as cluster kinds, as it seems the received view does, we have to account for the similarities between disorders being polythetic and being clusters *and* for these additional requirements. What does that mean for our example of MDD?

4.1) Properties in symptoms: To start, assuming that MDD is a cluster kind means that all the individual symptoms in the diagnostic criteria are properties. And the same would also apply for any mental disorder. The question of whether all the symptoms given above are the type of thing we would reasonably consider properties is a concern I will not consider at length here. I can simply say that if you think that disorder symptoms are not the type of thing we would reasonably consider properties, then your commitment to such a view would not necessarily undermine my ultimate conclusion in this project. So for now, let's just look at the symptoms we're given and how we might unpack them into properties.

For the case of a symptom, let's say MDD (8) "diminished ability to think or concentrate, or indecisiveness," it is evident that the symptom represents more than a single property. But the question of how we unpack the properties that the symptom expresses is open to interpretation. For our purposes here, it might be easier to unpack properties that we list as phenomena or behaviors the individual has or does, instead of what they do not have or do. So let's look at a single symptom from another disorder. Consider this symptom from ADHD: "Often fidgets with or taps hands or feet or squirms in seat."[98] On the most expansive interpretation of this sentence, there are five properties it expresses: fidgeting with hands, fidgeting with feet, tapping with hands, tapping with feet, and squirming in seat. On a more restricted interpretation of the symptom sentence, there are three

[98] APA (2017).

properties: fidgeting, tapping, and squirming. The expression of multiple properties in a single symptom listed in the diagnostic manual need not bear on our assertion that, if the disorders are cluster kinds, then the symptoms must be properties. That the symptoms disjoin the expressed properties does not entail that the symptom no longer expresses a property—it is just a convenient mode of communicating a number of properties, whatever that number ultimately is. And as we saw with MDD symptom (8), the expression of these properties framed as absences of phenomena or behaviors further complicates unpacking the properties involved. Still, such complications do not necessarily entail that the given symptom no longer expresses at least one property.

At this point, I want to note that I am not taking a particular stance on what counts as a property. The success of my argument here should not rely on such a commitment. Instead, I simply want to demonstrate that, under the received view of mental disorders as cluster kinds, the symptoms in the disorder must be understood in terms of properties.

4.2) Excluding adjudication conditions: That being said, I do want to make one assumption moving forward in our analysis of mental disorders. This assumption is about what I will call *adjudication conditions*. Adjudication conditions are those symptoms listed in a diagnostic category that I suggest serve only as "diagnostic guides." What do I mean by this?

Let's start by looking at the following symptoms from MDD, which represent what I think are paradigmatic examples of adjudication conditions:

> C. The episode is not attributable to the physiological effects of a substance or to another medical condition.
>
> *Note: The above criteria represent a major depressive episode.*
>
> D. The occurrence of the major depressive episode is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.
>
> E. There has never been a manic episode or a hypomanic episode.

> *Note: This exclusion does not apply if all of the manic-like or hypomanic-like episodes are substance induced or are attributable to the physiological effects of another medical condition.*[99]

(C), (D), and (E) all rule out some other phenomena as better explaining the presence of symptoms in an individual that would otherwise qualify them for diagnosis with MDD. It is a feature of the diagnostic manual that allows clinicians to discriminate between cases of disorder and, for example, cases where someone might simply need to revisit their medication plan. These conditions help clinicians diagnose by disqualifying patients that otherwise meet criteria for diagnosis under varied conditions. So these criteria help to adjudicate diagnosis, hence *adjudication conditions.*

However, it is my suggestion that this feature for the diagnostic manual is a hindrance for our analysis of mental disorders as natural kinds. Therefore, I will assume throughout that these adjudication conditions are not among the properties of the mental disorder. They were not meant to be properties of an individual for membership, but simply guardrails for diagnosticians. And so I will drop the adjudication conditions from here on out. This might not be an uncontroversial assumption.[100] But I will proceed as if it is. On this understanding, we can move forward and examine if an account of mental disorders as cluster kinds meets the further requirements for such an account.

5] Requirement for no individually necessary conditions.

Recall the first additional criteria for cluster kinds accounts: that there be no individually necessary but alone insufficient property required for membership in the kind. Now let's revisit the opening criteria for MDD:

---

[99] APA (2017).

[100] One might want to extend this analysis of the adjudication conditions to other properties expressed by the symptoms. Since diagnostic manuals are, by definition and purpose, diagnostic guides, dropping only the adjudication conditions might mean dropping any condition. But I would simply suggest that this is i) too quick an extension made by trading on a semantic quibble, and ii) prohibited explicitly by arguments I have made previously that require our diagnostic manuals to be interpreted as literal and constitutive (given the purpose driven aim of external validity in our diagnostic manuals). A better argument would pick up on the thread of adjudication conditions being seemingly arbitrary. But it is my view that there is at least significant prima facie reason to take the conditions I suggest as those that are best candidates for adjudication conditions. And so I take it as an assumption that these are the adjudication conditions and they can be, as a result, reasonably left out of analysis of a particular disorder and its properties.

> Major Depressive Disorder:
> A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.
> *Note: Do not include symptoms that are clearly attributable to another medical condition.*
> (1) Depressed most of the day, nearly every day as indicated by subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful)
> (2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by subjective account or observation)[101]

Immediately, we see an explicit requirement of an individually necessary but alone insufficient condition for MDD diagnosis—either depressed mood (1) or anhedonia (2) *must* be among the at least five symptoms an individual needs to express for diagnosis. So we have an immediate and explicit counterexample to the requirements for cluster kinds in MDD. You can have all the other symptoms described in the category, but if you do not have either anhedonia or depressed mood among them, you are not depressed. (1) or (2) is necessary for diagnosis with MDD. But if you have (1) or (2) alone, and no other symptoms you do not meet the requirements for diagnosis with MDD. So (1) or (2) are both alone insufficient for membership.

But not all disorders have this type of individually necessary but alone insufficient symptom. That is to say: it may reasonably be the case that MDD is the only disorder that has a requirement of this kind. Do we simply dismiss whatever disorders have this kind of requirement and save the rest? It seems a hard bullet to bite that we dismiss MDD alone on these grounds, since MDD is widely held as the most paradigmatically real among all the mental disorders. How? MDD is the default example in almost all philosophical discussions about the nature of mental disorder and is used as *the case* for any account to capture. But we could arguably bite that bullet, even if bitterly. Does that mean the rest of mental disorders are saved?

[101] APA (2017).

If this were the only case of individually necessary but alone insufficient symptoms for disorders and we were committed to maintaining mental disorders as cluster kinds, then yes. However, there are further criteria that violate the prohibition against individually necessary but alone insufficient properties for membership. Consider criterion (A) of MDD again:

> Major Depressive Disorder:
> A. Five (or more) of the following symptoms *have been present during the same 2-week period and represent a change from previous functioning*…[my emphasis][102]

Embedded into criterion (A) for MDD is this additional requirement for diagnosis: duration of symptoms. We can call this *the duration criterion*. In MDD this requirement is for (at least) a two-week period. In other disorders, the duration requirement differs. On those disorders, the amount of time required can be shorter or longer than the two weeks required by MDD. But all have an amount of time that the symptoms must be present as an additional, individually necessary but alone insufficient property for diagnosis. So have we found our violation in the duration criterion? Maybe not. Why?

We might reasonably dismiss this duration criterion on the basis that is purely a diagnostic guide. Recall from earlier that I will not use those criteria in a diagnostic category that I consider *adjudication conditions*. I understand adjudication conditions as those criteria in diagnostic category that serve only to help a professional *decide* about a diagnosis. It is reasonable to think that the duration criterion is one of these criteria that help a professional decide about a diagnosis. Surely if you are depressed, you meet the criteria for diagnosis if you have five of the symptoms, one of which is either anhedonia or depressed mood, and it's only been one week. If you are a member of a kind, you are surely a member of that kind the moment you meet the criteria, and not after a (seemingly and arguably arbitrary) probationary waiting period. The duration criterion is to help professionals differentiate between periods of profound sadness and depression, etc. So it is only there to help professionals decide about diagnosis. Therefore, if we were okay with dropping adjudication conditions above, and

[102] APA (2017).

we were, then we have to be okay with dropping an adjudication condition here. So the duration criterion does not violate the prohibition against any individually necessary but alone insufficient properties for membership, because it is not among the properties required for diagnosis.

But we are not out of the woods on these violations yet. In fact, MDD also shows another type of individually necessary but alone insufficient symptom:

> Major Depressive Disorder:
> ….
> B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

We might call this *the distress criterion*. The distress criterion is present in most, if not all, of the disorders—albeit with some changes in the wording for each disorder.[103] It is an individually necessary but alone insufficient criterion for diagnosis. That means that you might meet the symptoms for criterion (A) of MDD, but if these symptoms are not *distressing,* then you do not meet the requirements for diagnosis. And since we are assuming that meeting requirements for diagnosis is the same as meeting requirements for kind membership, then if the distress criterion isn't met, then it simply isn't MDD.[104]

So we can see that among these three requirements, all of the mental disorders have at least one individually necessary but alone insufficient property for membership: the distress criterion. And since one of the requirements for a cluster kind is that there is not an individually necessary but alone insufficient property, these mental disorders violate the core requirements for cluster kinds. So we can, tentatively, conclude that mental disorders are not cluster kinds. Why tentatively?

---

[103] Spitzer & Wakefield make this point well in their (2013) analysis, as well as in Narrow *et al.* (2009).

[104] I have here called this criterion the distress criterion, but it obviously can, as (B) indicates, be fulfilled by an individual's impairment in any of the important areas of functioning listed. I will focus on the distress element listed in (B) throughout, but you can substitute impairment for distress in my arguments *mutatis mutandis*.

It is not obvious that the cluster kinds account strictly requires that there be no individually necessary but alone insufficient properties in the cluster kind. And if that is the case, then there has been no violation of the core requirements of cluster kinds at all. Let's turn to this possibility now.

6] Cluster kinds that can make exceptions

What if I simply admit that I've made a mistake, that cluster kinds can reasonably include individually necessary but alone insufficient properties in the cluster? In fact, when we look back at our original example with cats, it seems that being a mammal is a necessary condition for being a cat, but being a mammal is alone insufficient for being a cat. So we have other natural kinds in the natural sciences that motivate the inclusion of individually necessary but alone insufficient properties in cluster kinds. An amendment that allows for such an inclusion would not be *ad hoc*—this amendment would not merely be made to avoid the issue that individually necessary but alone insufficient membership properties poses for mental disorders as natural kinds.

In fact, cluster account theorists have made similar moves in the past to account for kinds like that of gold—kinds whose members seemingly *always* share some of the same properties in common.[105] Consider our earlier example of gold. Gold is a special case in terms of such amendment, since having 79 protons in the nucleus will be *both* necessary *and* sufficient for membership. There are feasibly other properties in the cluster kind gold; it has an atomic mass of 197, it is solid at standard temperature and pressure, it is resistant to acids, etc. On the face of it, though, only having 79 protons in the nucleus can be candidates for individually necessary and sufficient conditions for membership in gold. Other elements are solid at standard temperature and pressure and are resistant to acids. What picks out the candidate properties for being necessary and sufficient for membership? The cluster theorist who favors an amendment like this would point to the mechanisms that mutually promote

[105] See especially Wilson *et al.* (2007).

the expression of the properties in the cluster. In this case she might, for example, indicate that the laws of nature function such that atomic mass and number are the only such candidate properties for gold. Why? They are the only candidate properties for being individually necessary and sufficient for membership *because they bear the strongest relation to the natural laws that establish the cluster*, whatever those are. So explanations like these depend on the mechanisms holding the cluster together, since they determine which properties are necessary, and sometimes also sufficient, for membership.[106]

What happens if we accept the amendment? My argument against mental disorders as cluster kinds so far would fail. The argument fails because there would be no requirement that there be no individually necessary but alone insufficient properties in a cluster. So mental disorders that have individually necessary but alone insufficient properties are not violating the core requirements for cluster kinds.

This concession would not be altogether damning for my argument though. We still have yet to see my second argument that states that mental disorders fail to be cluster kinds because they lack the required mechanism to mutually promote the expression of properties in the cluster. If this argument is still successful, then my argument against mental disorders as cluster accounts still stands. Even more so, since any amendment that allows for there to be individually necessary properties in the cluster will rely on the mechanisms to differentiate between clusters that have these kinds of properties and clusters that do not.[107] And if the cluster lacks these mechanisms, then they cannot be used to form the cluster, let alone support the inclusion of individually necessary but alone insufficient properties. Let's turn to this argument now.

---

[106] Bird (2007).
[107] *Ibid.*

7] No functions for mental disorder clusters

One requirement for cluster accounts of natural kinds that we cannot set aside is the requirement that there be some mechanism that promotes the mutual expression of the properties in the cluster. This mechanism holds the cluster together and makes it into the kind of grouping we can call a kind, so the reasoning goes. But what are these mechanisms? Let's look at our example of the gold atom. That a gold atom has the atomic number 79 is determined by the number of protons it has in its nucleus. Gold's property of being solid at standard temperature and pressure is linked to the structure of the gold atom, which is in part dependent on the number of protons in the nucleus. What makes it the case that the nucleus has 79 protons, and that gold is solid at standard temperature and pressure? According to the cluster theorist here, we must say that there is a mechanism that causally maintains these properties. What is that causal mechanism? In this case we might say it's a function of a natural law.

As we saw earlier, these mechanisms can be unpacked in a multitude of ways. But in each characterization of these mechanisms, they are importantly determinative; mechanisms make it the case that the properties in the cluster are expressed. The most common way that these mechanisms have been unpacked, as we just saw in our gold atom case, is as causal mechanisms.

Let's see how causal mechanisms might play out in our case of MDD from before. Given our discussions in §4.2, let's focus on the primary symptom list and the distress criterion.

> Major Depressive Disorder:
> A. Five (or more) of the following symptoms…
> (1) Depressed most of the day, nearly every day as indicated by subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful)
> (2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by subjective account or observation)
> (3) Significant weight loss when not dieting or weight gain (e.g., change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day
> (4) Insomnia or hypersomnia nearly every day
> (5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)

(6) Fatigue or loss of energy nearly every day
(7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
(8) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
(9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

We might easily fill in mechanisms that link some of the properties in this list. Insomnia (4) can cause loss of energy (6), psychomotor retardation (5), and diminished ability to think or concentrate (8). Since insomnia can cause those three symptoms, and is itself a symptom of depression, the mechanism that causes insomnia can appropriately link it to the remaining symptoms. If that mechanism can also causally determine either (1) depressed mood or (2) anhedonia, then we have a nice case of MDD with a causal mechanism that lends itself well to cluster kindhood. In fact, it seems reasonable that insomnia can causally determine (1) or (2) since there is plenty of data that suggests that insomnia can lead to depression.[108] And it looks like there's a reasonable causal link between the conjunction of these symptoms to (B) the distress criterion being met. So it looks like we've got a case of MDD as a cluster kind that works: MDD has clustered properties mutually expressed on the basis of a common causal mechanism—whatever the mechanism is that produces insomnia. But we haven't settled the issue just yet. Why?

7.1) Property overlap and mechanism individuation: For one, what makes this a case of depression rather than simply a byproduct of insomnia, which is itself a separate, diagnosable mental disorder? Remember that our causal mechanisms are supposed to be doing the work in determining the boundaries between cluster kinds. But now we see that the causal mechanism that picks out all the relevant properties in this cluster that *seems* entirely related to insomnia, *also* works as a mechanism for depression. Separate

[108] See Li, L., Wu, C., Gan, Y. *et al.* (2016) for a meta-analysis overviewing a multitude of studies on this connection.

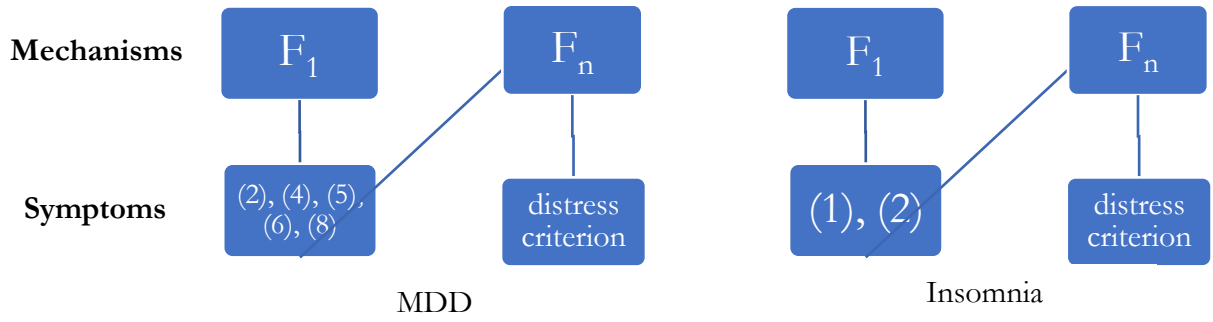
cluster kinds are supposed to be individuated by their causal mechanisms, and here we see that failing at the outset. The only reason we might think it's simply insomnia is because we *started* with insomnia in our example. Under those conditions, it seems clear that if we had to pick which (primary) disorder it is between depression or insomnia, we'd pick the latter. In fact, that seems to be the best option in response to a concern like the one I've just raised—simply look at the actual facts about the origin and progression of the properties, and you have your answer.[109] Still, the fact remains that there are two disorders that share at least one causal mechanism. So now we must appeal to something other than that mechanism to differentiate between the disorders. But the cluster kinds account *requires that the mechanism does in fact differentiate between disorders*. So any account that, given the same causal mechanism, relies on the order of presentation of symptoms to individuate between disorders is not purely a cluster kind account. So we need to shift to using more than one mechanism to maintain and demarcate separate disorders.[110]

Let's look at our example again. Let F_1 be the mechanism that produces the following symptoms of MDD: (2), (4), (5), (6), and (8). Let F_n be the mechanism that produces the distress criterion (B) that is required for MDD. Everything else remains the same as our previous example, so F_n will still be causally resultant from the presence of the other symptoms. The difference now is that F_n is an additional causal mechanism. And we will do the same thing for the mechanism for just insomnia and the distress criterion for that disorder too. Let's make our story for insomnia include the two symptoms difficulty falling asleep (1) and difficulty staying asleep (2), plus the distress

[109] See Tsou (2015). In particular, Tsou emphasizes etiological analysis of mental disorders to determine which are valid—where valid is understood in the ontological sense of carving nature at its joints.

[110] This was the assumed view at the start, that there would be at least one, but maybe more than one mechanism that maintains the cluster. We start here from the one mechanism view because it is useful in seeing why we shift to many mechanisms, and we will see the issues for one mechanism recur or compound in accounts of cluster kinds with many mechanisms.

criterion.[111] Now let's see if using more than one mechanism can resolve our inability to individuate between distinct disorders. You might already see where this is going generally, but let's see it explicitly.



We can see here that the use of two mechanisms does not resolve the problem of selecting between two distinct disorders. In order to individuate the disorders, even without our story, we would still need to appeal to the distinct symptoms that result under F_1. There are now two causal mechanisms that are shared between two distinct disorders, and still these two mechanisms cannot differentiate between the disorders. We can make this even more evident by imagining we do not know the respective symptoms each number represents in the list under F_1 and simply treat them as a placeholder. If we have only the placeholders and remove the disorder labels under each graph, then there is no way to tell the difference between these disorders besides pointing to the dummy symptoms.[112] And we have already established this is not a viable method to individuate disorders under a cluster kinds account. We must use the mechanisms.

[111] NCBI (2023).

[112] Tobin (2017) raises a similar concern about overlapping mechanisms, and how overlapping mechanisms pose similar issues for cluster kinds as overlapping properties do. Tobin's argument is toward a different goal than my own here, but much of her argument can be borne out on our cases as a practical example.

You might object at this point that these are not viable candidates for mechanisms. Many philosophers of science and metaphysicians have made objections to cluster theoretic accounts through similar concerns about mechanisms.[113] But my project here is not to argue against accounts of cluster kinds generally, so I will not spend time rearticulating these arguments. My focus is to determine whether mental disorders can reasonably fit the requirements for cluster kinds, given the most charitable account of cluster kinds. And, as we have seen so far, the mechanisms requirement presents a problem for cluster kinds even after we introduced more than one mechanism. And the mechanism, regardless of what it is, is playing two important roles: allowing for generalizations to be made about members and individuating discrete mental disorders. These roles are fundamentally explanatory ones. Without the mechanism, the cluster kind does not have the explanatory tools that are required to be a kind. Since mental disorders are things such that their properties exhibit such a significant amount of overlap, this role of the mechanism that allows for predictive inferences is even more important for mental disorders. So what can we do?

7.2) Compounding issues with property overlap: Without falling too far down the rabbit hole of mechanism discovery and its theoretical scaffolding, we can already see that so far the mechanisms on the table are too inclusive. They allow for more individuals to be members of a kind than they should—and potentially for them to be members of more kinds than they should. And without these mechanisms, under the assumption that expressing the properties for membership is equivalent to meeting the requirements for diagnosis, we can see how the practical implication is misdiagnosis. Why? With property overlap being a feature of cluster kinds, we can easily retell our story about insomnia versus depression as a story about insomnia and anxiety versus depression.

[113] See especially Craver (2009) and Tobin (2017).

In this new story, the properties the individual expresses in a case of both insomnia and anxiety are near perfectly overlapping with those that would be expressed in a case of depression. How? Let's take a look at the diagnostic criteria, minus the adjudication conditions, for generalized anxiety disorder.

> Generalized Anxiety Disorder
> A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).
> B. The person finds it difficult to control the worry.
> C. The anxiety and worry are associated with three or more of the following six symptoms (with at least some symptoms present for more days than not for the past 6 months).
> 1. Restlessness or feeling keyed up or on edge
> 2. Being easily fatigued
> 3. Difficulty concentrating or mind going blank
> 4. Irritability
> 5. Muscle tension
> 6. Sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep)
> D. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.[114]

Now imagine that you have complaints about persistent worries that you are not doing enough in work or school, and these concerns keep you up at night. This inability to fall asleep coincides with your persistent worries but has also itself become a worry. Since you are not sleeping, you struggle to do things you were once interested in, find yourself fatigued throughout the day, have trouble concentrating, but because you are worried about school or work you feel restless all day and can't stop moving.

It seems like our story is one in which you could reasonably have two distinct diagnoses. You could be a member of both the insomnia cluster and the generalized anxiety cluster, or you could simply be a member of the MDD cluster. And both stories are equally reasonable, since they both include a cluster of properties and relevant mechanisms that permit inductive inferences about an individual that expresses the properties in the story. But we cannot tell in this story if you've got

[114] NCBI (2023).

insomnia and anxiety or if you've got depression on the basis of the symptoms alone. And, as we saw before, there is a case to made that the mechanisms involved all also overlap. So we can only appeal to the way the story is told, the actual facts about symptom progression, to discern diagnosis. But mechanisms are supposed to be doing the entirety of the explanatory work to differentiate between one disorder and another. How do we fix this?

If we aim to fix this by hyper-specifying the mechanisms involved, say by picking out single mechanisms for each property the individual expresses, we simply push back the problem for individuating mental disorders. Not only that, but now the cluster does not have the tool it requires to explain why these properties are going around together. On the standard view, there must be a unifying mechanism that allows for there to be reasonable inductive inferences about the members of the cluster kind by ensuring that these 'sociable' properties in the cluster are mutually promoted.[115] Instead, by hyper-specifying the mechanisms we have gone from cluster kinds being too inclusive to too exclusive. Now mechanisms are only doing determinative work for establishing a single property. And a cluster account requires that there be a set of properties that each individual member could express, without requiring that every member express all those properties.

It is easy to get too bogged down in the details about mechanisms for mental disorders. So let's look at a contrast case where the use of cluster kinds works, in part, because the mechanism is neither too inclusive nor too exclusive. It functions exactly as the cluster theorist intends. A contrast case with a physical disorder that works in the cluster kinds account will show us, intuitively if not explicitly, how these same cluster kind accounts cannot capture mental disorders.

[115] Chakravarty (2011).

8] A contrastive case: multiple sclerosis

Take, for example, multiple sclerosis (MS). MS is a chronic illness that impacts the central nervous system. MS triggers an abnormal response in the human body that causes the immune system to target and damage the myelin and, eventually, the nerves underneath. This leads to disruptions in communication between nerves, which can result in a multitude of symptoms.[116] An unexhaustive list of the symptoms in the MS cluster includes: difficulty walking, numbness or tingling in limbs, difficulty with finding words, blurred vision, eye pain, dizziness, and fatigue.[117] Each individual with MS can experience a distinct set of symptoms at distinct levels of severity—the presence of individual symptoms will depend on the immune system response in each person.

Still, those with MS all have the same disorder. So we see clearly that MS meets both of the two requirements for a natural cluster kind. There is a mechanism that uniquely picks out the disorder from other disorders: the immune system attacking the myelin of the central nervous system. So an immune system response that targets the myelin in the peripheral nervous system, for example, is a distinct disorder, Guillain-Barré Syndrome.[118] So we have a mechanism that individuates the cluster from others. On top of that, this mechanism supports a cluster of symptoms that mutually promote the presence of the others, but none of which are individually necessary but alone insufficient. We saw that this requirement was met when we saw that each individual with MS can experience a distinct set of symptoms, all at varying degrees of severity.

8.1) Mechanism that acts destructively on body: It is these types of examples in physical illnesses that motivate treating mental illnesses through the same cluster kind lens. MS is a perfect fit for the cluster account of natural kinds, and we need not go far to see there are plenty of other physical cases that fit the

[116] NMSS (2023a).
[117] NMSS (2023b).
[118] NIH (2023).

requirements for natural cluster kinds too.[119] The motivation proceeds, in part, from the fact that we kept returning to the drawing board to research and discover the mechanism that makes the symptoms of MS cluster and appear together. And doing the work of discovery confirmed that MS was in fact a natural cluster kind, like we assumed. So the assumption is that wherever we see observable properties going around together like in the case of MS, with enough research we will discover the underlying mechanism for the cluster kind. And when we discover the underlying mechanism for that cluster kind, the presence of any of the symptoms will be fully explained. But in the case of mental disorders, there are problems with this assumption.

When we look for a mechanism for mental disorders like we do with something like MS, we tend to look for a singular mechanism that acts destructively on *a part of the body*. The view that there is a physical component that establishes the symptoms of a mental disorder is sometimes called the strong medical model. The strong medical model for mental disorders is, as a result, inherently reductive of mental disorders; it seeks to translate mental disorders into the kind of easily packageable cluster kinds we see in physical cases like MS through the discovery of physical mechanisms or, failing that, biomarkers.[120] Biomarkers are supposed to be biological signs for the presence of a condition, where we understand that condition to be a cluster kind with the appropriately unifying mechanism. Temperature, blood pressure, and body mass index are general biomarkers that give a window into abnormal processes in the human body. Specific biomarkers indicate the presence of specific conditions, like high blood sugar levels indicating type 2 diabetes,[121] or high levels of prostate-specific

[119] COVID-19, Guillain-Barré, Amyotrophic Lateral Sclerosis, and Celiac Disease, to name a few.

[120] The discovery of MS uses many methods of screening to determine the diagnosis. An MRI or spinal tap can yield screening biomarkers that indicate whether MS is the likely condition of interest. However, there is not a diagnostic biomarker for MS. Diagnostic biomarkers more directly indicate the presence of the condition of interest. The distinction between biomarker types is not especially relevant for our project here. It should be noted, however, that the most common type of biomarker is a screening biomarker rather than a diagnostic biomarker.

[121] Ortiz-Martínez, M. *et al.* (2022).

antigen indicating prostate cancer.[122] But there is not obviously going to be a single mechanism that acts destructively on a part of the body to produce the variety of symptom subsets present in the case of mental disorders. And assuming that there must be is, as I have said, reductive. Moreover, since there is not clearly a physical mechanism that does the work of holding the mental disorder cluster together, there is no reason to think that a biomarker would be present anyway.[123]

In fact, on top of the already precipitously declining popularity of the medical model even among theorists outside of philosophy,[124] there is a growing body of scientific evidence that suggests we abandon the search for biological mechanisms for mental disorder.[125] So it is relatively uncontroversial for me to claim now that we do not have the mechanisms necessary to support mental disorders as cluster kinds, and that it is unlikely we ever will. There are too many normative and social factors that causally contribute to mental disorder, whether etiologically or by mediation of an existing

---

[122] Hutch (2023).

[123] Biomarkers, as we see in these example cases, pick out the presence of other physical properties that appear as a result of the same mechanism that produce the symptoms of the disorder. Effectively, biomarkers are either by-products of symptoms in a cluster or a symptom in the cluster itself. So biomarkers are not the mechanism that does the work of holding the cluster together. Let's return to one of our example cases of biomarkers to see why. In type two diabetes, high blood sugar levels are another result of the mechanism for the disorder, which is the body's resistance to the using the insulin it produces to process the sugar in the bloodstream. So we see that biomarker is a symptom of the disorder. Biomarkers, as we understand them so far, only pick out symptoms or symptomatic corollaries that result from mechanisms. And if there is not a physical mechanism for any given mental disorder, then of course there would not necessarily be a biomarker for it either.

I have argued before that biomarkers are not the way to validate a mental disorder (2021). In fact, I am not the only one to have made arguments against the search for biomarkers for mental disorders. But the crux of my argument amounts to the following: looking for biomarkers *assumes* that there *must* be an underlying, unifying physical mechanism that acts destructively on the body and that mechanism always produces a biomarker. This is not to deny that there are some disorders that can and do produce as symptoms or symptomatic corollaries a biomarker. But the simple fact remains that for those disorders the presence of that biomarker can vary, and the disorder still be present in the individual. This variance means that biomarkers, even when present, are not a reliable measure for validating, let alone diagnosing that mental disorder. On top of this in-group variation for the presence of biomarkers, the presence of biomarkers for some disorders does not mean that every mental disorder can and does have a biomarker. These two facts together have led even researchers to conclude that the search for biomarkers in mental disorders has been unsuccessful. So the relentless pursuit of a biomarker in lieu of the discovery of that physical mechanism should be regarded as question-begging.

[124] The recent shift away from precision medicine even in the Research Domain of Criteria (RDoC), to start.

[125] For an excellently thorough account of these failures, see Anne Harrington's *Mind Fixers: Psychiatry's troubled search for the biology of mental illness* (2019).

disorder, to make clearcut biological mechanisms the foundation for a cluster.[126] But, one might reasonably object that claiming there it is unlikely there are such mechanisms is not the same as claiming it is impossible for mechanisms to do the work we need for a mental disorder cluster kind.

8.2) Mechanisms for unifying disorder: Let's return to our example case of MS. MS highlights another intuitive contrast problem for mental disorders as cluster kinds. The case of MS had a mechanism that obviously supported a number of possible subsets of properties that an individual might express in order to be a member of that kind. By contrast, in our example case of MDD the mechanism involved only reasonably produced one possible set of symptoms for membership. It was not obvious that the example mechanism could reasonably promote the mutual expression of the entirety of the properties possible for MDD membership—there is, of course, more than one way to have MDD! So we see that there might be need for more than one unifying mechanism for the possible combinations of properties in the symptoms that an individual might express for membership. And if that is the case, then we would need a mechanism that unifies all these mechanisms. Then such a mechanism must in turn unify all expressions of subsets of the properties in the disorder, in order to be a cluster kind. If that is the case, and it seems it likely would be, then we have a similar problem for MDD as we had when we hyper-specified our mechanisms for each individual property in the cluster.

When we hyper-specified our mechanisms in §7, we had to find a reasonable unifying mechanism for all these individual mechanisms that can explain why these properties are going around together. Now we would have to find one that can explain why these subsets are all instances of the same disorder rather than individual disorder clusters.[127] It is acceptable on our current understanding of MDD that individual members have distinct experiences of the disorder, like with MS. It would be

[126] Oakley, Chapter 1.

[127] This concern has been raised in other forms in the literature surrounding problems with mechanisms for cluster kinds, particularly in Tobin (2017) and Craver (2009). I have here parsed it out in terms of mental disorders to highlight the issue of explanatory unity in cases of distinct disorder experiences.

contrary to our understanding of MDD, for example, that everyone with distinct subsets of the properties in the cluster *actually have distinct disorders*. And if it must, then this is another place where the contrast between MDD and MS highlights the failure of mechanisms for mental disorders as cluster kinds. MS as a cluster kind gets an explanation of distinct experiences of the disorder for each individual from the same mechanism. When we have to hyper-specify in cases of unique experiences of MDD into distinct disorders because of distinct mechanisms, we have missed our mark. If we do not, it is hard to see how any mechanism that can cover the distinct experiences of mental disorder between individuals can be a mechanism that acts destructively on the body.

So we can see that the original concern about the individuation of mental disorders, and attempts to resolve it, have come at the expense of the unity of a cluster. The unity of a cluster allows for generalizations about members, and not for generalizations about non-members of a cluster kind. Moreover, our understanding of the mental disorder is not commensurate with distinct experiences of the disorder as requiring distinct disorders. We are stuck with two related issues for mechanisms for mental disorders as cluster kinds: First, the growing body of scientific data that suggests that we cannot appeal only to such mechanisms. Second, even if we could find the requisite mechanism, it would have to be so far removed from the individual experience of the disorder to allow it to cover any possible instance of the disorder.

Such a level of removal is hard to view as solely determinative and explanatory of the presence of symptoms, causally or otherwise. And that is what the cluster account would require for a mechanism. So mental disorders violate the requirement for mechanisms in cluster kinds. Without such mechanism we cannot make a reasonable exception to prohibition against individually necessary but alone insufficient properties for membership in a cluster kind. So mental disorders initial violation

of this first requirement for cluster kinds reasserts itself. Because they violate both minimal requirements for cluster kinds, mental disorders cannot be captured in cluster kinds accounts.

But again, one might reasonably still object: the claim that it is unlikely we will find mechanisms for mental disorders is not the same as it is impossible to find the requisite mechanisms. Of course, I agree. So we cannot rightly rule out mental disorders as cluster kinds. Yet. Let's pretend like we can find such mechanisms for mental disorders and see what happens.

9] Can idealized mechanisms accommodate exceptions? No.

Suppose there is some unifying and fully explanatory mechanism *F* that causally determines all the symptoms of some mental disorder *D*, save the distress criterion. The mechanism *F* can be fleshed out however you like, so long as it fully determines the symptoms of whatever disorder *D*. We no longer are violating the mechanism requirement for cluster kind accounts. But we are still left with the problem of the distress criterion. We must explain the presence of the distress criterion with the mechanism *F*. So the question now is, given that this mechanism *F* that fully explains the presence of all the other symptoms of a disorder *D*, does this mechanism also explain the presence of the distress criterion in the appropriate way? I argue that no. It cannot. Why?

First we must ask: what is the appropriate way that some mechanism *F* must explain the presence of the distress criterion? Recall from §6 that the relationship between the mechanism and a property that is individually necessary but alone insufficient must be the strongest in the cluster. What does this mean for our purpose here? It means that the presence of the exceptional property must be *fully explained by the mechanism alone*. Let's see if the distress criterion can be captured by our mechanism *F* in this way. Before that, though, we will need a little more information about the distress criterion than what we already have.

Recall that we are treating "meeting the requirements for diagnosis" as equivalent to "exhibiting the properties required for membership." Now the distress criterion, as a reminder from our MDD case, reads as follows:

> Major Depressive Disorder:
> ….
> C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Most importantly, we must remember that you *must meet the distress criterion* to be diagnosed. You can exhibit all the symptoms of a disorder and without the distress criterion being met, you are not disordered. This is what makes the distress criterion an individually necessary, but alone insufficient property required for membership.

Should we specify what counts as distress? We will operate on a general agreement that distress, while vague in boundary, is clear enough for a working understanding for our purposes here. I will assume that distress is something more than being stressed, bothered, or annoyed, and that it is something that brings an individual a type of pain—but specifying *when* someone reaches distress beyond this working construal is the subject of another argument entirely. As such, I will not be concerned with the *when* question of distress here. Moving forward from this assumption, we must get clear on *how* the distress criterion can be met.

Functionally, the distress criterion can be met in three distinct ways. The first is that you have symptoms that are distressing someone else. The second is that your distress is reported for you. The third is that you are distressed. Now that we know how you can meet the distress criterion, let us examine whether our mechanism *F* can adequately capture each way the distress criterion can be met. Let's start with the first.

9.1) The distress of others: Can the mechanism *F* adequately capture the distress of someone else? To help us answer this question, let's look at an example case. Meet Timmy. Timmy is an eight year old

boy. Timmy has some mechanism *F*. That mechanism *F* causally determines that Timmy also has some set of symptoms *S*. Timmy's mechanism *F* fully explains the presence of his symptoms *S*. Those symptoms together constitute the requirements for diagnosis with a disorder *D*, save the distress criterion. All that Timmy lacks is meeting the requirement for the distress criterion. Now meet Timmy's mother. Timmy's mother finds Timmy's symptoms distressing. We have all heard this kind of story before, so it is easy to imagine how this distress is impacting Timmy's mother. She is annoyed, sure, and she is bothered—but it has gone beyond that into distress. Timmy's mother's distress together with Timmy's mechanism F and symptoms S now qualify Timmy for diagnosis with disorder *D*. It is only once this distress criterion is met that Timmy can be actually diagnosed. Without Timmy's mother's distress, Timmy is not disordered.

Before we even analyze whether our mechanism *F* can capture this fulfillment of the distress criterion, we have a few problems. The first is that you might think, reasonably, that it is strange that Timmy's diagnosis is dependent upon his mother's distress. To put it alternatively, in order for Timmy to be an individual member of a kind, you have to consider both his properties and the properties of another individual. An individual's membership in a kind is, we tend to think, determined by the properties of *that* individual—particularly when the kind in question is supposed to be a natural one. Surely any member of a natural kind cannot reasonably rely on the properties of a member outside the kind to be considered a member. And yet here we have Timmy's membership, or diagnosis, being partially determined by the properties of his mother. We can abbreviate this issue as *the metaphysical contention* for the first way the distress criterion can be met.

The second problem is that there are other, perhaps even more normative, social, practical, and even political reasons why others would also want to set aside this method for fulfilling the distress criterion. Others have raised and examined these issues with greater care and detail than I can give

here.[128] And while there are also normative, social, practical, and even political reasons for including the distress of others in mental disorder categories,[129] it is not at all clear that these outweigh the reasoning against doing so. Given these concerns, together with the metaphysical contention raised above, we will abandon this first way of meeting the distress criterion. Now only two remain. So now the question is whether these two methods of meeting the distress criterion can be captured by mechanism *F*.

9.2) The Report of Distress in an Individual: Consider Timmy again. Timmy is still age eight, he still has some mechanism *F* that is causally determinative of some set of symptoms *S*. That mechanism still fully explains the presence of this set of symptoms. Those set of symptoms *S* together constitute what is required for diagnosis with some disorder *D*, save the distress criterion. Enter Timmy's mother. Timmy's mother calls the school psychologist, and she reports that Timmy is distressed. She reports, he has this set of symptoms *S* and he is distressed. We can imagine how this conversation goes: Timmy is distressed, it is affecting his schoolwork, it must be because of these symptoms *S*.

Now, since we are in the business of determining whether Timmy is disordered (because we are in the business of determining whether Timmy meets the conditions required for kind membership), we must decide if this report constitutes meeting the distress criterion. What it seems like, from this report, is that we can conclude, at <u>most</u> that Timmy is *possibly* disordered. Why? With the report of the distress, we have three options for Timmy's disorder status.

1. Timmy's mother reported his distress wrongly. He has the mechanism *F*. He has some set of symptoms *S* that are causally determined by *F*. But he is not distressed. So Timmy is not disordered.

---

[128] See, for example, Phillips (2009), Spitzer &Wakefield (1999), Bolton (2013).

[129] For example, if you were to discharge the distress criterion requirement by looking at the impairment or disruption of individuals around you in important areas of functioning—say Timmy's mother cannot keep a steady job because her son's mental disorder is such that she must be available to help him cope, take him to various therapies, etc.—you can see why this criterion reasonably extends to individuals outside the diagnosis.

2. Timmy's mother reported his distress correctly. Timmy has some mechanism *F*. Timmy has a set of symptoms *S* that are causally determined by *F*. Together with the distress criterion being met, as indicated correctly by the report, Timmy is disordered.
3. Timmy's mother reported his distress correctly. Timmy has some mechanism *F*. Timmy has a set of symptoms *S* that are causally determined by *F*. But Timmy is not distressed by *S*. He is simply distressed about an unrelated set of phenomena—he is being bullied on the playground at school. The distress criterion is *not* met, although Timmy's distress was indicated correctly by the report. So Timmy is not disordered.

In only one of these cases does the report constitute an instance in which the distress criterion is actually met. As we can see in (1) and (2), the truth of the report matters to whether Timmy is really a member of the kind. But, as we can see in (3), it is not *just* that the report of distress is a true one. It matters that the distress is of a relevant sort. What sort?

The distress must *determined* by the set of symptoms *S*. Put alternatively, Timmy's distress *must be about S*. If Timmy is simply distressed, as we see in (3), it is not sufficient for diagnosis. Remember, if the symptoms are not distressing, it is not a disorder. That is what makes the distress criterion a property required for membership that violates the prohibition of individually necessary but alone insufficient properties for cluster kinds. It is the symptoms that causally determine the stress. It is even written explicitly into the specific version of the distress criterion we started our examination with when we looked at MDD.

So where does that leave us with our analysis of whether our mechanism *F* can adequately capture the distress criterion as met in a report of an individual's distress? Since we can only determine if an individual has a disorder if the report of distress is both i) a correct report and ii) the distress is about the symptoms, then *F* fails to adequately capture the presence of the distress criterion. Why? Recall that the relationship between the mechanism and a property that is individually necessary but alone insufficient must be the strongest in the cluster, if we are to permit an exception to the exclusion of such properties. So the exceptional property must be fully explained by the mechanism *alone*. And here we see that the distress criterion, and whether it is met, is not fully explained by the mechanism

alone. We must seek out some story to determine whether the distress is directed at or the result of the set of symptoms *S*. Furthermore, in virtue of its role in mental disorder criteria, the distress criterion must always be mediated by *S*. So the distress criterion is not a property that bears the strongest relationship to *F*, and cannot be explained solely by it. So it cannot be an exception to the rule that prohibits properties that are individually necessary but alone insufficient in cluster kinds. Since mental disorders violate these minimum requirements of cluster kinds, mental disorders cannot be cluster kinds.

Is this too quick? We might push back here by denying that reports of distress are a viable method for meeting the distress criterion. After all, we only get, at best, some individual is possibly disordered on the basis of these reports. But, with our last way of meeting the distress criterion, I show that pushing back in this way only forestalls the issue. The distress criterion cannot be captured by our mechanism *F* alone, and the report of distress by another only serves to highlight the issue that exists in cases where the distress is simply present. Let's turn to that next.

9.3) Distress in the individual: Timmy is now a grown man. Timmy has some mechanism *F* that is causally determinative of some set of symptoms *S*. This mechanism *F* fully explains the presence of this set of symptoms *S*. Those set of symptoms *S* together constitute what is required for diagnosis with some disorder *D*, save the distress criterion. Furthermore, Timmy is distressed. Is Timmy disordered?

It is obvious by now that we must ask another question before we can answer this one. We must ask: Is Timmy distressed *about S*. If yes, then Timmy is disordered. If no, then Timmy is not disordered. To answer that question, we have to tell a story about the relationship between Timmy's distress and his set of symptoms *S*. When we tell such a story, we are forced to make an appeal to something outside of the mechanism alone to explain the presence of the distress. We know that on a cluster account, an exception is only allowed for individually necessary but alone insufficient

properties whenever they are fully explained by the presence of the mechanism alone. Yet our mechanism *F*, that otherwise fully explains and determines the rest of the symptoms *S*, cannot explain the presence of Timmy's distress without the help of the symptoms. So the mechanism does not fully explain the presence of Timmy's distress alone. Therefore, even our toy mechanism *F* cannot adequately capture the distress criterion. So the distress criterion cannot be an exception to the rule that prohibits properties that are individually necessary but alone insufficient in cluster kinds. Since mental disorders violate these minimum requirements of cluster kinds, mental disorders cannot be cluster kinds.

I reached this conclusion using an idealized mechanism *F*. I have already argued in §8, there are plenty of reasons to think that a mechanism that can play the role required for cluster kinds is *extremely unlikely*. Now I have shown that, even if there were such a mechanism, the special quality of the distress criterion—the quality that makes it an individually necessary but alone insufficient property for diagnosis—explicitly prohibits *any* such idealized mechanism from adequately capturing distress at all. So it is impossible for mental disorders to meet the minimum requirements for cluster kinds. It is impossible even with an idealized, provisional mechanism; one that is already a concession I made to avoid a previous failure for mental disorders to meet the minimum requirements for cluster kinds. So mental disorders are not cluster kinds, and due to the unique quality of the distress criterion, cannot be cluster kinds.

10] Mental disorders are not natural kinds.

I began this project by saying that if mental disorders are natural kinds, then they are either essential kinds or cluster kinds. I concluded early (§2) that mental disorders are not essential kinds. This conclusion is not controversial. The received view of mental disorders as natural kinds has them as cluster kinds (§3-4). I have now shown that mental disorders are not cluster kinds (§5-9). It follows,

therefore, that mental disorders are not natural kinds. And so long as the distress criterion remains, they cannot be.

As I said at the outset, that does not mean that I affirm that mental disorders are not natural. Nor does it require that I assert that mental disorders are not real. In fact, I assumed from the outset that we must always take the default position that mental disorders are very real. And I assert that they are also natural. Nothing about the revocation of their capacity to be captured in our best natural kinds accounts necessarily contravenes these commitments. So my view is not an anti-psychiatric one. It is simply one that shows that mental disorders are not candidates for natural kindhood. My argument also shows that further serious attempts to give accounts of mental disorder must either accommodate and explain the presence of the distress criterion or give good theoretical and practical reasons for its abandonment. Neither of these projects are small tasks, though. And, I suspect, neither will make a dent in the mess.

III.

So far I have argued that any account of mental disorder must start from a specific interpretative position of our diagnostic manuals. I have argued that this position is required because of the goals, both in the field of philosophy and those of psychology and psychiatry, for our best accounts of mental disorder. Furthermore, I have argued that this interpretative position is ultimately one that precludes our ability to avoid a significant issue for our diagnostic categories: the problem of reification. Given that the problem of reification will, as a result, apply broadly to *whatever* account we give of mental disorders, I suggest that we still attempt to refine our conceptual understanding of the nature of mental disorders. Conceptual analysis need not fall prey to reification in the same way that empirical analysis might.[130]

I moved on to examine popular accounts of the nature of mental disorder. I argued that modern accounts that attempt to encapsulate mental disorder through an appeal to biological or biochemical mechanisms alone will fail. Such accounts fail because mental disorders are incapable of being disentangled from social mechanisms, and are often determinatively established through them. Ultimately, I suggest that any successful account of mental disorder will have to include both social and biological mechanisms. Even still, given that all these accounts require we take a particular stance to interpret diagnostic categories, we still cannot establish the validity of mental disorders through an exploration of both biological and social mechanisms.[131]

But this is not a new suggestion. My argument was simply a new path to an established conclusion; one that showed that the reason why we end up at that conclusion is because of how we *must* interpret our diagnostic categories. On the understanding that we must treat the disorder

[130] Oakley, Chapter 1. See especially §1-5.
[131] *Ibid.* See especially §6-13.

categories in a particular light, I then took up the skeptical project of critically examining the standard received view of mental disorders. The goal with this project was not to debunk the reality of mental disorders. Nor was the goal to reify them, since I operate on the assumption that mental disorders *generally* are real, not that *any particular* mental disorder must be real. The goal was to simply assess whether the current philosophical framework we typically apply to mental disorders could capture them. I concluded that they could not.[132] I argued that there is a special feature of mental disorders that prohibits them being captured in the received view, regardless of any idealization or concessions we could make to try and force the fit.[133]

It is from these conclusions that I pick up the examination of accounts of mental disorder now. Here, I will examine what I think is the next most plausible step that we can take to give an account of mental disorder. I will take the contributions from the conclusions of these earlier arguments to inform the metrics for success for our new account. Ultimately, I argue that this new account also fails to capture mental disorder by nearly the same metrics.

The trajectory of the argument is as follows. First, I explain the diagnostic manual that guides my examination, the two ways that we can interpret that manual, and my previous argument that concludes only one of these interpretations is viable (§1-2). Taking this interpretive stance, I assert that there are only two primary parts of mental disorder, explain their relationship to each other, and establish their relationship to disorder (§3). Next I argue that one of the two parts plays the distinctive role of "disorder-maker." I illustrate how it plays such a role with example cases. These cases intuitively highlight the difference between cases of disorder and non-disorder and explain how this is commensurate with the professional psychological understanding of disorder (§4). Emphasizing the "disorder-maker," I revisit one failed account of mental disorder to motivate the next most promising

[132] Oakley, Chapter 2. See especially, §1-8.
[133] *Ibid.* See especially, §9-10.

account—mental disorders as interaction effects—and subsequently unpack it (§5-6). However, I demonstrate that this account cannot distinguish between cases of disorder and misdiagnosis, and that further attempts to resolve this failure only raises more problems for the account (§7-8). So I conclude that an account of mental disorders as interaction effects fails, and suggest that further accounts of mental disorders will be met with similar difficulties (§9-10).

1] The DSM:

The *Diagnostic and Statistical Manual* (henceforth, DSM) is a tool for mental health professionals. The DSM helps mental health professions by describing the diagnostic criteria that their patients must meet for any given mental disorder. This descriptive approach—the listing of the diagnostic criteria for each mental disorder—allows mental health professionals to give more uniform diagnoses and treatments to their patients. These are, to date, our best working definitions of mental disorders.[134]

2] The constitutive position:

Many mental health professionals have complaints about the DSM generally.[135] These complaints have spurred a recent shift in professional attitudes about the DSM, moving beyond viewing the tome as a "diagnostic bible" into a more antagonistic, if not begrudging, working acceptance of the reference book. But even understanding and, often, agreeing with the complaints made about the DSM, I have argued that we still cannot interpret the DSM any differently than as fully descriptive of the nature of disorder.[136] Why?

---

[134] Quickly I'll just state that the Research Domain of Criteria (RDoC) and the International Classification of Diseases (ICD) systems are not used for diagnostic practice in the USA with any consistent prevalence. Furthermore, the ICD system, where it is used, is indexed particularly to the existing criteria in the DSM. So the ICD relies on the DSM in its system. As for the RDoC, its creation was guided heavily by the existing diagnostic criteria in the DSM.

[135] Broad complaints include, but are not limited to: the reference book is supposedly descriptive of the way the world is but is ultimately formed and resultant from one early school of psychological thought; there (still) are sexist diagnostic categories included in the DSM that serve only a heteronormative, patriarchical agenda; the boundaries for setting constraints on individual autonomy are set too low in cases that require diagnosis for forced intervention; and there are too many "disorders" in the manual and more are introduced with each edition.

[136] Oakley, Chapter 1.

The DSM has two primary aims. The first aim, and the reason for the DSM's creation and subsequent revisions, was to establish internal validity.[137] For the DSM, internal validity is simply "inter-rater reliability." Inter-rater reliability ensures that a patient that presents with the same set of symptoms would be diagnosed near uniformly among mental health professionals. To put it alternatively, internal validity ensures that individuals with similar symptoms will be diagnosed such that each are members of the same diagnostic category regardless of who diagnoses them. The second aim is external validity. External validity, for the DSM, means that the DSM takes itself to be describing facts about the way the world is—that it's diagnostic categories are "accurate in representing the true state of nature."[138]

Those who find fault with the DSM often wish to take more interpretative liberties with the DSM's diagnostic categories. I have called this interpretative stance of the DSM the *heuristic position*. Broadly speaking, taking this position would allow, for example, a clinician to appeal to things listed outside of the diagnostic category as important for establishing a disorder in an individual. By contrast, taking a *constitutive position* on the DSM requires that we only look to what is given in the diagnostic criteria. Having some disorder simply *is* to meet the criteria listed in the manual because the disorder simply *is* the criteria listed in the diagnostic category. In this way, diagnostic criteria are literal and definitive descriptions of disorders. I have argued that the only interpretative position that is commensurate with *both* aims advanced by the DSM is the constitutive position.

The argument can be summed up succinctly when you consider whether the heuristic position can uphold both aims. If a diagnostician can appeal to things outside the diagnostic category for diagnosis, then the DSM fails in its aim of inter-rater reliability: it cannot ensure uniformity of diagnosis across professionals for each individual, or individuals presenting with similar symptoms.

---

[137] Frances (2013).
[138] Blacker and Endicott (2000), pg. 7.

Furthermore, if you take up the heuristic position, you must deny that the diagnostic category is "accurate in representing the true state of nature." So if you take up the heuristic position, then the DSM must fail to uphold its aim of external validity as well. So the only interpretative stance that is consistent with the DSM's stated goals is the constitutive position. Therefore, we must view our diagnostic categories as literal and definitive.[139]

3] The parts of mental disorder:

At their most basic, mental disorders have two parts.[140] The first of these parts are the symptoms that we point to when we talk about the disorder in everyday conversation, as we learn about them in coursework, or how we identify them in diagnosis. For ADHD, we look at figeting, frequent and inappropriate talking, or interruption, or the inability to sit still for long periods of time (when that is the social expectation), etc.[141] For Major Depressive Disorder, we talk about fatigue, or a lack of interest in doing things that were once enjoyable, or sleeping too little or too much, etc.[142] These are the bricks of the disorder—they represent the largest portion of our diagnostic parameters and set boundaries on the possible set of experiences of one disorder from another. I will call this first part the symptom set.

The second part of mental disorder is something I have called the distress criterion. It is also, technically, a symptom—and one that is required for diagnosis with whatever mental disorder symptom set you have. What is the distress criterion? Formally, it is that whatever symptom set an individual has, that symptom set causes "significant distress or impairment in social, occupational, or

---

[139] This goes a little faster than my original argument for this conclusion and, as a result, leaves space for some objections to be raised. For a more careful and detailed articulation of this argument, please see Oakley, Chapter 1.

[140] Here I am setting aside a third thing that appears in the diagnostic criteria but I do not consider (a) genuine elements of disorder, and so (b) subject to the constitutive view. Elsewhere, I call these conditions *adjudication conditions*. These are those criteria that serve *only* to eliminate confounding factors (exclusion criteria) and guide a professional's diagnostic judgement. See Oakley (Chapter 2) for a more detailed explanation.

[141] APA (2017).

[142] APA (2017).

other important areas of functioning."[143] Put another way, it is the requirement that a person with some symptom set *S* is distressed or impaired *by S*.

The explicit inclusion of this distress criterion began with the DSM-III. Physicians, becoming concerned with the prevalence rates of certain disorders, introduced a threshold requirement for "caseness," or what could be considered an instance of disorder versus normal variance.[144] The distress or impairment requirements worked their way into specific criteria language in the DSM-IV for nearly a third of the disorders in the manual.[145] Now in its latest iteration, the DSM-V-TR includes both the explicit language for the distress criterion and the inherent threshold requirements (sometimes called contextual criteria) baked into the symptom set.[146] So if the distress criterion is not found explicitly in the language of a diagnostic category, then that does not mean it isn't there. In fact, it is present explicitly or implicitly in all mental disorder categories.[147] So I will use the distress criterion in its explicit language as previously stated for the sake of simplicity.

I will take it for granted that "distress" offers us a minimal working understanding of the severity of the threshold—it is something worse than bothersome or annoying, and therefore causes a person some kind of pain. How we might unpack the "impairment" requirement is a complicated project of its own, though the troubles it raises for our medical understanding are structurally similar to worries about dysfunction in physical health analyses. Since this debate is ongoing,[148] I will run with distress, assuming that whatever applies to distress will also reasonably apply to impairment.

---

[143] APA (2017).

[144] Narrow *et al.* (2009).

[145] Spitzer & Wakefield (1999).

[146] APA (2017).

[147] An extensive argument for this claim is outside of the scope of my project here. Others, philosophers and mental health professionals alike, have made extensive arguments about this claim. See, for example, Narrow *et al.* (2009).

[148] For one specific argument about the failures of impairment accounts see Schwartz (2007). For a summary of the state of the debate surrounding the failures of impairment accounts of physical health, as well as specific arguments against impairment accounts, see Barnes (2023).

I also want to highlight an important feature of the distress criterion. Distress does not have to involve a person being directly upset about their symptom set, or even that a person be aware that their symptom set is causing them distress. The distress criterion can be met whenever the individual suffers some kind of harm as a result of the symptom set. Of course, the distress criterion is often exemplified in the kind of pain I previously mentioned. But this way of talking about the distress criterion is not exhaustive of how it can be met.

So now we have our two main parts of mental disorder: the symptom set and the distress criterion. The distress criterion alone is not sufficient for a mental disorder: there must be some symptom set that is distressing to you. Likewise, a person seeking a mental disorder diagnosis must meet the distress criterion to be diagnosed with a disorder. Simply exhibiting the symptom set is not enough. It is these two parts together that make a mental disorder; the symptom set and the distress determined by or directed at that symptom set. In this way, the distress criterion is the mortar that holds the mental disorder together. But hold it together in what shape?

4] The distress criterion as the disorder-maker:

We know these are mental disorders component parts, by our best working definitions. And, since I have argued that we must take these working definitions as literal and constitutive accounts of what a mental disorder is, we must take these component parts as the fact of the matter about particular mental disorders. Now it seems like we must also take literally the idea that the distress criterion holds together a disorder in a way that is more ontologically important than we might have originally thought. How?

Consider the following: Remy has been recently diagnosed with Attention-Deficit Hyperactivity Disorder (ADHD). Remy has, persistently, been one of those people that loses items of importance—keys, wallet, phone, homework assignment sheets, glasses. This has cost Remy, over the

course of their life, upwards of thousands of dollars and emotional and social harm. Remy, regardless of organizational system, cannot keep track of important dates, is always running late, and often fails to finish tasks. As a result, Remy has been left out of social groups and is regarded as unreliable among friends. At Remy's most recent family dinner, Remy's Aunt Marge suggests that Remy is being overmedicalized. Marge asserts that "everyone is a little bit ADHD—I've lost my car keys too!" Aside from the normative issues with raising this point directly to Remy at family dinner, Marge raises a common concern about mental disorders.

This is where the distress criterion comes in. Remy is distressed by the symptom set. Marge, reasonably annoyed or inconvenienced by losing her car keys that *one time*, is not obviously distressed. This is not to mention that Marge only experiences a small subset of the symptom set associated with ADHD in this case—a separate concern that we won't examine here. If we assume that Marge experiences a sufficient number of the symptoms in the symptom set, if she is not distressed by them, she is not disordered. She is right to say, in some sense, that everyone is a "little ADHD" if all that we are concerned with is meeting the symptom set requirement. Perhaps Marge has distress with her symptoms and simply is wrong about *everyone* being a little ADHD and, in fact, Marge could reasonably require diagnosis.[149] But assuming she is not distressed, where Marge goes wrong with her assertion, at least metaphysically, is that she thinks that you can have ADHD without the distress.[150] And by extension, we can apply such reasoning to any other disorder met with concerns that *everyone* exhibits its requisite symptom set (at least, *sometimes*). Sure, everyone might have some depressed mood from

[149] A lot of family members discover that their lived experiences constitute neurodivergence or mental disorder only after learning about the realities of living with these disorders through a close relative recently diagnosed. This is highlighted especially well through data surrounding parents that discover they are on the autism spectrum only through the process of their child being diagnosed with autism. See, for example, Loftus (2022).

[150] (Where Marge goes wrong *normatively* is likely more complicated. It seems like she goes wrong in one of several ways. The first is that Marge could be assuming that to find these symptoms distressing is out of the realm of possibility. The second could be that Marge assumes that her being bothered by her experience of these symptoms counts as distress, and since it has had little impact on her that Remy's distress is being overblown. The third could be that Marge doesn't think that Remy is distressed or, worse, that Remy should be distressed by the symptom set.)

time to time. But having *depression* is—as anyone who has experienced it already knows—significantly distinct from the occasional depressed mood friends might chat about over brunch. And the same for post-traumatic stress disorder, anxiety, insomnia, etc. The distress makes the difference. The distress makes the disorder.

But all of these disorders seem to bring about distress in a particular way—distress about the disorder that is a clear and obvious result of the symptoms. In fact, most of these listed have included the explicit language of distress in their disorder criteria. But let's consider a case that doesn't explicitly yield feelings of distress in the individual but instead distress is met through what is sometimes called contextual criteria. The idea here is that contextual criteria introduces the element of *harm* to the individual, like I touched on in §4, which functionally serves the role of the distress criterion by introducing threshold requirements for symptoms. To understand what this looks like, let's examine a case of addiction.

Addictions make up a class of disorders in the DSM, so let's focus particularly on alcoholism. Alcoholics will often not see their addiction and the consequences of it for quite some time.[151] So distress will not arise for the alcoholic in the same way that we might expect it to in the ADHD or depression cases—alcoholics often enjoy their addiction, they rely on it emotionally or physically, and it is often soothing to them. But we *know* that alcoholism does, inevitably, meet the distress criterion, even if the individual is not *distressed*.

We also know that there is a distinction between cases of alcoholism and cases in which individuals are heavy drinkers. If alcoholism were simply heavy drinking, then most college students would be alcoholics. The fact that some college students are alcoholics and can be distinguished from their peers that still drink heavily underscores the distinction. Heavy drinkers can and often do meet

[151] Stewart & Birch (2019).

the base symptom set for alcoholism, but they are still not necessarily alcoholics. Heavy drinkers are not necessarily disordered drinkers. It is the relationship to alcohol, the role it plays for the drinker, and the way it impacts their individual lives that makes the difference. These are the contextual criteria that we appeal to when we differentiate the disordered drinkers from otherwise normal, albeit heavy drinkers. And the implicit element in the contextual criteria is the element of individual harm.

Since, as I have said before, I will not be looking at impairment for this project here, I will not unpack these contextual criteria in these terms. But it is easy to see that we could unpack the harm here in terms of individual impairment. But the individual harm need not be packaged solely as impairment, and especially so in the case of alcoholism. Functional alcoholism highlights well that the individual harm of an alcohol centered addiction does not require that an individual be impaired in the important areas of their everyday functioning. And yet there is still harm for functional alcoholics. So we can see that in these cases, the contextual criteria are fulfilling the role of the distress criterion. When we are deciding between a case of disordered drinking and heavy drinking, we look to the contextual criteria to fill in for the explicit inclusion of distress language in the symptom set. With this in mind, let's run our family dinner table conversation again.

Remy lets the family know that they are an alcoholic. Aunt Marge suggests that Remy is being overmedicalized. Now, assuming that everyone at the table has at some point imbibed, if Aunt Marge were to say something like "we're all a little bit of an alcoholic" at family dinner, the effect of her voicing that opinion would be significantly different. First, more people at the dinner table might feel awkward, like this is a strange thing for Aunt Marge to say in response to Remy's announcement. It would seem more like Marge is perhaps saying something about herself she might not have meant to disclose, or perhaps something about the family as a whole. This is because, second, Aunt Marge is

saying something that rings false.[152] At this table of, say, social wine drinkers, it rings false because of a fact about alcoholism that most people readily accept: alcoholism is pre-theoretically distinct from drinking (even regularly).

One way that we can vindicate this pre-theoretical distinction between mere drinking and alcoholism is to point to some feature that differentiates the two. One such feature that might reasonably differentiate the two is the heritability component of addictive disorders, particularly alcoholism. The genetic predisposition to alcoholism, or other addictive behavioral patterns, is one that has become widely known and accepted. Knowledge that alcoholism is, at least in part, tied to some determinative mechanism lends weight to this mental disorder diagnosis that is not as present in one like ADHD.[153] When we are secularly aware of some mechanism that makes it possible to explain the presence of a mental disorder, we are secularly aware of at least some of the contextual criteria that would take someone from "heavy drinker" to "alcoholic." So if Marge were to make a claim that "everyone's a little bit of an alcoholic," and everyone at the table understands the heritability component of that disorder, her claim cannot sideline considerations of distress since the contextual criteria are baked into the disorder criteria. In this case, meeting the symptom set means you also meet the distress criterion. For a disorder like alcoholism, the symptom set bakes in distress by establishing the degree of severity required to be considered disordered.

On the other hand, the fact that risk factors like heritability are present when considering whether a case meets contextual criteria does not mean that the mere presence of risk factors

---

[152] Of course, it is false if a person has never had a drop of alcohol. Then of course that person could not possibly be a little bit of an alcoholic. But since I am assuming that everyone at the table has had alcohol, we can set that aside. It is also obviously false that anyone who has ever had any alcohol is a little bit of an alcoholic—the relationship to alcohol fails to be one characteristic of addiction.

[153] To be clear, recent studies suggest that ADHD is among the disorders that have high levels of heritability (Borgen *et al.,* 2021). The difference I am pointing to between ADHD and alcoholism on this score is a matter of degree—we have longer standing research, and so a greater quantity of established scientific data, to support the heritability claim for alcoholism.

establishes disorder. Let's take another family member at the table, Martin. Martin is Marge's brother and Remy's father. Martin has a family history of alcoholism. Martin also works as a crisis counselor, so his job is often high stress. Martin also was a bit of a partier in college, and so has at one point been a binge drinker. But Martin has never had even a hint of a problem with his drinking. While he has many of the risk factors for alcoholism, Martin never meets the contextual criteria for disorder. So Martin was never remotely 'a little bit of an alcoholic' either.

So, importantly, having the predisposition to alcoholism does not automatically make you an alcoholic either. If you have the predisposition for alcoholism and you are a drinker, you are not automatically an alcoholic. You must meet the symptom set *and* the contextual criteria in order for you to shift over into disorder. You might exhibit some of the symptom set required for diagnosis with alcoholism, but if you do not meet the threshold for the contextual criteria, then you are not disordered. And the same extends to any other disorder with these built-in contextual criteria in their symptom set.

And so it is not the case that everyone could be a little bit of an alcoholic—you cannot tease apart the symptom set from the distress. If you meet the symptoms in this case, you are disordered. And while the severity of the disorder can come in degrees, meeting the contextual criteria in the symptoms cannot: you do or you don't. The contextual criteria provide thresholds that must be met for the symptom to be met. As a result, if you are a little bit of an alcoholic, then you are an alcoholic.

It is also worth noting that I am interpreting Marge's claim that 'everyone's a little bit *x*' here in a particular way. Marge's claim rings true or false because, in these cases, it seems like she is in fact making a claim about *being disordered*. Most clearly in the ADHD case, Marge is claiming that people generally exhibit the symptoms, maybe even the sufficient symptom set, for the disorder—she is simply downplaying the impact the disorder has. We can contrast this with statements like "I'm

depressed about the recent Supreme Court rulings" or "I'm a little OCD about formatting citations." Neither of these kinds of popular use invocations of disorder are claims about *being disordered*. One is merely a claim about being really rather bummed about some Supreme Court decisions, and perhaps even in a way that is somewhat persistent. The other is a claim about being rather particular about meeting the exacting standards of citation style guidelines. A little weird, to be sure, but nothing on the order of claiming to meet some specific symptom set, even if there is some symptom being met. Neither of these kinds of popular use claims are claims about the presence of disorder in the individual. But Marge's claim about 'everyone being a little *x*' inherently involves a claim about, at the very least partially, meeting some symptom set.

At its core, Marge's comment, at least most obviously in the ADHD case, is a simple and true suggestion. The suggestion is that it is likely that each individual has some set of properties that, on their own, might be sufficient for meeting some symptom set in the DSM. And to that I say: of course they do! Everyone *does* lose their car keys sometimes, everyone *can* have trouble sitting still for long periods of time, and everyone *can* have periods of fatigue or depressed mood sometimes. We might even go so far as to assume that everyone, at some point in their lives, reasonably exhibits the properties sufficient for meeting a symptom set. And in the case of disorders that have included contextual criteria, we might assume instead that one exhibits the property expressed by the symptom but not to the threshold required for meeting that symptom. My suggestion here is that simply exhibiting those symptoms is not enough *on their own*, and simply meeting them *does not ensure, with any regularity that an individual will have a disorder*. It is only when these symptoms interact in a particular way that they get distress. It is only when these symptoms meet a certain threshold that they meet the contextual criteria. And it is only when we have met the distress criterion, in either of these ways, that we get disorder.

My suggestion is not new or radical. It is, in fact, implicit[154] in the construction and maintenance of the DSM. And to those theorists who already accept non-skeptical positions on mental disorder, it might seem like I am simply preaching to the choir. However, the point that I am making here with distress is that its role as the disorder-making quality in our disorder criteria contravenes our previous working understanding of mental disorder. It contravenes our previous understanding because distress as the disorder-making quality means that any attempt to give an account of disorder that works from lower-level mechanisms to establish and maintain the boundaries of disorder cannot succeed. Distress is the *result of, directed at, or included in* symptoms, and so occurs at the same level or higher than the symptom set. The disorder-maker is not determinative of the symptoms but determined by or within them.

Moreover, distress that is caused by or baked into the symptom set is not something that arises with invariant regularity from a symptom set. We *want* to approach mental disorder as something that is produced with invariant regularity that gives rise to such and such symptoms with so and so effects. But we see now that disorder only results when the properties that make up the symptom set together produce a very specific type of effect: distress. And that effect is not invariably present as a result of these properties—not everyone has ADHD despite the fact that everyone might reasonably present the symptom set *without the distress*. So these symptoms have to interact in a specific way, in a specific individual, under specific conditions—and even then a mental disorder is not guaranteed. So what is a mental disorder?

[154] I would also suggest it is often so incredibly explicit, as well. The primary or secondary differential diagnosis for any given mental disorder in the DSM is "normal variance." This just means that the creators and caretakers of the DSM recognize that meeting a symptom set is not the pure arbiter of disorder in all and every case—we must get distress explicitly or through the contextual criteria thresholds.

5] Accounts of mental disorder, striking a balance:

As it often goes with these analyses, it is much easier to say what mental disorders are not. It is fortunate for us philosophers, then, that what fails to account for phenomena can be equally, if not more, informative about their nature in our analyses. I have argued previously that mental disorders cannot be natural kinds. Of the two most viable accounts of natural kinds on offer, one account was too restrictive to capture mental disorders as we currently understand them. As a result, it would fail to accommodate distinct expressions of disorder among those with the disorder. The other of the two most viable accounts of natural kinds was too permissive—for example, it could not explain the difference between two people with distinct disorders but overlapping symptoms. Moreover, it could not adequately capture and explain the presence of the distress criterion, which we see now is the disorder-making quality among the two parts of disorder. So it should be clear, given this failure, that a natural kinds account could not capture disorder because it could not capture the disorder-maker: the distress criterion.

This argument together with the above exploration of the disorder-making quality of the distress criterion will inform the progress of our analyses from here. We know, foremost, any successful account of mental disorder must capture the distress criterion. And we know that a successful account of mental disorder, given that we must interpret diagnostic categories constitutively, must include and explain the presence of *both* parts of disorder. We know that we need an account that is not too restrictive; it must accommodate distinct expressions of the same disorder. We also need an account that is not too permissive. Our account must be able to tell the difference between a case of one disorder or another, and particularly between cases of disorder and those that are not.

A successful account must, therefore, strike a very fine balance both conceptually and practically. A successful account cannot be overly specified so as to not be applicable in day-to-day diagnostic practice nor descriptive of some feature of the world rather than a particular individual. A successful account cannot be over-generalizable, which would only obfuscate or obstruct the goal of diagnostic practice: successful treatment. Lastly, this balance must be struck using the established elements of disorder to which we are already committed and, so, must accommodate.

6] Mental disorders as interaction effects:

The natural next step is to give an account of mental disorder that accommodates the two parts—the symptoms set and the distress criterion—in a way that adequately captures their required relationship. In my view, such an account would view the symptom set as properties that mediate each other, such that the ultimate effect is the distress criterion. In the case of contextual criteria, the properties of the symptom set would mediate one another in such a way to potentiate the degree to which each symptom is expressed, thereby meeting the contextual criteria threshold. I suggest we understand the way that the symptoms work together in specific ways to produce distress, and thereby, mental disorders, as *interaction effects*. Interaction effects are not determinative of the symptoms but determined by or within them. And wherever the interaction effect is distress, and so meets the distress criterion, the interaction effect establishes a disorder.

What are interaction effects? An interaction effect occurs whenever the effect of one thing is changed by the presence of an effect from another. Since we are using ADHD, let's take a look at what an interaction effect is in terms of a common ADHD medication. Dextroamphetamine is a drug used in the treatment and management of ADHD. It is a central nervous system stimulant.[155] Caffeine is a drug used in the daily management and functioning of most adult lifestyles. It is a central nervous

[155] Shoar *et al.* (2023).

system stimulant.[156] If one were to take dextroamphetamine and wash it down with a cup of coffee, an interaction would occur. Since both of these drugs work as central nervous system stimulants, they will each amplify the effects of the other. Both are used, for example, to force wakefulness. If you were to take both at the same time, the length and intensity of a wakeful period would be significantly amplified. You produce a result greater than what you would reasonably expect from the use of either drug alone. We call this kind of interaction potentiation. It is an interaction effect.

You can also produce an effect that is distinct from the effects you would expect from the presence of each individual component. Take dextroamphetamine again. It is a central nervous system stimulant. It is used to enable focus, arouse wakefulness, etc. Take grapefruit juice. It is a nice refreshing citrus juice. If you were to wake up, take your dextroamphetamine and wash it down with grapefruit juice, the two will interact and produce an unexpected effect. Grapefruit juice is acidic, and so changes the level of acid in you stomach. This level of acid can prevent the absorption of dextroamphetamine. So if you were expecting to sit down and get some work done after your morning breakfast, in this case, the aid in focus and wakefulness effect of the dextroamphetamine will be significantly limited.[157] This is an interaction effect.

I have suggested above that we view the symptoms as individual elements of disorder that, when they meet and interact, produce an unexpected or exacerbated effect. I suggest that the magnitude of such an effect is greater than, or the presence of the effect is distinct from what would reasonably occur from the presence of each individual symptom alone. And I suggest that wherever such an effect occurs from the interaction of symptoms in a symptom set, we get the distress criterion.

[156] Princeton (2023).
[157] *Dextroamphetamine* (2023).

In our ADHD case with Remy and Aunt Marge, we can see immediately how an account of mental disorder as interaction effects can play out. It is possible, sure, that Aunt Marge has some of the symptom set for ADHD. And perhaps even Aunt Marge exhibits enough symptoms of the symptom set for ADHD that, if she had distress, she would have ADHD. But for Aunt Marge, her symptoms do not interact in a way that produces the effect of distress. For Remy, their symptoms do interact in such a way that it determinatively produces distress and therefore disorder.

So the distinction between Remy and Aunt Marge illustrates one way that an account of mental disorder as interaction effects succeeds. If we understand mental disorders as interaction effects, then we capture the aspect of mental disorder that is highlighted in diagnostic practice. An interaction effect account can explain how under some circumstances, in some individuals, symptoms will interact with one another such that they produce distress, and therefore disorder. And it will accommodate the difference between Remy and Aunt Marge, as well as explain why Aunt Marge's dinner table comment is a reasonable misunderstanding of mental disorder.

But now, why use interaction effects? We saw in §5 that an account of mental disorder as natural kinds failed ultimately because the theoretical machinery of natural kinds could not adequately accommodate all the nuanced aspects of mental disorder. So it is my view that the most deflationary account of mental disorder would be one that makes use of *only those parts of disorder to which we are already committed*. Using just these elements and their interaction is not only structurally simplified, but also commensurate with the understanding of mental disorders implicated the DSM. While this understanding might not be shared by every professional in the psychological or psychiatric fields, it is the working basis for their collective understanding of disorders. The received view of mental disorders as natural kinds seemed to result from theoretical analysis that *imported* philosophical kinds frameworks—along with all that respective baggage. It was importing these theoretical frameworks

that led us to our previous problems, some of which arguably result from a misapplication or misunderstanding of the kinds framework itself.[158] Better to use only what we have in our diagnostic manuals than to unnecessarily expand our ontological commitments for an account that still fails to capture mental disorders.

7] The failure of an interaction effects account:

We see that mental disorders as interaction effects holds some promise. Great! But we know from our examination of mental disorders as natural kinds that any successful account of mental disorder must strike a difficult balance. Successful accounts of mental disorder must be permissive enough to capture genuine instances of disorder and exclusive enough to rule out those cases that are not. Put alternatively, a successful account of mental disorder must not over-generalize *and* it must not over-specify. Accounts of mental disorder must tell us that Remy has one, while Aunt Marge does not. Accounts of mental disorder must tell us that Remy's ADHD is not the only way to have ADHD, but simply one way to have ADHD. So if an account fails to strike the balance, on either side, then it fails as an account of mental disorder. So how does mental disorder as interaction effects fair?

It is important, at this juncture, to recall that mental disorders include aspects from across various domains. Among the component parts of mental disorder, various explanations can be given for the presence of symptoms in a symptom set. Often there can be more than one explanation for a single symptom. Some of these explanations will be biological, but of course a solely biological explanation for any disorder will be insufficient. Symptoms can be determined by social mechanisms as well in case like, for example, the trauma response reactions in post-traumatic stress disorder (PTSD).[159] These social mechanisms of, or social contributions to, disorder, are part of a set of social

[158] Oakley, Chapter 2, does in fact make this argument, albeit in different terms.
[159] NIMH (2023).

facts. Social facts establish an individual's context and are an important element that we must also consider when we are looking at cases of potential disorder.

So for this, we will require some context. To start, we will be looking again at ADHD. Currently, ADHD is among the most over-diagnosed mental disorders.[160] This means that ADHD diagnoses outpace the expected prevalence rate for the disorder. It is also among those disorders that are used as examples of how pharmacological incentives contribute to overdiagnosis, where the primary stimulant medications used in pharmacological interventions for the disorder are wildly profitable.[161] Couple this with the fact that schools reporting higher levels of students that have ADHD accommodations can mitigate test score accountability requirements,[162] and you end up with a near perfect storm: an explosion of ADHD diagnoses in students of high school age and younger.

But there's also more localized factors that can impact a student's likelihood of meeting the requirements for ADHD diagnosis. Children of parents in the lowest socioeconomic percentiles are more likely than those in the highest percentiles to meet requirements for diagnosis with ADHD.[163] And there is evidence to suggest that the likelihood of meeting the requirements for ADHD diagnosis is also elevated for children with *parents* that are unemployed,[164] or received social assistance.[165] All of these factors about the *parents* of children constitute risk factors *for the children receiving diagnosis with ADHD*. And all these risk factors also reasonably constrict the places where parents live—and so establish their children's public school district. Furthermore, national averages indicate that public school students have higher rates of meeting the ADHD criteria (approximately 32%) than private

[160] Frances (2013).
[161] *Ibid.*
[162] Miller (2021).
[163] Froehlich *et al.* (2007).
[164] Davis *et al.* (2010).
[165] Anderson (2018).

school students (approximately 22%).[166] Public schools in poorer districts often have higher student to teacher ratios, and class size has had a documented impact on the likelihood of diagnosis with ADHD as well.[167]

Rural schools, especially those in states like West Virginia, meet all these conditions: the socioeconomic features of the average parent in this region is consistent with *all* these mentioned factors and the schools, struggling to receive adequate funding, often have larger class sizes. These things, taken together, lead to West Virginia students having among the highest prevalence rates of ADHD diagnosis in the United States at over 11%, significantly higher than the national average.[168]

Now if you are a young boy in West Virginia, or nearly any predominantly rural state, you have a *significantly* increased likelihood of meeting the requirements for diagnosis with ADHD. To start, young boys are more likely to be diagnosed with ADHD in the first place—1.4 times more likely than young girls.[169] At last count, the CDC estimated that young boys have a 20% rate for ADHD diagnosis nationally—that is one in five boys of high school age or younger diagnosed with ADHD.[170] This diagnostic rate has a lot to do with the socialization of young boys, as well as the social expectations of long days of schooling.[171]

So let's consider Timmy. Timmy is a young white boy in a rural, poorer public school district. Timmy is having trouble sitting still during his lessons. He is often out of his seat and distracting his peers by interrupting his teacher or talking during quiet working times. He is also doing poorly on standardized tests and often forgets his homework at home. Does Timmy have ADHD? It could be ADHD. Or it could be that Timmy is just a young boy doing the things that any person his age would

---

[166] Abbasi *et al.* (2023).
[167] Havey *et al.* (2005).
[168] Visser *et al.* (2014).
[169] Abbasi *et al.* (2023).
[170] Miller (2021).
[171] Froehlich *et al.* (2007).

do. Or the teacher might simply not have enough time or resources to give Timmy the structure he needs to be challenged and engaged in class.[172] Timmy might not be good at taking standardized tests—test taking is a skillset that not everyone has! Timmy might have a rocky home environment that precludes his ability to complete homework. But it seems that, given what we know, for every one in five cases like Timmy's, a diagnosis of ADHD is given. Even when there are numerous other ways to explain the presence of the symptoms in the symptom set that precludes the possibility of diagnosis.

Let's hold the same symptom set we have for Timmy fixed. If we were to take Timmy's case and drop it into a rich, private school community, with parents in the upper socio-economic percentiles, Timmy's chances of being similarly diagnosed drop. If Timmy were a young Black boy, his chances would be higher—in both the poorer and richer scenarios.[173] If Timmy were a young White girl, then the chances of ADHD diagnosis drop significantly. And if Timmy were a young Asian girl, the chances drop even further.[174] But in all of these cases, we are assuming that *the exact same symptom set* is met. The only descriptive features that change about the individual are either racial, gendered, sexed, or socioeconomic. So nothing about the symptom set is changing, and yet the likelihood for diagnosis does change. So it stands to reason that the factors that lead to diagnosis with disorder here are these "risk factors." If these risk factors are the difference makers for disorder, the claim is that they are difference makers for distress. You cannot be disordered without distress. So they would have to contribute to the establishment of distress in order to establish disorder.

But it is clearly *not* the case that these factors are difference makers in the case of disorder. Risk factors are not interchangeable with the distress criterion nor the symptom set; these diagnostic

---

[172] So-called "giftedness" in young students often appears with similar behavioral symptoms as those of ADHD. This further complicates the already troubled process of diagnosis. See especially Mejia (2016).

[173] Cénat *et al.* (2021).

[174] Wong & Landes (2021).

disparities are likely driven by *other factors*, like diagnostician bias or inadequate resources to support teachers, and so constitute likely misdiagnosis. This likelihood is further underscored when we account for the heritability component of ADHD and that ADHD has among its multifactorial etiology a neurobiological process.[175] So these are misdiagnoses that count non-disordered students as disordered. When we have too many cases being diagnosed without actually meeting the distress criterion, we have *overdiagnosis*.[176] And when we have overdiagnosis of mental disorder, we have an overgeneralization in the account of that mental disorder. So an account that uses interaction effects to capture disorder goes wrong by including too many extra-symptomatic features to determine disorder, and still fails to capture distress.

That is not to say, however, that these cases are *always* instances of misdiagnosis. Some of these cases, even that of Timmy's original case, are genuine cases of disorder. And for some of these cases, the risk factors are making a difference in whether there is disorder. There are good reasons to think that risk factors like poverty constitute genuine factors that can push the threshold of symptom sets to distress. Chief among these is, for example, the well-documented and well-studied relationship between depression and poverty.[177]

Part of the problem in discerning genuine cases of disorder under these circumstances with risk factors is due to the fact that it can be hard to discriminate between facts of social context and social contributions to disorder. Recall that at the very beginning of this project I explained that social factors cannot be cleanly disentangled from mental disorders and should not be. Sometimes social context plays the role of a social contributor to disorder. This is what happens in the mentioned cases

---

[175] Borgen *et al.* (2021).

[176] In the case of young girls, especially young Asian girls, we should note there are likely cases of *missed* diagnosis; the disparity between prevalence rates and diagnosis for these young girls constitutes a case of *underdiagnosis*.

[177] See Ridley *et al.* (2020), Heflin & Iceland (2009), and Kim *et al.* (2013) for a few examples highlighting this relationship.

of poverty and depression. A social fact plays the role of social contributor, and that establishes symptoms for disorder or pushes symptoms to distress creating levels. Post-traumatic stress disorder (PTSD) is among the mental disorders that most obviously includes facts about social context in the contributing role, and somewhat reliably in that role. This is because we include facts about a persons social context to establish the precipitating event or experience that establishes a trauma response. We can contrast this with depression and poverty, where poverty could be the precipitating social fact, or it could be another fact that contributes, or it could be entirely unrelated to the depression. In this last case, we see that sometimes social context does not play a contributing role to disorder. If we understand risk factors as social context, and I am suggesting we do, then it is clear that risk factors come apart from determinative factors for disorder. It is the fact that they sometimes do not come apart that complicates things in the interaction effects account. And since it can be exceedingly difficult to parse out when risk factors make the difference and when they do not, including them as the elements that tip the scales toward disorder is a mistake. And when we do, we conflate cases of disorder with cases that are not.

The problem with the interaction effects account of mental disorder is that it would count misdiagnosed cases and genuine cases alike. And a good account of mental disorder, we recall, is supposed to explain the difference. So I will just reiterate: an account of mental disorder as interaction effects does not give us the machinery to tell the difference between cases where risk factors come into play. All that an interaction effect account can tell us is, in cases like Timmy's, that the case meets the symptom set criteria, not that it is or isn't an instance of disorder.

An interaction effects account of mental disorder would, therefore, assert space between the diagnostic criteria and the disorder it aims to describe. Since we have already committed to the idea that the diagnostic criteria can only be considered literal and constitutive (§2), we must resolve this

disconnect. As far as I can reason, we can only resolve this disconnect in one of two ways. First, we can specify which risk factors under which conditions should be considered as "diagnostic factors" in determining caseness; basically, we can include more specific accounts of risk factors as a component of disorder. Second, we abandon the interaction effects account and reassert the descriptive accuracy of our diagnostic categories. I argue that the first option only raises more issues for an account of mental disorders as interaction effects. Let's turn to this argument now.

8] Problems with including risk factors as part of mental disorder:

We have now established that mental disorders as interaction effects fails to strike the balance by failing to prevent over-generalization. Mental disorders as interaction effects has, perhaps unwittingly, folded in what are otherwise considered mere risk factors as *the key elements* for establishing whether the symptom set interacts to produce and meet the distress criterion. And risk factors are *not* in the symptom set for a reason—they are simply those external factors that *might* contribute to an individual's expression of the symptom set. Risk factors do not guarantee that a symptom or symptom set might arise, they simply raise the probability that a symptom set *might*.

Consider Aunt Marge again. The family might have an established history of alcoholism. So Marge might have a genetic risk for developing alcoholism. If Marge is a heavy drinker, even with the added heritability risk factor of alcoholism, it would be wrong to diagnose Marge with alcoholism on the basis of her drinking and genetic risk alone. It would be wrong in the same way it would be wrong to say that all college students are alcoholics: the relationship that the drinker has to alcohol is the most important feature, not the risk factor.

So we see in our example of Marge that there is a reason that risk factors are not treated on a par with symptoms. But perhaps, one might press against this distinction. We saw earlier that symptom sets do not guarantee that the distress criterion will be met—the presence of the symptom set alone

only raises the possibility that distress could arise. One might suggest that it is hard to see how we can functionally differentiate between symptom sets and risk factors. Both merely raise the possibility of expressing the symptom set or meeting the distress criterion, respectively, after all. But symptom sets and risk factors are distinguished, at least, by their locus and their outcomes. Symptom sets interact to produce distress, and risk factors interact to produce symptom sets. Symptom sets are supposed to be located within the individual. Risk factors are, typically, located outside of the individual. Even still, one might push back: if risk factors *do* produce symptom sets, and those symptom sets *do* produce distress, it is hard to see how we can deny that risk factors could produce distress.

But we *do* deny that risk factors *determinatively* produce the symptoms or symptom set, even when they contribute to the probability that the symptoms will present. This is clear in our earlier cases with Aunt Marge and with Timmy. The presence of risk factors was not enough to differentiate between cases of disorder and cases of normal variance. And this is because risk factors cannot bring about the presentation of symptoms with any sufficient regularity required to establish a robust relationship between symptom and risk factor. We deny that the relationship between risk factors and symptom presentation are causal, and in many cases we deny that they are even more broadly determinative.[178] So we must deny that risk factors can produce distress, which on an interaction effects account must be established *determinatively* through the interaction of the symptom set. Since risk factors cannot be so determinative, they cannot be included in the establishment of distress.

Even if we were to include risk factors, what would that look like? We have seen that risk factors cannot be the difference maker in the case of alcoholism, but that they might reasonably make the difference in Timmy's ADHD, or in cases of depression. The first thing we would have to commit to is that risk factors *cannot* be symptoms. It should immediately be clear that something is wrong if

[178] Sciberras *et al.* (2017).

we suggest that we elevate risk factors, which include race, class, and gender, to the level of symptoms. But perhaps instead we say that we must include risk factors among the component elements of mental disorder. There are two problems with that. Let's look at each in turn.

8.1) Risk factors, over-general or inadequately specific: The first problem with including risk factors in the components of mental disorder is that risk factors are too broadly applicable. We saw in our previous cases that it was part of the risk factor being too broad that made it impossible for interaction effects to pick out all and only genuine instances of disorder. Thus, risk factors are too generalizable to function as a component of mental disorder. If we attempt to make risk factors more specified, we run the risk of making risk factors either unpalatably targeted toward specific groups—and run afoul of the above concerns—or so specified that they are practically infeasible. Let's look at the problems with specification.

How ever we specify these risk factors, these specifications must eventually be included in the diagnostic criteria. And surely there are near innumerable combinations of risk factor variables that would need to be listed in our criteria if we were to adequately specify them to avoid over-generalization.[179] Such a list would obscure the diagnostic criteria to the point where the diagnostic manual would not be practically useful. Since this is the diagnostic manuals primary aim, to aid in the standardization of diagnosis and treatment, this is not a palatable outcome.[180] But, practical concerns aside, a complete and fully specified list would only work if the risk factors so specified were partly

[179] My concern here is combinatorial. Let's say we can pinpoint the various risk factors that we must consider together to establish some adequately specified "risk vector" for a set of individuals. That combination must be reiterated until we exhaust the possible combinations of risk factors that an individual might meet. And then combinations must also include deliberate omissions of risk factors in some cases where, for example, mothers did not smoke during pregnancy, etc.

[180] If we were to specify risk factors in this way, we would be essentially enacting a more unwieldly correlate to a Boorsean reference class (1987, 1997). But Boorsean reference classes in his biostatistical theory of health have already met significant rebuke (Kingma, 2014). In general, for creating effective groupings with a high enough degree of specificity, it is not clear how many of the "variables" or the risk factor vectors must overlap to constitute a useful treatment category. If such groupings fail for the more obviously quantifiable components of physical health, the problems will compound in less obviously quantifiable modalities for risk factors in mental health.

determinative of disorder membership. And we have already seen that risk factors cannot be. So such a specified list fails both practically and conceptually.

8.2) Risk factor inclusion and consistency: The second problem with including risk factors as a component of mental disorder is that we are already committed to assumptions that prohibit this inclusion! We arrived at the interaction effects account on the assumption that there are two parts of mental disorder: the symptom set and the distress criterion. We arrived at this commitment to two parts because of our commitment to the constitutive interpretation of the DSM; our descriptive diagnostic criteria must be taken as literal and definitive. Our current DSM does not list risk factors in the disorder criteria for each disorder.[181] Now if we include a third part to help an interaction effects account explain when a case is a disorder, then we undermine our original assumption. Moreover, since we now know that the inclusion of risk factors fails practically and conceptually, there is no good reason to include a feature that is incompatible with our starting assumptions.

9] Mental disorders are not interaction effects:

So far we have seen that an account of mental disorders as interaction effects fails because it cannot distinguish cases of genuine disorder from cases that meet only the disorder's symptom set. Since distress is the disorder-maker (§4), any account that treats identically cases of disorder and cases that only meet the symptom set fails by over-generalizing. It appeared that mental disorder as interaction effects could capture the difference between Aunt Marge and Remy. But when I introduced risk factors, the interaction effects account could not distinguish between a case of disorder and a case of likely misdiagnosis, as illustrated in Timmy's case.

[181]The most current edition of the DSM, the DSM-V-Text Revision, includes risk factors in the descriptive text for disorders, not the disorder criteria (APA, 2023).

So the interaction effects account counts as equal cases where there is disorder and cases where only the symptom set is met. This conflation arose because, since we held the symptom set equal in Timmy's cases, the inclusion of risk factors highlighted that the interaction effects account introduced space between the diagnostic criteria and the disorder it aimed to literally describe. Since I am committed to the constitutive position, I cannot endorse any account of mental disorder that generates space between the diagnostic criteria and the disorder. I offered one suggestion to resolve the disconnect. I have now argued that we must not try to resolve it by incorporating risk factors into diagnostic criteria. Therefore, we are forced to abandon the account of mental disorders as interaction effects.

10] The constitutive position, further considerations for accounts of mental disorder:

I have argued that mental disorders cannot be interaction effects. I offered an account of mental disorders as interaction effects because it was the most deflationary account that I could use that also could support the necessary components of mental disorder. While the account seemed like it was promising, it took only a little investigation to discover that it could not explain the difference between cases of disorder and cases of misdiagnosis. Moreover, mental disorders as interaction effects could not distinguish between social facts relevant to diagnosis, and social context that did not play a contributive role. This failure illuminated the problem with such an account: without descriptive difference between two cases in which the symptom set is met, an interaction effects account looks to features outside of the component parts of disorder. In doing so, an interaction effects account abandons the constitutive interpretation of mental disorder in an attempt to preserve the account of mental disorder. This inconsistency is incompatible with the original assumptions I took up to critically examine the account, and so it must be abandoned.

The heart of the issue with any account of mental disorder, it seems, will always be that we are forced to take the constitutive interpretation of our diagnostic categories. I have here examined this account using the constitutive interpretation of the DSM. But the constitutive interpretation will apply broadly to any diagnostic categorization system. It remains the only way that we can uphold the commitment to the reality of disorder, that our diagnostic manuals track that reality, and that we can provide the uniformity in treatment that is necessary for patient care. And so it is my view that any attempt to provide an account of mental disorder, using any diagnostic categorization system, will be a project of putting out perpetual fires. These fires all originate in the requirement for the constitutive position, and its resulting resistant reification problem. While reification does not rear its head in our conceptual analyses, its source creates distinct problems for them.

I have said before that the constitutive position as our default position is the best we can do with what we have.[182] But I also acknowledge it is the likely root of the problem. Practically, this bears out as treating disorders as valid until proven otherwise. Socially, these commitments can bear out as harmful, discriminatory, and often unpalatable "diagnoses." Conceptually, now, it seems that the assumption of the constitutive position will likely undermine the success of any account of mental disorder. So, do we abandon it?

I do not think that we can. It is this view that is established in the DSM. It seems to be the minimum consequence of any view that is non-skeptical about the reality of mental disorders and does not want to reinvent the diagnostic wheel. However, that any account of the nature of mental disorders is potentially undermined by the assumption that they are real seems like an unavoidable, and uncomfortable tension. I am not suggesting that my conclusion here demonstrates this definitively. I am, however, suggesting that my conclusion here will likely extend across any account of mental

[182] Oakley, Chapter 1.

disorder. And I concede that I cannot see any clear, obvious, and unified way to resolve the potential tension if that is the case. The only way to avoid the problems raised by the constitutive position is to abandon it, and I have argued that is not tenable either. So, we best get used to the tension.

Conclusion

I have argued that there is only one way to interpret our current diagnostic manuals: constitutively. In taking up this position, I have argued that accounts that attempt to validate mental disorders using biological or social mechanisms will fail. Giving a purely biological account of mental disorder fails anyways, but even one that uses both social and biological mechanisms will still not have a principled method for selecting out all and only genuine mental disorders. Furthermore, I showed that the problem for validation, the problem of reification, persists in our best accounts precisely because we must take the constitutive interpretation of our diagnostic manuals. All of this took place in chapter one.

But the commitment to the constitutive position continued to have consequences in chapters two and three. In chapter two I argued that there is a special feature of mental disorders generally, the distress criterion, that prohibits mental disorders being captured by natural kinds. This distress criterion results from taking our diagnostic criteria constitutively, so this commitment to the constitutive position ultimately prohibits an account of mental disorders as natural kinds. Lastly, in chapter three I argued that even the most deflationary account of mental disorders is likewise undermined by our commitment to the constitutive position. The only way, it seems, to avoid the problems that arise for various accounts of the nature of mental disorder is to abandon the constitutive position.

But as I indicated at the end of chapter three, and as I concluded in chapter one, abandoning the constitutive position is not an acceptable option. To abandon the constitutive position, and avoid both the reificiation problem and the conceptual troubles it raises, is to sacrifice internal validity for our diagnostic categories. And if we abandon the internal validity of our diagnostic categories, they lose any meaningful semblance of being *diagnostic*. When our categories are appropriately diagnostic,

as they are in the constitutive position, we can reasonably collapse the distinction between the diagnostic category and the disorder. And that facilitates our engagement with mental disorders in a way that can illuminate features of disorder, like the relationship between the distress criterion and the symptom set, that can help us clean up our conceptual understanding of them.

Our diagnostic categories are meant to help us to identify and treat problems that arise in patients. The aim of standardized diagnostic outcomes across individuals and mental health professionals is, as we know by now, the first aim of the DSM. Moreover, the only way we can ensure internal validity is serving that ultimate purpose is to also uphold external validity—the idea that our diagnostic categories track something real in the world. And the only way we can uphold *that* commitment is to stick with the constitutive position.

So it seems that the well-meaning basis of our commitment is ultimately the foundation of our conceptual troubles in giving an account of mental disorder. Thus, there is no clean resolution to the problems at hand. On the one hand, with this commitment, we can have no obviously successful account of what mental disorders *are*. On the other hand, we will not have any obviously successful account of what mental disorders are *without* the commitment to the constitutive position. We are effectively, conceptually stuck.

But I hope that the troubles I suggest we will have, despite persisting, are not enough to dissuade further attempts to get clearer on the nature of mental disorder. As I said in chapter three, it is often the failed attempts to give accounts of something in our philosophical practice that are more illuminating than our successful ones. And I hope here to have pulled back the curtain on at least one central aspect of mental disorders that is making them so difficult to pin down.

So I will leave this project with this suggestion: we make do, we work with what we have, we make the best of what we've got. Here, for me, this means that future philosophers and mental health

professionals should analyze mental disorders in a particular way. Moving forward, future analysis should work with the commitments that we are given in our best existing diagnostic manuals, import as few theoretical frameworks from outside the discipline as possible, and apply the best tools of the philosophical trade to whatever we've got under those conditions. I don't think an account of mental disorder will ever be as perfectly unpacked as philosophers would like, but unwieldy concepts have never stopped us before. The fact that they will likely stay unwieldy shouldn't stop us either. Mental disorders are worthy entities for continued critical engagement, regardless of how messy they might remain.

Works Cited

Abbasi, L.N., Mazzawi, T., Abasi, L., et al. (2023). The Prevalence and Associated Factors of Attention Deficit Hyperactivity Disorder Among Primary School Children in Amman, Jordan. *Cureus, 15*(4), e37856. doi:10.7759/cureus.37856

Aftab, A., & Sharma, M. (2021). How not to think about biomarkers in psychiatry: Challenges and conceptual pitfalls. *Biomarkers in Neuropsychiatry, 4*. https://doi.org/10.1016/j.bionps.2021.100031

American Psychiatric Association (APA). (1952). *Diagnostic and Statistical Manual: Mental Disorders*. Washington, DC: American Psychiatric Association.

APA. (1968). *Diagnostic and Statistical Manual of Mental Disorders, Second Edition: DSM-II*. Washington, DC: American Psychiatric Association.

APA. (1980). *Diagnostic and Statistical Manual of Mental Disorders, Third Edition: DSM-III*. Washington, DC: American Psychiatric Association.

APA. (2000). D*iagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision: DSM-IV -TR*. Washington, DC: American Psychiatric Association.

APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition: DSM-5*. Washington, DC: American Psychiatric Association.

APA. (2017). *Diagnostic and Statistical Manual of Mental Disorders,Fifth Edition, Text Revision: DSM-5-TR*. Washington, DC: American Psychiatric Association.

APA. (2023). About the DSM-5-TR. *American Psychiatric Association.* https://www.psychiatry.org/psychiatrists/practice/dsm/about-dsm

Anderson, L.R. (2018). Adolescent mental health and behavioural problems, and intergenerational social mobility: A decomposition of health selection effects. *Social Science & Medicine, 197*, 153–160.

Barnes, E. (2023). *Health Problems*. Oxford: Oxford University Press.

Beebee, H. (2011). Necessary Connections and the Problem of Induction. *Noûs, 45*, 504–527.

Beebee, H., & Sabbarton-Leary, N. (2010). Are psychiatric kinds 'real'? *European Journal of Analytic Philosophy, 6*, 11–27.

Bird, A. (2007). *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Bird, A. (2012). Are Any Kinds Ontologically Fundamental?. In T. Tahko, T. (Ed.). *Contemporary Aristotelian Metaphysics*. Cambridge: Cambridge University Press.

Bjorklund, P. (2006). No man's land: Gender bias and social constructivism in the diagnosis of borderline personality disorder. *Issues in mental health nursing, 27*(1), 3-23.

Blacker, D., & Endicott, J. (2000). Psychometric properties: concepts of reliability and validity. In Rush, J.A.(Ed.). *Handbook of psychiatric measures*. Washington, D.C.: American Psychiatric Association, pp. 7-14.

Boksa, P. (2013). A way forward for research on biomarkers for psychiatric disorders. *Journal of Psychiatry & Neuroscience*, *38*(2), 75–77. http://doi.org/10.1503/jpn.130018

Bolton, D. (2013). Overdiagnosis problems in the DSM-IV and the new DSM-5: can they be resolved by the distress-impairment criterion?. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *58*(11), 612–617. https://doi.org/10.1177/070674371305801106

Boorse, C. (1987). Concepts of Health. In Van De Veer, D. & Regan, T. (Eds.). *Health Care Ethics: A Introduction*. Philadelphia: Temple University Press, pp. 359–393.

Boorse, C. (1997). A Rebuttal on Health. In Humber, J. M., & Almeder, R. F. (Eds.). *What Is Disease?* Totowa, NJ: Humana Press, pp. 1–134.

Borgen, N.T., Frønes, I., & Raaum, O. (2021). Impact of the School Environment on Medical Treatment of Attention Deficit Hyperactivity Disorder: A Population-Wide Register Data Study of School-Wide Positive Behavioral Interventions and Supports. *Child Development, 92*, 2089-2105. https://doi.org/10.1111/cdev.13574

Boyd, R. (1985). Observations, explanatory power, and simplicity: toward a non-Humean account. I Achinstein, P., & Haraway, O. (Eds.). *Observation, Experiment, and Hypothesis in Modern Physical Science*. Cambridge, MA: MIT Press, pp. 47–94.

Boyd, R. (1988). How to Be a Moral Realist. In Sayre-McCord, G. (ed). *Essays on Moral Realism*. Ithaca: Cornell University Press.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies, 61*, 127–148.

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

Boyd, R. (2010). Realism, natural kinds, and philosophical method. In Beebee, H., & Sabbarton-Leary N. (Eds.). *The Semantics and Metaphysics of Natural Kinds*. New York: Routledge, pp. 212–234.

Cénat, J.M., Blais-Rochette, C., Morse, C., et al. (2021). Prevalence and Risk Factors Associated With Attention-Deficit/Hyperactivity Disorder Among US Black Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry, 78*(1), 21–28. doi:10.1001/jamapsychiatry.2020.2788

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge University Press.

Chakravartty, A. (2011). Scientific Realism and Ontological Relativity. *The Monist*, *94*(2), 157–180.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.

Colston-Nepali, L., & Leigh, D. (2019). Ligers and Tigons and Grolars, Oh My! Hybridization, and How It Affects Biodiversity. *Frontiers for Young Minds, 7*, 113. 10.3389/frym.2019.00113.

Cowen, P.J. (2008). Serotonin and depression: pathophysiological mechanism or marketing myth? *Trends in Pharmacological Science,* 29, 433–436.

Cowen, P.J., & Browning, M. (2015). What has serotonin to do with depression?. *World Psychiatry*, 14(2), 158–160. http://doi.org/10.1002/wps.2022

Craver, C.F. (2007). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. New York : Oxford University Press,: Oxford University Press, Clarendon Press.

Craver, C.F. (2009). Mechanisms and natural kinds. *Philosophical Psychology,* 22, 575–594.

Culp, S. (1994). Defending robustness: the bacterial mesosome as a test case. In: Hull, D., Forbes, M. & Burian, R.M. (Eds.). *PSA 1994: Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association*, vol. 1. East Lansing, MI: Philosophy of Science Association, pp. 46–57.

Culp, S. (1995). Objectivity in experimental inquiry: breaking data-technique circles. *Philosophy of Science* 62, 430–450.

Davenport, M. (2016). Why bearcats smell like popcorn. *C&EN Global Enterprise*, *94* (17), 11. DOI: 10.1021/cen-09417-scicon004

Davis, E., Sawyer, M.G., Lo, S.K., Priest, N., & Wake, M. (2010). Socioeconomic risk factors for mental health problems in 4-5-year-old children: Australian population study. *Academic Pediatrics, 10*(1), 41–47.

De Bellis, M., & Pan, L. (2003). Biologic findings of post-traumatic stress disorder and child maltreatment. *Current Psychiatry Reports,* 5, 108-117. DOI: 10.1007/s11920-003-0027-z

Dextroamphetamine: Side Effects, Dosage, Uses, and More. (2023). Dextroamphetamine: Side Effects, Dosage, Uses, and More. https://www.healthline.com/health/drugs/dextroamphetamine-oral-tablet

Dupré, J.A. (1981). Natural kinds and biological taxa. *Philosophical Review,* 90(1), 66–90.

Dupré, J.A. (1993). *The Disorder of Things.* Cambridge, MA: Harvard University Press

Ellis, B. (2001). *Scientific Essentialism.* Cambridge: Cambridge University Press.

Frances, A. (2013). *Saving normal: An insider's revolt against out-of-control psychiatric diagnosis, DSM-5, Big Pharma, and the medicalization of ordinary life.* William Morrow & Co.

Franklin, A., & Howson, C. (1984). Why do scientists prefer to vary their experiments? *Studies in History and Philosophy of Science, 15*, 51–62.

Froehlich, T.E., Lanphear, B.P., Epstein, J.N., Barbaresi, W.J., Katusic, S.K., & Kahn, R.S. (2007). Prevalence, recognition, and treatment of attention-deficit/hyperactivity disorder in a national sample of US children. *Archives of Pediatrics and Adolescent Medicine, 161*(9), 857–864.

Garb, H.N. (2021). Race bias and gender bias in the diagnosis of psychological disorders. *Clinical psychology review*, *90*, 102087. https://doi.org/10.1016/j.cpr.2021.102087

García-Gutiérrez, M.S., Navarrete, F., Sala, F., Gasparyan, A., Austrich-Olivares, A., & Manzanares, J. (2020). Biomarkers in Psychiatry: Concept, Definition, Types and Relevance to the Clinical Reality. *Frontiers in psychiatry*, 11, 432. https://doi.org/10.3389/fpsyt.2020.00432

Goodwin, G.M. (2015). The overlap between anxiety, depression, and obsessive-compulsive disorder. *Dialogues in Clinical Neuroscience*, *17*(3), 249–260.

Hacking, I. (1995). The looping effects of human kinds. In Sperber, D., Premack, D., & Premack, A.J. (Eds.). *Causal Cognition: A Multidisciplinary Debate*. Oxford: Clarendon Press, pp. 351–383.

Hacking, I. (1999). Madness: biological or constructed? In Hacking, I. (Ed.), *The Social Construction of What?* Cambridge, MA: Harvard University Press, pp. 100–124.

Hacking, I. (2007). Kinds of people: moving targets. *Proceedings of the British Academy, 151*, 285–318.

Harrington, A. (2019). *Mind fixers: Psychiatry's troubled search for the biology of mental illness*. New York, NY: W. W. Norton & Company.

Haslam, N. (2002). Kinds of kinds: a conceptual taxonomy of psychiatric categories. *Philosophy, Psychiatry, and Psychology, 9*(3), 203-217.

Havey, J., Olson, J.M., McCormick, C., & Cates, G.L. (2005). Teachers' perceptions of the incidence and management of attention-deficit hyperactivity disorder. *Applied Neuropsychology, 12*(2), 120 -127. doi:10.1207/s15324826an1202_7

Hawley, K., & Bird, A. (2011). What are Natural Kinds? *Philosophical Perspectives*, *25*(1), 205–221.

Heflin, C.M., & Iceland, J. (2009). Poverty, Material Hardship and Depression. *Social science quarterly*, *90*(5), 1051–1071. https://doi.org/10.1111/j.1540-6237.2009.00645.x

Hempel, C.G. (1965). Fundamentals of taxonomy. In Hempel, C.G. (Ed.), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, pp. 137– 154.

Hutch (2023). *Research Areas*. Diagnostics and Biomarkers. June 25, 2023. https://www.fredhutch .org/en/research/research-areas.html

Hyman, S. (2010). The diagnosis of mental disorders: the problem of reification. *The Annual Review of Clinical Psychology, 6*, 155-179.

Kendler, K.S., Zachar, P., & Craver, C. (2010). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*, 1143–1150.

Kendler, K.S. (2012). Epistemic iteration as a historical model for psychiatric nosology: promises and limitations. In Kendler, K.S., & Parnas, J. (Eds.), *Philosophical issues in psychiatry II: nosology*. Oxford: Oxford University Press, pp. 305-322.

Kendler, K.S. (2014). Toward a limited realism for psychiatric nosology based on the coherence theory of truth. *Psychological Medicine, 45*, 1115-1118.

Kendler, K.S. (2017). DSM disorders and their criteria: how should they inter-relate? *Psychological Medicine, 47*, 2054-2060.

Kim, J., Richardson, V., Park, B., & Park, M. (2013). A multilevel perspective on gender differences in the relationship between poverty status and depression among older adults in the United States. *Journal of women & aging*, *25*(3), 207–226. https://doi.org/10.1080/08952841.2013.795751

Kingma, E. (2014). Naturalism about Health and Disease: Adding Nuance for Progress. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 39*(6), 590–608. https://doi-org.proxy01.its.virginia.edu/10.1093/jmp/jhu037

Kornblith, H. (1993). *Inductive inference and its natural ground*. Cambridge, MA: MIT Press.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Li, L., Wu, C., Gan, Y., Qu, X., & Lu, Z. (2016). Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. *BMC psychiatry*, *16*(1), 375. https://doi.org/10.1186/s12888-016-1075-3

Lichtblau, N., Schmidt, F.M., Schumann, R., Kirkby, K.C., & Himmerich, H. (2013). Cytokines as biomarkers in depressive disorder: Current standing and prospects. *International Review of Psychiatry*, 25(5), 592-603, DOI: 10.3109/09540261.2013.813442

Loftus, Y. (2022). Autism Statistics You Need To Know in 2022. *Autism Parenting Magazine*. https://www.autismparentingmagazine.com/autism-statistics/

van Loo, H.M., & Romeijn, J. (2015). Psychiatric comorbidity: fact or artifact*?*. *Theories in Medical Bioethics,* 36, 41–60.

Machamer, P., Darden, L., & Craver, C.F. (2000). Thinking about mechanisms. *Philosophy of Science,* 67(1), 1-25.

Matthen, M. (2009). Chicken, Eggs, and Speciation. *Noûs, 43*, 94–115.

Mayr, E. (1969). *Principles of Systematic Zoology*. New York: McGraw-Hill.

Mejia, A.R. (2016). Appalachian Teachers' referral for ADHD in Contrast to Giftedness in Selected Vignettes. *Theses, Dissertations and Capstones*. 1030. http://mds.marshall.edu/etd/1030

Mellor, D.H. (1977). Natural Kinds. *British Journal for the Philosophy of Science, 28*, 299–331.

Mill, J.S. (1884). *A System of Logic*. London: Longman.

Miller, C. (2021). Are Schools Driving ADHD Diagnoses? *Child Mind Institute*. https://childmind.org/article/schools-driving-adhd-diagnoses/

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

Murphy, D. (2017). "Can Psychiatry Refurnish the Mind?" *Philosophical Explorations*, *20*(2), 160–174.

Narrow, W.E., Kuhl, E.A., & Regier, D.A. (2009). DSM-V perspectives on disentangling disability from clinical significance. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *8*(2), 88–89. https://doi.org/10.1002/j.2051-5545.2009.tb00222.x

National Institute of Mental Health. (2020a). *Domain: Cognitive Systems*. Research Domain Criteria Initiative. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/cognitive-systems

National Institute of Mental Health. (2020b). *Potential Threat ("Anxiety"), RDoC Classification*. Research Domain Criterion Initiative. https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/potential-threat-anxiety

National Institute of Mental Health (NIMH). (2023). *Post-Traumatic Stress Disorder*. NIMH Information Research Center. https://www.nimh.nih.gov/health/topics/post-traumatic-stress-disorder

NCBI. (2023). Table 3.15: DSM-IV to DSM-5 Generalized Anxiety Disorder Comparison – Impact of the DSM-IV to DSM-5 Changes on the National Survey on Drug Use and Health – NCBI Bookshelf. *National Library of Medicine*, National Institute of health. https://www.ncbi.nlm.nih.gov/books/NBK519704/table/ch3.t15/

Nevin, R.L. (2009) Low validity of self-report in identifying recent mental health diagnosis among U.S. service members completing Pre-Deployment Health Assessment (PreDHA) and deployed to Afghanistan, 2007: a retrospective cohort study. *BMC Public Health, 9*, 376. https://doi.org/10.1186/1471-2458-9-376

NIH. (March 2023). Guillain-Barré Syndrome. *National Institute of Neurological Disorders and Stroke*, www.ninds.nih.gov/health-information/disorders/guillain-barre-syndrome.

NMSS. (2023a) What is MS? *National Multiple Sclerosis Society*. https://www.nationalmssociety.org/What-is-MS

NMSS. (2023b). MS Symptoms. National Multiple Sclerosis Society. https://www.nationalmssociety.org/Symptoms-Diagnosis/MS-Symptoms

Ortiz-Martínez, M., González-González, M., Martagón, A.J., Hlavinka, V., Willson, R.C., & Rito-Palomares, M. (2022) Recent Developments in Biomarkers for Diagnosis and Screening of Type 2 Diabetes Mellitus. *Curr Diab Rep,* 22, 95–115. https://doi.org/10.1007/s11892-022-01453-4

Phillips, M.R. (2009). Is distress a symptom of mental disorders, a marker of impairment, both or neither?. *World psychiatry : official journal of the World Psychiatric Association (WPA), 8*(2), 91–92.

Princeton Health Services. (2023). Caffeine. *University Health Services*. https://uhs.princeton.edu/health-resources/caffeine

Putnam, H. (1973). Meaning and Reference. *Journal of Philosophy*, *70*, 699–711.

Quine, W.V. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science (New York, N.Y.)*, *370*(6522), eaay0214. https://doi.org/10.1126/science.aay0214

Salmon, N. (1982). *Reference and Essence*. Oxford: Basil Blackwell.

Schwartz, P.H. (2007). Defining dysfunction: natural selection, design, and drawing a line. *Philosophy of Science, 74*(3), 364-385. http://dx.doi.org/10.1086/521970

Sciberras, E., Mulraney, M., Silva, D., & Coghill, D. (2017). Prenatal risk factors and the etiology of ADHD-review of existing evidence. *Curr Psychiatry Rep, 19,* 1. 10.1007/s11920-017-0753-2

Slater, M. (2015). Natural Kindness. *British Journal for the Philosophy of Science*, *66*(2), 375–411.

Spitzer, R.L., & Wakefield, J.C. (1999). DSM-IV diagnostic criterion for clinical significance: does it help solve the false positives problem?. *The American journal of psychiatry*, *156*(12), 1856–1864. https://doi.org/10.1176/ajp.156.12.1856

Staley, K.W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science, 71*, 467– 488.

Stewart, S.H., & Birch, C.D. (2019). "Psychology Works" Fact Sheet: Alcoholism. *Canadian Psychological Association-the National Voice for Psychology in Canada.* https://cpa.ca/psychology-works-fact-sheet-alcoholism/

Strawbridge, R., Young, A.H., & Cleare, A.J. (2017). Biomarkers for depression: recent insights, current challenges and future prospects. *Dovepress, 13*, 1245-1262

Stegenga, J. (2009). Robustness, discordance, and relevance. *Philosophy of Science,* 76, 650–661.

Szasz, T.S. (1960). The myth of mental illness. *American Psychologist, 15*(2), 113–118. https://doi.org/10.1037/h0046535

Szasz, T.S. (2009). *Antipsychiatry: Quackery squared*. Syracuse University Press.

Tabb, K. (2015). Psychiatric progress and the assumption of diagnostic discrimination. *Philosophy of Science, 82*(5), 1047-1058.

Tabb, K. (2017). Philosophy of psychiatry after diagnostic kinds. *Synthese*, https://doi.org/10.1007/s11229-017-1659-6

Tekin, Ş. (2016). Are Mental Disorders Natural Kinds?: A Plea for a New Approach to Intervention in Psychiatry. *Philosophy, Psychiatry, & Psychology*, *23*(2), 147–163. doi:10.1353/ppp.2016.0013

Tobin, E. (2017). "Mechanisms and Natural Kinds" in *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. Abingdon: Taylor and Francis.

Tsiga, E., Panagopoulou, E., Sevdalis, N., Montgomery, A., & Benos, A. (2013). The influence of tim pressure on adherence to guidelines in primary care: an experimental study. *BMJ Open 3,* e002700. doi: 10.1136/bmjopen-2013-002700

Tsou, J.Y. (2015). DSM-5 and psychiatry's second revolution: descriptive vs. theoretical approaches to psychiatric classification. In Demazeux, S., & Singy, P. (Eds.). *The DSM-5 in Perspective: Philosophical Reflections on the Psychiatric Babel*. Dordrecht: Springer, pp. 43–62.

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

Ussher, J.M. (2013). Diagnosing difficult women and pathologising femininity: Gender bias in psychiatric nosology. *Feminism & Psychology*, *23*(1), 63-69.

Visser, S.N., Danielson, M.L., Bitsko, R.H., Holbrook, J.R., Kogan, M.D., Ghandour, R.M., & ... Blumberg, S.J. (2014). Trends in the parent-report of health care provider diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 20032011. *Journal of The American Academy of Child & Adolescent Psychiatry, 53*(1), 34-46. doi:10.1016/j.jaac.2013.09.001

Wakefield, J.C. (1992). The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values. *American Psychologist*, *47*(3), 373–388. doi:10.1037/0003-066X.47.3.373

Wakefield, J.C. (1997). Diagnosing DSM-IV—Part I: DSM-IV and the Concept of Disorder. *Behaviour Research and Therapy*, *35*(7), 633–649. doi:10.1016/S0005-7967(97)00018-1

Wakefield, J.C. (2012). Editorial: DSM-5: Proposed Changes to Depressive Disorders. *Current Medical Research and Opinion*, *28*(3), 335–343. doi:10.1185/03007995.2011.653436

Whewell, W. (1858). *Novum Organon Renovatum*. London: John W. Parker.

Wilson, R.A. (1999). Realism, Essence, and Kind: Resuscitating Species Essentialism? in *Species: New Interdisciplinary Studies*, R. Wilson (ed.). Cambridge, Massachusetts: MIT Press.

Wilson, R.A. (2005). *Genes and the agents of life*. Cambridge: Cambridge University Press.

Wilson, H., Giordano, B., Turkheimer, F.E., Chaudhuri, K.R., & Politis, M. (2018). Serotonergic dysregulation is linked to sleep problems in Parkinson's disease. *NeuroImage: Clinical,* 18, 630 -637.

Wimsatt, W.C. (1981). Robustness, reliability, and overdetermination. In Brewer, M.B., & Collins, B.E. (Eds.). *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass, pp.124–163.

Wimsatt, W.C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Wong, A.W.W.A., & Landes, S.D. (2021). Expanding Understanding of Racial-Ethnic Differences in ADHD Prevalence Rates among Children to Include Asians and Alaskan Natives/American Indians. *Journal of Attention Disorders.* doi:10.1177/10870547211027932

Wylie, A. (2002). *Thinking about Things: Essays in the Philosophy of Archaeology*. Berkeley, CA: University of California Press.

Yasue, I., Matsunaga, S., Kishi, T., Fujita, K., & Iwata, N. (2016). Serotonin 2A Receptor Inverse Agonist as a Treatment for Parkinson's Disease Psychosis: A Systematic Review and Meta

-analysis of Serotonin 2A Receptor Negative Modulators. *Journal of Alzheimer's Disease, 50*(3), 733-740.

Young, J.J., Silber, T., Bruno, D., Galatzer-Levy, I.R., Pomara, N., & Marmar, C.R. (2016). Is there Progress? An Overview of Selecting Biomarker Candidates for Major Depressive Disorder. *Frontiers in Psychiatry*, *7*, 72. http://doi.org/10.3389/fpsyt.2016.00072

Zachar, P. (2000). *Psychological Concepts and Biological Psychiatry: A Philosophical Analysis.* Amsterdam: John Benjamins Publishing Company.

Zachar, P. (2008). Real kinds but no true taxonomy: an essay in psychiatric systematics. In Kendler, K.S., & Parnas, J. (Eds). *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology.* Baltimore, MD: John Hopkins Press, pp. 327–367.

Zachar, P. (2014). *A Metaphysics of Psychopathology.* Cambridge, MA: MIT Press.

Zachar, P., & Kendler, K.S. (2017). The philosophy of nosology. *The Annual Review of Clinical Psychology, 13,* 49-71.