Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

Demarcation, instantiation, and individual traits: Realist social ontology for mental disorders

Polaris Koi

To cite this article: Polaris Koi (2021): Demarcation, instantiation, and individual traits: Realist social ontology for mental disorders, Philosophical Psychology, DOI: 10.1080/09515089.2021.2016674

To link to this article: https://doi.org/10.1080/09515089.2021.2016674

PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2021.2016674

Routledge
Taylor & Francis Group

ORIGINAL PAPER

OPEN ACCESS Check for updates

Demarcation, instantiation, and individual traits: Realist social ontology for mental disorders

Polaris Koi

Philosophy, University of Turku, Turku, Finland

ABSTRACT

Realists about mental disorder have been hasty about dismissing social explanations of how mental disorder is constituted. However, many social ontologies are realist ontologies. In order to create a meaningful distinction between realism and social metaphysics about mental disorder, I propose that realism about mental disorder is best understood as Individual Trait Realism (ITR) about them. For ITR, mental disorders exist in virtue of traits. I defend the view that ITR is compatible with social metaphysics, arguing that, in asking whether constituents in the social sphere figure the metaphysics of psychopathology, we are asking questions on three different strata of explanation: the strata of demarcation, instantiation, and individual traits. Distinguishing between these strata allows for nuanced realism that need not reject the social constitution of mental disorder.

ARTICLE HISTORY
Received 10 October 2019
Accepted 7 December 2021

KEYWORDS
Mental disorder; psychiatry; realism; social ontology; social construction; demarcation; instantiation; individual traits

1. Introduction

There is a rich ontological debate concerning whether, and how, mental disorders are *real*. For the purposes of this paper, "mental disorder" is used as an umbrella term encompassing both mental disabilities, characterized by their relative persistence, and more transient forms of pathologized mental functioning. Here are two examples of mental disorder.

Example 1: ADHD. Pete often feels restless, including in settings where this infringes on his ability to complete various goals. He is impulsive and has considerable trouble concentrating on tasks at work. He often speaks out of turn, causing problems in social settings, including romantic relationships. Pete was diagnosed with Attention Deficit/Hyperactivity Disorder as a child and continues to have ADHD.

Example 2: MDD. Emma experiences sadness or hopelessness almost every day. She suffers from insomnia, and sometimes, she thinks about self-harm. Emma's friend has suggested she visits a mental health professional,

CONTACT Polaris Koi polaris.koi@utu.fi Philosophy, University of Turku, Turku, Finland

but she thinks it pointless because she does not believe that she could be helped. If Emma were assessed by a clinician, Emma would be diagnosed with Major Depressive Disorder (MDD).

Many philosophers feel that there are pressing reasons to argue for realism about mental disorders, such as defending the legitimacy of psychiatry as a science and improving psychiatric praxis.[1] By realism about mental disorders, I mean the stance that for at least some mental disorders, their description is not exhausted by reference to our beliefs and that they therefore are a legitimate target of study and intervention for psychiatry.

Mental disorders like Emma's MDD and Pete's ADHD are sometimes called social constructs: phenomena whose existence is in some sense dependent on the social sphere. Social constructionism is the view according to which various phenomena arise from the social sphere and could not exist independently of it.

Sometimes, social constructionism is taken as the *causal* claim that the referent of *X* is a product of the social sphere, either because social factors brought it about or because social factors shaped it to be a specific way. This can be contrasted with social constructionism in the *constitutive* sense, which claims that *X* denotes something that is (or necessarily involves) a social structure. Since the causal construction claim about mental disorder is not controversial, in this paper, the focus is on constitutive social construction, i.e. *social ontology*.[2]

Realism about mental disorder is often contrasted with social constructionism. Social constructionist stances are sometimes described as antirealist, albeit that is not always the case. Realists about mental disorder,[3] however, want to defend a conception of mental disorder that is not exhausted by social structures: a conception of mental illness in which it, like somatic illness, is at least partially constituted by states of affairs pertaining to the biological individual, such as anatomical structures, mechanisms, behavioral traits, and symptoms. Realists about mental disorder are dissatisfied with stances that characterize mental illness as human artifice, even if this artifice is a robust social structure inescapably embedded in our culture.

The purpose of this paper is to make two connected contributions: first, to show that the debate concerning the metaphysics of mental disorders is not concerned with one but with three distinct explanatory challenges: disorder demarcation, determining whether a given case instantiates a given disorder, and explaining the individual traits that constitute mental disorders and second, to suggest that in order to make sense of the debate, the pool of views that are usually labeled as realism about mental disorders are best understood as Individual Trait Realism (ITR) about them. For ITR about mental disorders, these disorders are metaphysically dependent on

individual traits. The metaphysical commitments uniting realists about mental disorder are, for this reading, primarily concerned with the traits that constitute disorder.

"Individual traits," here, is shorthand for *phenotypic traits subject to individual variance within a population*. Any phenotypic trait that is subject to variance within a population (such as height, hippocampal volume, handedness, extraversion, suicidal ideation, or being divorced) is an individual trait in this sense. For mental disorders, some individual traits are labeled as the symptoms of these disorders: for example, high impulsivity is a trait that is also a symptom of Attention Deficit/Hyperactivity Disorder.

In my analysis, proponents of ITR want some of the individual traits that Pete has, such as his restlessness, his impulsivity, or his anomalous cortical thickness, to play a constitutive role in his ADHD. They want to say ADHD is metaphysically dependent on at least some such traits or combinations of traits (although ADHD may not be exhausted by these traits). They furthermore want to say that if Pete had no such traits, Pete would not have ADHD; and were no one to have such traits, ADHD would not exist. As a result, for proponents of ITR, improving psychiatry requires looking at the traits that patients have, and patients have mental disorders on grounds of the individual traits that they have.

In what follows, I will first demonstrate that realism in the standard sense does not stand in a meaningful contrast to social metaphysics and introduce the distinction between the strata of demarcation, instantiation, and individual traits in section 2. In section 3, I will then develop ITR and explicate the concept of trait involved and justify locating realist commitments in traits rather than the other strata discussed. Section 4 is devoted to the metaphysics of traits, whereas in section 5, the metaphysics of demarcation and instantiation is discussed. The analysis presented in this paper demonstrates that even as realism about mental disorders, understood as ITR, forms a meaningful contrast to social metaphysics, they are robustly compatible.

2. Social construction and the realist challenge: Not one but three controversies

For realists about mental disorder, Emma's MDD and Pete's ADHD are something our theorizing and scientific efforts seek to properly capture, rather than something constructed by these pursuits. By realism about mental disorders, I mean the stance that for at least some mental disorders, their description is not exhausted by reference to our beliefs and that they therefore are a legitimate target of study and intervention for psychiatry. A standard approach to the notion of "real" would be something that is independent of our beliefs about it (see, e.g., Kendler et al., 2011, p. 1145).

In their quest to ground mental disorder on belief-independent facts, realists about mental disorders often find themselves pressed to reject social accounts of mental disorder, which are sometimes associated with attempts to delegitimize psychiatry and to instead endorse a biology-centric conception of mental disorder (see, e.g., Fellowes, 2019; Kendler et al., 2011; Schaffner, 2013). This move is motivated by the notion that belief-independent facts are more readily found within a biologically minded methodology than within an approach emphasizing the social sphere.

According to the above standard notion, "real" facts are belief-independent facts. However, it would be a mistake to confuse social metaphysics with belief-dependence. For example, in Haslanger's (2016) structuralist account, facts about social structures would still obtain even if we, as a culture, were sociologically inept and unaware of these facts. Haslangerian social structures are therefore "real" in this standard sense.[4] In other words, realism in the standard sense of holding that the phenomena at hand are belief-independent does not form a meaningful contrast to social constructionism because some theories of social construction fall under realism in that standard sense. Such theories are unsatisfactory for those who wish to defend mental disorders as something that is not exhausted by the social sphere, but that also necessarily involves culture-independent states of affairs – in the case of mental disorders, facts about human (neuro)biology. However, a meaningful contrast to social construction can be drawn by qualifying realism about mental disorder as locating its commitments in traits subject to individual variation – a task I will take on in section 3.

Mental disorders are often seen as both biological and social, as exemplified by the "biopsychosocial model." The idea that belief-independent states of affairs have a foundational role in the metaphysics of social phenomena is widespread in both scientific and lay discourse concerning mental disorders. It is common to talk of "the physiological basis," "the genetic correlates" or "the underlying neurobiology" of mental disorders, while granting that these disorders are largely socially caused or constituted. Schaffner and Tabb (2014) call such stances inclusionary social constructionism, by which they mean stances where the social construction of mental disorder is seen as compatible with granting physiological etiologies and constituents for mental disorder as well as with treating psychiatry as a legitimate science; they contrast these with what they term exclusionary social constructionist stances, which assert that mental disorders are mere social structures and better suited for topics of sociological than medical research. For inclusionary social constructionism, social constructs can be real *and* biological.

However, as the lack of contrast between realism in the standard sense and social constructionism demonstrates, what it is for a disorder to be *real* remains unclear. This is because multiple objects of inquiry are conflated in this question. In asking whether mental disorder is real, or

how biological and social phenomena figure in the ontology of mental disorder, we are asking – and sometimes conflating – a variety of questions, such as what is the ontology of diagnostic *categories*? What is the ontology of Pete's ADHD? How come Emma instantiates MDD? When does the clinical community consider a mental disorder real? When is it *right* in considering it real?

To make better sense of this pool of questions, I suggest we distinguish between three strata of metaphysical explanation, each with its own distinct explanandum:

> (1) The demarcation stratum, on which the ontology of mental disorders *qua categories* is analyzed.
>
> (2) The instantiation stratum, on which explanation is sought concerning on which grounds certain individuals, and not others, are picked out as members of the category at hand.
>
> (3) The individual trait stratum, on which we examine the ontology of the traits that ground diagnosis.

There are two reasons why philosophers ought to distinguish between these three strata. First, making this distinction enables a split view concerning the ontology of mental disorder. A philosopher may, e.g., wish to defend a social ontology of mental disorder in terms of instantiation and demarcation yet be a neuroreductionist concerning the stratum of individual traits. Indeed, as I will argue in section 3, realist commitments about mental disorder are best understood as attached to the individual trait stratum, as the claim that certain traits subject to individual variance are the necessary constituents of mental disorder.

Second, while it is possible to defend, e.g., a social ontology of mental disorder across all three explanatory levels, extending the same analysis from one stratum to other strata may yield implausibly heavy-handed results. For example, the pragmatist stance that mental disorder categories are human artifices, conceptual tools to be negotiated and refined based on their usefulness (Varelius, 2009; Zachar, 2002) is a stance concerning mental disorder on the demarcation stratum. However, it does not necessarily follow that individual traits are similarly constructed. And vice versa, it is often pointed out that the ontology of most or all individual traits crucially involves biological facts such as facts concerning the individual's genotype and neuroanatomy. Yet it does not follow that the ontology of the *demarcation* of the mental disorders associated with these traits should involve these traits; if it is, a distinct explanation is needed for why that is so. The metaphysics of mental disorder categories are often disanalogous with the metaphysics of the relevant traits.

I will now turn to the topic of individual traits, explicating what they are, why they are so central to realism about mental disorders, and outlining aspects of their metaphysics. Toward the end of this paper, I will return to the strata of demarcation and instantiation.

3. Realism about mental disorder is best understood as Individual Trait Realism about it

While realists about mental disorder hold a range of heterogenous views, there is a point of agreement, though often implicit, among realists about mental disorder. By explicating this point of agreement, realism can be qualified to form a meaningful contrast to exclusionary social constructionism. This implicit point of agreement is that *individual traits are necessary constituents of mental disorder*. That there is a strong relationship between traits and mental disorders has been vindicated in various, contrasting ways, e.g., in terms of essences (Kendell & Jablensky, 2003) and using Boyd's (1991) concept of homeostatic property clusters (Kendler et al., 2011; Beebee & Sabbarton-Leary, 2010; see also Fellowes, 2019).

Realists about mental disorder share the stance that real mental disorders necessarily have some individual traits as their metaphysical constituents. I term this position Individual Trait Realism (ITR for short). Realists about mental disorder are best understood as Individual Trait Realists about it. I formulate ITR as follows:

Individual Trait Realism: For X to be real is for X to exist in virtue of individual traits Y.

Thus, for ITR about mental disorders, mental disorders X exist in virtue of individual traits Y. The set of traits Y should not, however, be arbitrary; instead, whether a given trait or set of traits qualifies as a constituent of mental disorders, is subject to controversy among realists.

To explicate what this entails, I will expand on each aspect of this formulation, starting with the concept of individual trait that is doing the work here.[5]

In biology, "trait" denotes various phenotypic features of an organism, ranging from the physiological (such as bone density, hair color, or the mass of the cerebellum) to the behavioral and psychological (such as nesting behavior in birds or neuroticism in humans). Some phenotypic traits are persistent (e.g., Pete's forgetfulness, height, and eye color), whereas others are only displayed transiently (e.g., Emma's suicidal ideation, marriage, and infant wheezing). Both the genotype and the environment play a robust role in the production of traits. Some traits are shared features of a species. Other traits, however, are subject to individual variation. Traits that are subject to individual variation – individual traits for short – are what psychopathology is interested in since shared traits will not do as markers of illness.[6]

Individual traits relevant to psychopathology include symptoms (such as Pete's forgetfulness and Emma's suicidal ideation), personality traits (such as Pete's impulsivity), and physiological traits (such as atypical cortical thickness). In brief, the concept of trait is here used in its broad, biological sense.[7]

The trait concept in biology has itself been a subject of some confusion (see, e.g., Violle et al., 2007). Here, I adopt the standard definition of trait as any state or feature of the individual organism, whether morphological, physiological, or behavioral, that is measurable at the individual level. Measurability, here, denotes simply that there are some means of assessing the presence of a trait or in the case of quantitative traits, its state. For example, extraversion can be measured using psychometric questionnaires, and the height, weight, and bone density of an organism can be measured using various methods in biometrics.

Phenotypic traits are contrasted with the genotype. This contrast was originally twofold: in addition to a contrast between the cause and the effect, traits were contrasted to genes in that traits were observable in a way that genes, historically, were not. With the advent of improved genetic technologies, this latter contrast no longer holds true. That the genotype is not considered a trait is, rather, a matter of convention.

When ITR asserts that mental disorders X are real when they exist in virtue of individual traits Y, it simply means that they exist in virtue of certain traits of individual organisms that are subject to variance within a population. For example, for ITR, Major Depressive Disorder exists in virtue of a set of traits that may include such as traits depressed mood, suicidal ideation, dopamine dysregulation, and insomnia. For ITR, a mental disorder is real when there is in fact such a set of traits that it exists in virtue of and that are subject to individual variance rather than universal to the human species.

Even as this claim (unlike the standard realist claim) stands at a stark contrast to exclusionary social constructionism, it is a modest claim. ITR does not assert that X would exist *solely* in virtue of Y or be *reducible* to Y – those are stronger claims that are not involved in ITR, but rather, are compatible with it. To reiterate, for ITR, X may be metaphysically dependent on other states of affairs *in addition to* traits Y. Additionally, the contingent constituents of X may include a wide variety of states of affairs and traits in addition to the traits Y that it exists in virtue of. However, what ITR does assert is that traits Y are required for X. According to ITR, if, for example, MDD turned out not to exist in virtue of individual traits Y, it would not be a real mental disorder: if it existed in virtue of traits shared by humans, it would be an aspect of the human condition rather than a disorder. If it existed solely in virtue of social structures and institutions, it would be a cultural artifact rather than a disorder.[8]

Now that the formulation of ITR has been clarified, I turn to controversy, rather than consensus, among realists. Realists about mental disorder further agree that the set of traits Y must be qualified in some way, to foreclose ways to pick out sets of individual traits Y that would be a poor fit for a realist approach to psychopathology. Without such a qualifier, ITR would be overly inclusive. There is a need to restrict what *sorts* of traits qualify as constituents for mental disorders for ITR: otherwise, the group of traits Y may be selected via means that are either arbitrary or otherwise unfit for the task given that mental disorders are intended to be viable targets for scientific inquiry.

One such unfit approach would be institutionalism, for which the traits relevant for mental disorder would be picked out as constituents of X by institutional agreement: for example, that trouble staying on task is a constituent of ADHD solely because that was agreed by majority vote at a meeting of the Psychiatric Association. Most realists about mental disorder would be unhappy with institutionalism, as they hold that mental disorders are not, or at least are not solely, a product of institutional exercises of power but rather something that institutionalized practices, such as medical science, seek to understand. If psychiatrists agree that trouble staying on task is a constituent of ADHD, realists hope that this is because they have *discovered* that this is so, not solely because they have exercised their institutional power to decree that this is so.

There are multiple competing realist approaches to restricting what sorts of traits or sets of traits qualify for inclusion in the set of traits Y. Here, I remain agnostic about what exactly is the best qualifier for selecting traits Y. Instead, for the remainder of this section, I will describe and discuss some prominent candidate qualifiers.

One candidate qualifier is the one put forth within neuroreductionist stances about mental disorder. For it, *a fitting individual trait is a trait reducible to neurobiology* (in whatever sense of "reducible" the reductionist endorses). But ITR need not be reductionist. The mechanistic property cluster model (Kendler et al., 2011) and the network model (Borsboom et al., 2019) are examples of nonreductionist ITR stances about mental disorder, for each of which the relevant qualifier is that there are the right sorts of causal connections among the candidate traits. For example, Borsboom et al. (2019) profess that "mental disorders are not brain disorders at all" (ibid.: 2). They found this assertion on a network approach to mental disorder where mental disorders are caused by and consist of causally related symptoms. These symptoms may be generated by neurobiological dysfunction, but they may also arise from external states of affairs. For the network model, what triggered the symptom in the first place is less important than the causal links among symptoms that cause the disorder to continue to exist.

The network approach draws on the work of Kendler et al. (2011), who offer a rejection of both essentialism[9] and social constructionism about mental disorder, as well as a nuanced argument for accounting for it in terms of mechanisms. Kendler, Zachar, and Craver reject social construction on grounds that while it "offers a revealing view of how psychiatric symptoms are related to social context," it also "neglects insight into the underlying genetic, physiological and psychological factors that are often shared among particular cases" (ibid.: 1146). (These charges are valid concerning exclusionary social constructionism, but it is a mistake to apply them to social constructionism *tout court*.) They then proceed to apply a concept developed for describing such complex mechanisms, that of *mechanistic property cluster* (MPC), to analyze mental disorder. On the MPC account, developmental, genetic, and other causal mechanisms together produce, not a single set of traits, but a loose and imperfectly shared cluster of traits. Kendler, Zachar, and Craver use MDD as an example, where the symptoms of suicidal ideation, feelings of guilt and depressed mood are causally interrelated (2011: 1147). For the MPC account of mental disorder, mental disorders X exist in virtue of individual traits that cluster together in a self-sustaining causal process. This causal clustering is what qualifies these traits as ITR constituents for mental disorders.

Causal clustering and reducibility are two examples of qualifiers that realists can propose for traits Y in order to rule out arbitrary and institutional selection of traits Y. For both qualifiers, certain features of the traits, such as the reliable correlation between a neural phenomenon and an overt trait or the reliable causal clustering of a set of traits, are the features that make these traits fitting constituents for mental disorder.

While realisms about mental disorder are varied, they come together in vindicating the necessary role of individual traits in the metaphysics of psychopathology. To meaningfully distinguish these realisms from the application of realist, exclusionary social ontologies to mental disorder, accounts ordinarily described as realist accounts of mental disorder (such as reductionist, essentialist, MPC, and network accounts) are best understood as ITR accounts of mental disorder.

4. The metaphysics of traits

In section 2, I claimed that the debate concerning the metaphysics of mental disorder ought to be understood as three distinct debates: one about disorder demarcation, one about instantiation, and one concerning the metaphysics of the individual traits that ground disorder. Above, I have clarified the concept of individual trait and set out the conception of individual traits that I believe unites realism about mental disorder. I next discuss the metaphysics of individual traits as constituents of mental disorder. If the

conception of individual trait is that which is outlined above, what metaphysical commitments follow? Must individual traits be in some sense "biological?" Can they instead be socially constituted? Is there space for both?

While the metaphysics of individual traits seems like a fairly theoretical topic of discussion, it remains central for debates concerning mental disorder, its demarcation, and proper care. For example, within the public debates surrounding ADHD and MDD, typical charges are that some traits, such as hyperactivity or feelings of hopelessness, really only denote ordinary behavior under certain circumstances: boys without sufficient support at home may act out in school, and are then described as "hyperactive," whereas young adults, facing global warming and employment difficulties, experience hopelessness that is warranted but are mistakenly told that instead of a societal crisis, they are facing a personal one. The friction in these examples stems from disagreement regarding whether characteristics such as Pete's hyperactivity and Emma's hopelessness really qualify as individual traits or whether they instead collapse into shared traits (the "human condition"), into social structures and norms, or into a combination of the two.[10]

It is harder to give a sweeping account about the ontology of individual traits than it is of the ontology of mental disorders. This is because individual trait is an extremely broad concept. Fashioning a detailed unitary description of the ontology of traits therefore appears a futile exercise: it should be obvious that there are meaningful differences between the metaphysics of traits such as cortical volume and traits such as Emma's sadness or Pete's hyperactivity. However, some broad strokes can be painted, and some distinctions made.

For ITR, mental disorders qualify as real when they exist in virtue of individual traits Y. Proponents of ITR must hold that these individual traits are real because for X to exist in virtue of something else, that something must also be real. Likewise, for ITR accounts that propose mechanistic or interactive relationships among the traits, such as network and MPC models, the individual traits must be real in order to act as causes (Eronen, 2019).

What, then, is meant by the assertion that an individual trait is real? I am using biological language to describe individual traits and ITR, and this may lead one to wonder whether I am endorsing the idea that individual traits must be "biological." The answer is "yes" only in a trivial sense. As humans are organisms, all human traits, from being a redhead to being a bank teller, are biological in the trivial sense of the word. But insofar as "biological" is used to denote something narrower, such as "physiological," traits need not be ."biological." The concept of trait includes traits such as being a bank teller, a widower, or a juvenile delinquent: all of these are features of humans subject to individual variance, and noncontroversially belong in the social

sphere. However, a proponent of ITR may opt to use "biological" qualifiers in restricting what sorts of sets of traits are suitable as constituents of mental disorder: one may, for example, choose to only include traits that are measurable using biometric means.

While such a physiology-centric view would qualify as a form of ITR, it is not my intention to endorse it. Rather, my intent is to demonstrate that mindfulness of human biology need not foreclose social ontology. There are multiple ways in which social factors may enter into mental disorders. Many of these pertain to instantiation and demarcation and are not terribly controversial. But perhaps surprisingly, social constituents may also factor into the very individual traits that constitute mental disorder.

First, there are various social structures, such as structures of stigmatization, that aggravate the burden of persons with mental disorders. For some mental disorders, aggravated distress will result in an aggravated disorder, as various forms of distress are characteristic of most mental disorders. For those mental disorders, the distress may be partly caused *and* constituted by a response to the stigmatization experienced by the individuals at hand; in such cases, distress-related traits are partly or wholly social. For Emma, such traits are feelings of worthlessness and self-blame: in addition to her other feelings of worthlessness and self-blame, she may have further such feelings that are a response to her stigma. For Pete, his restlessness may be such a trait: part of Pete's restlessness may be a response to anxiety about meeting social expectations.

Social structures have a constitutive role in some or all individual traits *beyond* the distress experienced by persons with the disorders. For example, Pete's tendency to speak out of turn is a phenomenon located in the social sphere: speaking out of turn is a social phenomenon, as it is the social conventions that determine whose turn it is to speak.

Above, I have provided evidence that some individual traits that are constituents for mental disorders under the current classification are social traits. However, a turn toward individual traits can also enable an examination of the nature of these traits with less emphasis on current diagnostic categories, which some hope would enable the generation of novel diagnostic categories that line up with neurobiological traits rather than social traits. The Research Domain Criteria (RDoC) movement is one example of this trend (Insel & Cuthbert, 2015; Tabb, 2019). Unhappy with our present demarcation of mental disorder, proponents of RDoC wish to assert that real disorders exist in virtue of the neurobiological individual traits they consist in, even if our present ideas concerning these disorders are imperfect. The hope of the RDoC movement is that purely physiological and causal descriptions of mental disorders would yield their redefinition as brain circuit disorders (Insel & Cuthbert, 2015). Such strict reductionism is one form of ITR: it is conceptually consistent to narrow down the group of

traits Y to physiological ones. However, as we will see, it is not viable as a full description of mental disorder. Even as it is a conceptually coherent stance on the stratum of individual traits (albeit one that will require empirical substantiation), it is unsuitable for the strata of instantiation or demarcation.

5. Demarcation and instantiation

It is not unusual to expect disorder demarcation or instantiation to be equivalent to, or analogous with, the stratum of individual traits in terms of their metaphysics. Here, I will first describe the strata of demarcation and instantiation more fully and then proceed to discuss some instances of how the strata are muddled in the literature, suggesting that distinguishing between these three strata alters the implications of the cases discussed.

Most philosophical work on mental disorder has been focused on the demarcation stratum. Ontological questions on the demarcation stratum revolve around the ontology of the *category* of mental disorder generally or around specific disorder categories such as Major Depressive Disorder. We may, for example, ask whether these are natural kind categories, i.e. whether these categories are based on a grouping that is independent of language and human interaction, or whether these categories emerge solely from such interaction.

That the ontology of demarcation is social is trivially true because language is social, and demarcation is something we do with language. The debate about demarcation does not, however, concern this trivial sense of social ontology, but rather, the more philosophically interesting question of whether our categories necessarily are shaped by social factors, such as the cultural history of these categories or the practical work we aim to do with them.

The demarcation of some phenomena seems to be guided not only by sociohistorical contingency but also by the belief-independent nature of the universe that any language or culture would strive to capture. One might assume, for example, that any society possessing an advanced scientific enterprise would sooner or later have concepts for protons, electrons, and neutrons. If they would define the atom differently enough, we might respond that they have not observed it correctly: there are certain stable facts about atoms that science can perceive and that are consistent across time, space, and culture. The thrust of the demarcation debate has been that our present demarcation of mental disorder appears not to be directed by correspondingly stable facts. While some, such as Zachar (2014), are ready to embrace this state of affairs, others (such as Insel & Cuthbert, 2015) have called for a radical revision of our psychiatric categories.

Assuming that the correspondingly stable facts driving the demarcation of mental disorder are individual traits, a proponent of ITR could argue that disorder categories exist in virtue of certain individual traits, arising from our need to describe what we observe in the world. However, while demarcation should evolve as our understanding of mental disorders evolves, excising the social from the demarcation of mental disorders appears as an overly ambitious aim. Here, I make use of Kendler's (2016) discussion on disorder demarcation. In their efforts to biologize psychiatric categories, proponents of reductionism are prone to overlook that categories denoting complex biological phenomena do not, and *cannot*, neatly match culture-independent states of affairs. Kendler's charge is that the project of revising psychiatric categories to be independent of the social sphere is impossible because mental disorders, like many if not all biological categories, characterize heterogeneous sets with fuzzy boundaries.[11]

This does not entail that the broad concept of mental disorder or the more specific concept of MDD would be invalid or that it would not pick out any individual traits. Rather, the upshot is that due to its fuzzy boundaries and internal heterogeneity, the scope of the concept will necessarily be subject to negotiation and renegotiation. Therefore, the social sphere necessarily enters into the demarcation of mental disorders. Disorder categories are nontrivial social constructs and will be such even if a RDoC-style revision is carried out.

So much about demarcation; what, then, about instantiation? Ontological questions on the instantiation stratum concern how individuals are identified as members of categories – for mental disorder, how they are identified as having a disorder. Instantiation seems straightforward: with the demarcation criteria in mind, we look at a given population and pick out people who satisfy the demarcation criteria. Pete is picked out as having ADHD because of individual differences and traits – e.g. that Pete is impulsive – that figure as diagnostic criteria for the diagnostic category of ADHD, and this constellation of traits then is an instance of ADHD. However, as I will argue below, the instantiation of mental disorders additionally necessarily includes the social sphere.

Even if a purely physiological description of a given mental disorder could be fashioned, such a description would not be a full or apt description of the individual traits that ground the disorder. Specific neurobiological traits, in the relevant social contexts, can produce the kinds of behavioral traits that constitute ADHD, but it is these behaviors – characterized by their lack of conformity to specific social expectations about individual behavior – that form the grounds on which ADHD, as we typically understand it, is instantiated.

Consider a scenario where the neural underpinnings of ADHD have become magnificently clear, and the disorder can now be demarcated with reference to neuroanatomical traits alone: a veritable (and plausibly

altogether unfeasible) feat for science. Now, consider a person who has the signature neurobiology but has never displayed any of the behavioral features of ADHD. It would be counterintuitive to say that such a person has ADHD, or if one were to insist that these neural traits suffice for ADHD, the term ADHD would then denote two distinct concepts – one neurobiological and one behavioral – which would not always line up. It is hard to predict which concept we would give priority to in such a case; however, describing neurobiological variance in the absence of pathologized functioning as a *disorder* seems false. As a result, behavioral traits appear to have explanatory primacy for ADHD even if neural traits are de facto constituents thereof.

For behavioral traits, the social sphere plays a crucial role in assessing whether these traits constitute features of a pathology. Consider the description of Major Depressive Disorder in DSM-V, which states that "responses to a significant loss (e.g., bereavement, financial ruin, losses from a natural disaster, a serious medical illness or disability)" (DSM-V: 161) can elicit an affective and behavioral response resembling Major Depressive Disorder. DSM-V further advises that whether the individual is depressed in addition to grieving should be considered by the clinician "based on the individual's history and the cultural norms for the expression of distress in the context of loss" (ibid.). For example, whether Emma has MDD depends in part on her social circumstances. Were Emma to have the same symptoms of sadness, hopelessness, insomnia, and suicidal ideation in response to losing her home and family in a military conflict, those individual traits would stand in a very different relationship to the social environment, since these behavioral traits would likely be seen as nonpathological in such a harrowing situation. Whether Emma instantiates MDD therefore depends not just on disorder demarcation and on the individual traits at hand, but also on Emma's immediate social context.

Davies (2016) has argued for the same effect by claiming that whether behavioral dispositions become manifest as behavioral traits also hinge on the social sphere. To this effect, Davies presents a thought experiment where an individual, possessing every disposition to act in a manner constitutive of Oppositional Defiant Disorder (ODD), is placed in an egalitarian utopia where there are no opportunities for authority-defying behavior. Davies concludes that since the individual will not display these behavioral traits in such a utopia, it is not possible for the individual to have ODD in such a social environment and that this is evidence that mental disorder involves a "relation to certain kinds, events, individuals, practices or institutions" (Davies, 2016, p. 291) in the individual's social and natural environment.[12] While Davies presents the argument as one about mental disorder generally, it is, more specifically, an argument about disorder instantiation.

The assertion that the instantiation of mental disorder involves the social context is often countered with the objection that some mental disorders, such as schizophrenia, present similarly across cultures (see, e.g., Kendler et al., 2011). These observations, however, are compatible. The empirical claim that schizophrenia presents similarly across cultures does not mean that all cultures would have the same concept of schizophrenia: it is not a claim about demarcation. It can be consistently read as a claim about instantiation: we could postulate that even if the concept of schizophrenia would not be available to a specific individual nor to her surrounding culture, this counts as an instance of *our* concept of schizophrenia. Yet its most intuitive application is as a claim about individual traits. A neuroreductionist would articulate this claim as follows: the neural underpinnings of schizophrenia are the same regardless of culture. A proponent of the MPC or network model could instead say that specific individual traits – the ones we associate with the disorder – cluster together across cultures. That this cross-culturality would entail that *the same disorder* is instantiated across cultures is an attractive claim and a useful hypothesis. However, this hypothesis is not confirmed simply by the observation regarding the cross-cultural clustering of individual traits unless one makes the further claim that such clustering is *sufficient* for something to *instantiate* mental disorder X. That claim, while consistent, seems unnecessarily heavy-handed. A more nuanced picture of disorder instantiation describes it as metaphysically dependent not just on individual traits, but also on the social sphere.[13]

The idea that the instantiation of mental disorders would be dependent on social contexts is resisted largely because it is taken to entail that instantiation would be arbitrary and to lead to a cultural relativist view on mental disorders. However, this worry is unwarranted. First, one can hold instantiation to occur in virtue of both social and nonsocial constituents. Second, even the claim that the instantiation of some disorder is *exclusively* social in ontology does not entail that instantiation would be arbitrary because culture is not arbitrary: it, too, is subject to a variety of causal structures.

With the three strata of demarcation, instantiation, and individual traits having been clarified, I wish to show that this clarification has not been in vain as these strata are sometimes confused, conflated, or expected to be analogous. Here, I raise two examples of such conflation, one constructionist and one reductionist, from within the debate concerning the metaphysics of ADHD.

Vehmas and Mäkelä (2009), in their discussion of the metaphysics of disabilities, assert that while disorders like Down syndrome/trisomy 21 involve both language-independent and social facts, ones like ADHD where no unitary biological cause has been found "are wholly language dependent because the corresponding facts are language dependent" (2009: 50) – the corresponding facts, for Vehmas & Mäkelä, being the referents of the diagnostic criteria. In claiming so, they confound the

individual trait stratum with the demarcation stratum: demarcation criteria and the relevant individual traits underlying mental disorders may not always match. A range of traits, including physiological ones, can underlie syndromes without there being a need to include those traits in demarcation criteria. For example, individual traits relevant to Pete's ADHD may include various traits pertaining to his structural and functional neuroanatomy such as anomalies in cortical thickness and in the default mode, frontoparietal and ventral attention networks (Park et al., 2018), traits pertaining to executive functioning, such as impairments in functions like behavioral inhibition, working memory, and cognitive flexibility (Barkley, 1997), and behavioral traits ranging from motor hyperactivity to speaking out of turn (DSM-V).

However, defining a disorder for diagnostic purposes is not simply about making an inventory of all the relevant individual traits. Many of the individual traits relevant to ADHD are not specific to ADHD. For example, atypical functioning of the default, frontoparietal, and ventral attention networks is not only relevant for ADHD but also for schizophrenia (Park et al., 2018). Second, many of the relevant individual traits are not consistent among people who are affected by the disorder: if the disorder is subject to multiple realizability, the group of relevant traits may also be subject to some individual variance. Third, the heterogeneity and richness of many mental disorder categories merit some reduction when it comes to demarcation criteria. An *exhaustive* description of ADHD would be a great scientific accomplishment, but it would be too elaborate to be clinically utilizable. Demarcation may not be a *mere* tool (depending on the degree to which one prioritizes pragmatist considerations in demarcation), but it is subject to the practical demands of clinical applicability.

A similar expectation of analogy can be found on the reductionist side of the debate. For example, Hoogman et al. interpret the results of their "mega-analysis" of brain physiology in people with ADHD as demonstrating that "patients with ADHD have altered brains; therefore, ADHD is a disorder of the brain" (Hoogman et al. 2017: 311). Yet while they found structural individual traits about ADHD, it does not follow from these findings that ADHD would not have a social ontology on any of the explanatory strata discussed above. Its demarcation and instantiation can depend metaphysically on the social sphere even if it also, whether necessarily or contingently, involves "altered brains."

Is there a way out of social ontology for the reductionist? Yes, *on the stratum of individual traits*: the reductionist may eschew behavioral traits in favor of traits on lower levels of biological organization than behavioral traits are and ground mental disorder on such traits alone. These include, e.g., structural traits, such as cortical volume, neural networks, and ultimately cells, chemicals, and molecules. Biological

constituents can be found for all mental disorders because all mental disorders are possessed by organisms; the trouble with reductionism is that it is very hard to find low-level constituents that would be *specific* to the mental disorders at hand. In the face of empirical research, there appear not to be dedicated physiological phenomena underlying specific mental disorders, but instead, *many* general processes, structures and mechanisms contribute to their causation and constitution (Kendler, 2005, 2008; see also Koi, 2021).

As a result, there is considerable difficulty in demarcating mental disorders directly by reference to low-level traits or in inferring their instantiation from the same. It may be epistemically necessary to involve traits on the behavioral level. Nonetheless, whether something is a necessary constituent for mental disorder does not depend on whether it is a *dedicated* constituent for it. For the neuroreductionist, the lack of dedicated low-level biological traits highlights the importance of critically examining whether these need to be complemented with social constituents on the demarcation and instantiation strata, even if one holds that they suffice as the constituents of mental disorder. In other words, the reductionist may need to explicitly include the social sphere to do the work of demarcation and instantiation, since that work cannot be done by reference to low-level biological processes alone.

6. Conclusion

The worry inherent in much of psychiatric reductionism is that were psychiatry dealing with social constructs, psychiatric medicine could not improve. But rather than denying the constitutive role of the social sphere in mental disorder, medicine would do well to seek to better understand it.

Above, I have established that realism about mental disorders is best understood as ITR about them and made the further claim that ITR is compatible with social ontology. This is not a particularly radical claim: many proponents of ITR endorse a conception of mental disorder where social norms and structures, institutional practices, history, etcetera play *some* sort of a part in the ontology of mental disorder. However, neither is this claim noncontroversial. Neuroreductionism about mental disorder still figures prominently in professional and lay debates concerning mental disorder. Not just neuroreductionists but also some proponents of complex models of mental disorder such as Kendler et al. (2011) reject social metaphysics in order to make space for their approach. However, as I have demonstrated, rejecting the social metaphysics of mental disorder is an unnecessary move for defending ITR accounts.

A plausible objection is that "biological" constituents may nevertheless be in some sense more important for mental disorder than social constituents are. Yet this criticism is harmless, because the present argument makes no claims regarding the relative *weight* of each constituent, whether "biological" or social.

I have argued that to meaningfully distinguish realism about mental disorders from exclusionary social ontology, it ought to be conceived as ITR about them. However, while ITR enables a meaningful contrast to exclusionary social ontology, it is compatible with inclusionary social ontology. Realists pursuing an accurate metaphysical description of mental disorder ought to carefully consider and account for its social constituents on any or each of the three strata of demarcation, instantiation, and individual traits.

Notes

1. See, e.g., Kendler's (2016) eloquent protest to nonrealist views in the following quote: "Over history, many cultures have done a poor job of properly seeing the other in those who are psychiatrically ill. It has been too easy to deny their humanity, to say they are not really sick. I continue to feel an obligation to counter this position and argue for the reality of mental illness." (Kendler, 2016, p. 6.)
2. The distinction between causal and constitutive social construction was made by Haslanger (1995). For a causal social constructionist account of mental disorder, see Church (2004).
3. While this paper focuses on explicating points of agreement among realists and assessing their compatibility with social metaphysics, realists about mental disorder are not a unified camp. Topics of debate among realists include, e.g., whether a correspondence theory or a coherence theory of truth is best suited for psychopathology (Eronen, 2019; Kendler, 2016), and to what extent mental disorders are reducible to brain disorders – some (e.g., Insel & Cuthbert, 2015) urge hard reductionism about psychopathology, while others (Borsboom et al., 2019; Davies, 2016) are externalists about it.
4. This point is also made in Barnes (2017).
5. In the following, I will build on a biological conception of "trait." However, this is not to be taken as an endorsement of the misconception that disorders would be real to the extent that they are biological; I will explicitly criticize that conception in section 4, below.
6. Indeed, this is a criticism leveled against some diagnostic categories: for example, the reality of ADHD is questioned by leveling the charge that traits associated with ADHD in fact are traits shared among children (or, according to Timimi and Taylor (2004) among boys).
7. Note that a different, narrower conception of individual trait is used in personality psychology; in this paper, my intention is to use the individual trait concept in biology rather than that in personality psychology.
8. Notice that neither conclusion need entail that the suffering related to it ought not to be eased by medical and other caregiving.

9. Essentialism about mental disorder (e.g., Kendell & Jablensky, 2003) is also an ITR stance. In it, individual traits that cluster together via their connection to a shared essence act as constituents of the mental disorder in question. I will not engage essentialism in depth in this essay because there is a widespread consensus that essentialism about mental disorders is false (see, e.g., Kendler et al., 2011; Zachar, 2014).
10. For example, Timimi and Taylor (2004) bemoans how the diagnostic category of ADHD "leads to all of us – parents, teachers and doctors – disengaging from our social responsibility to raise well-behaved children. We thus become a symptom of the cultural disease we purport to cure" (ibid, p. 8). According to Timimi, Pete's tragedy is that the incompetent upbringing he received as a child is misinterpreted as a nonexistent medical condition. Neuroreductionism has been one attempt to decisively resolve this matter in favor of grounding mental disorder on individual traits.
11. Kendler compares mental disorders to species, where "The features of a species typically vary over its range, and at its limits the dividing line between sister species can become indistinct [. . .] The species we know about only exist in our biosphere and are temporally limited, existing only between their emergence and extinction [. . .] species have no essence. There is no one thing that defines a species that makes a walrus, robin, or drosophila. Fourth, not all members of a species are identical to one another (2016: 6). For Kendler, mental disorders resemble species on all four counts.
12. Davies does not name his stance a social metaphysics, preferring the term externalism. However, Davies' argument concerning the instantiation of mental disorder makes no use of the nonsocial environment. That there is no ODD in Davies' utopia is due *specifically* to the differences in social structures.
13. As a further remark, that we take certain individuals with specific traits, regardless of their local culture, to have a mental disorder does not entail that the disorder would necessarily need to be presocial: rather, it can entail that similar cultural practices arise in various cultures due to factors such as cultural exchange and the sociobiological characteristics of *Homo sapiens.*

Acknowledgments

This research was funded by the Finnish Cultural Foundation and the John Templeton Foundation.

The author would like to thank Åsa Burman, Markus Eronen, Neil Levy, ChongMing Lim, Guido Robin Löhr, Juha Räikkä, Jukka Varelius, Sam Fellowes, and audiences at the European Society for Philosophy and Psychology annual conference for helpful comments on earlier drafts of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Polaris Koi is a senior researcher in Philosophy at the University of Turku, Finland. More on his work can be found at www.polariskoi.com

ORCID

Polaris Koi http://orcid.org/0000-0001-6152-2230

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, 5th edition*. American Psychiatric Publishing.

Barkley, R. A. (1997). *ADHD and the Nature of Self-Control*. Guilford Press.

Barnes, E. (2017). Realism and social structure. *Philosophical Studies*, *174*(10), 2417–2433 doi:10.1007/s11098-016-0743-y

Beebee, H., & Sabbarton-Leary, N. (2010). Are psychiatric kinds real? *European Journal of Analytic Philosophy*, *6*(1), 11–27. https://eujap.uniri.hr/casopis/abstracts/6_1_1.pdf

Borsboom, D., Cramer, A. O. J., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, *42*(e2), 1–63. doi:10.1017/S0140525X17002266

Boyd, R. (1991). Realism, antifoundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, *61*(1), 127–148. doi:10.1007/BF00385837

Church, J. (2004). Social constructionist models: Making order out of disorder — On the social construction of madness. In J. Radden (Ed.), *The philosophy of psychiatry: A companion* (pp. 393–406). Oxford University Press .

Davies, W. (2016). Externalist psychiatry. *Analysis*, *76*(3), 290–296. doi:10.1093/analys/anw038

Eronen, M. I. (2019). Psychopathology and truth: A defense of realism. *Journal of Medicine and Philosophy*, *44*(4), 507–520. doi:10.1093/jmp/jhz009

Fellowes, S. (2019). Scientific realism, antirealism, and psychiatric diagnosis. In S. Tekin, &R. Bluhm (Eds.), *The bloomsbury companion to philosophy of psychiatry* (pp. 467–484). Bloomsbury.

Haslanger, S. (1995). Ontology and social construction. *Philosophical Topics*, *23*(2), 95–125. doi:10.5840/philtopics19952324

Haslanger, S. (2016). What is a (social) structural explanation? *Philosophical Studies*, *173*(1), 113–130. doi:10.1007/s11098-014-0434-5

Hoogman, M., Bralten, J., Hibar, D. P., Mennes, M., Zwiers, M. P., Schweren, L. S. J., van Hulzen, K. J. E., Medland, S. E., Shumskaya, E., Jahanshad, N., de Zeeuw, P., Szekely, E., Sudre, G., Wolfers, T., Onnink, A. M. H., Dammers, J. T., Mostert, J. C., Vives-Gilabert, Y., Kohls, G., Oberwelland, E., & Seitz, J. (2017). Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis. *The Lancet Psychiatry*, *4*(4), 310–319. doi:10.1016/S2215-0366(17)30049-4

Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? Precisely. *Science*, *348*(6234), 499–500. doi:10.1126/science.aab2358

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160*(1), 4–12. doi:76/appi.ajp.160.1.4

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*(6), 1143–1150. doi:10.1017/s0033291710001844

Kendler, K. S. (2005). Toward a philosophical structure for psychiatry. *American Journal of Psychiatry*, *162*(3), 433–440. doi:10.1176/appi.ajp.162.3.433

Kendler, K. S. (2008). Explanatory models for psychiatric illness. *American Journal of Psychiatry*, *165*(6), 695–702. doi:10.1176/appi.ajp.2008.07071061

Kendler, K. S. (2016). The nature of psychiatric disorders. *World Psychiatry*, *15*(1), 5–12. doi:10.1002/wps.20292

Koi, P. (2021). Genetics on the neurodiversity spectrum: Genetic, phenotypic and endophenotypic continua in autism and ADHD. *Studies in History and Philosophy of Science*, *89* (October 2021), 52–62. doi:10.1016/j.shpsa.2021.07.006

Park, M. T. M., Razhanan, A., Shaw, P., Gogtay, N., Lerch, J. P., & Chakravarty, M. M. (2018). Neuroanatomical phenotypes in mental illness: Identifying convergent and divergent cortical phenotypes across autism, ADHD and schizophrenia. *Journal of Psychiatry & Neuroscience*, *43*(3), 201–212. doi:10.1503/jpn.170094

Schaffner, K. F., & Tabb, K. (2014). Varieties of social constructionism and the problem of progress in psychiatry. In K. S. Kendler, and J. Pargas (Eds.), *Philosophical issues in psychiatry III* (pp. 85–106). Oxford University Press .

Schaffner, K. F. (2013). Reduction and reductionism in psychiatry. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. StanghelliniT. Thornton (Eds.), *The oxford handbook of philosophy and psychiatry* (pp. 1003–1022). Oxford University Press .

Tabb, K. (2019). Philosophy of psychiatry after diagnostic kinds. *Synthese*, *196*(6), 2177–2195. doi:10.1007/s11229-017-1659-6

Timimi, S., & Taylor, E. (2004). ADHD is best understood as a cultural construct. *British Journal of Psychiatry*, *184*, 8–9. doi:10.1192/bjp.184.1.8

Varelius, J. (2009). Defining mental disorder in terms of our goals for demarcating mental disorder. *Philosophy, Psychiatry & Psychology*, *16*(1), 35–52. doi:10.1353/ppp.0.0215

Vehmas, S., & Mäkelä, P. (2009). The ontology of disability and impairment. In K. Kristjansen, S. Vehmas, and T. Shakespeare (Eds.), *Arguing about disability: Philosophical perspectives* (pp. 42–56). Routledge .

Violle, C., Navas, M., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, *116*(5), 882–892. doi:10.1111/j.0030-1299.2007.15559.x

Zachar, P. (2002). The practical kinds model as a pragmatist theory of classification. *Philosophy, Psychiatry and Psychology*, *9*(3), 219–227. doi:10.1353/ppp.2003.0051

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