

Supervenience and Psychiatry: Are Mental Disorders Brain Disorders?

Charles M. Olbert
Fordham University

Gary J. Gala
University of North Carolina School of Medicine

Prominent psychiatrists have moved to rebrand psychiatry as clinical neuroscience and rechristen mental disorders as brain disorders. Recent shifts in research and funding priorities have followed suit, privileging neuroscience over psychological and behavioral research. With the possible exception of identifying general paresis with advanced syphilitic brain infection, however, no theorized identities between mental and brain disorders have been empirically corroborated. Consequently, we regard the thesis that mental disorders are brain disorders as an ontological hypothesis. Any robust formulation of the hypothesis that mental disorders are brain disorders logically requires the minimal thesis that mental disorders supervene upon brain disorders. A mental disorder supervenes upon a brain disorder if there could be no change in the mental disorder without a change in the brain disorder. In this paper we analyze contemporary diagnostic criteria used to individuate certain mental disorders to argue that at least some mental disorders fail to supervene upon brain disorders. Hence, we conclude that at least some mental disorders are not and cannot be (merely) brain disorders. This conclusion highlights a basic heterogeneity in psychiatry's subject matter: some mental disorders constitutively involve psychological experiences or sociocultural relationships to the external environment that cannot be identified with or reduced to brain states or functioning. We propose that establishing cases of supervenience failure represents a method for discriminating between more robustly mental (as opposed to brain) disorders.

Keywords: psychiatric disorders, philosophy of psychiatry, mind-brain identity, *DSM-5*, Research Domain Criteria (RDoC)

Recent trends have brought the vexed question of the nature of mental disorders into focus by privileging neuroscientific and medical paradigms. In the past decade, thought leaders in psychiatry have moved to rebrand psychiatry as

clinical neuroscience (Insel & Quirion, 2005; Murphy, 2006; Reynolds, Lewis, Detre, Schatzberg, & Kupfer, 2009) and rechristen mental disorders as brain disorders or brain circuit disorders (Insel & Cuthbert, 2015). Researchers and policymakers often claim, for example, that mental disorders are brain disorders, have established biological causes, stem from chemical imbalances in the brain, or are known to be medical illnesses just like any other (for an overview, see Deacon, 2013).

The movement to rebrand psychiatry as clinical neuroscience and to identify mental disorders with brain disorders presupposes a disease model of mental disorders that privileges physiology over psychology. If mental disorders are indeed brain disorders, it seems to follow that mental disorders can, in principle, be conceptualized, diagnosed, and treated without essential reference to physical or social conditions out-

This article was published Online First July 27, 2015.

Charles M. Olbert, Department of Psychology, Fordham University; Gary J. Gala, Department of Psychiatry, University of North Carolina School of Medicine.

An earlier, abbreviated version of this article was presented at the 2014 Annual Midwinter Meeting of the Society for Theoretical and Philosophical Psychology. The authors are grateful to Daniel D. Moseley, the UNC Chapel Hill Philosophical Issues in Psychiatry Reading Group, and Kathryn Tabb for helpful discussion of these ideas.

Correspondence concerning this article should be addressed to Charles M. Olbert, Department of Psychology, Fordham University, Dealy Hall 218A, 441 East Fordham Road, Bronx, NY 10458-9993. E-mail: colbert@fordham.edu

side a person's brain. According to a disease model it would, for example, be possible to diagnose depression via a blood test (see, e.g., Redei et al., 2014), obviating the need to rely upon symptoms based in psychological, social, and cultural factors or experiences. Such privileging of brain circuits over psychological constructs represents a logical extension of the neo-Kraepelinian framework, which holds that psychiatry ought to focus on the biological aspects of mental disorders (Deacon, 2013; Klerman, 1978; Patil & Giordano, 2010). Indeed, many theorists have argued that psychiatry's putatively atheoretical classification system, which has its roots in the neo-Kraepelinian framework, rests upon a biomedical view of diagnostic categories in which disease entities are presumed to underlie descriptive symptom criteria (e.g., Cooper, 2004; Deacon, 2013; Nelson-Gray, 1991). The Research Domain Criteria (RDoC) initiative exemplifies the hope of discovering such disease entities by aiming to supplant the current diagnostic paradigm with more basic neuroscience-based classifications (Insel et al., 2010).

Claims equating mental disorders with brain disorders should not be regarded as theoretical or rhetorical marginalia, but rather as tangible threats to psychological research and practice. Such claims motivate an overarching framework for research into and treatment of mental disorders: if mental disorders are brain disorders, and if brain disorders should be treated via neuroscientific interventions (targeting the neural, chemical, or genetic underpinnings of brain pathophysiology), then it follows that mental disorders should be treated via neuroscientific interventions (and correlatively, it suggests that mental disorder research should focus on neuroscience). This framework represents policy, not rhetoric. Under the banner of the RDoC initiative (Insel et al., 2010), the National Institute of Mental Health (NIMH) has proposed a strategic plan in which funding priorities focus on biological factors in mental disorders at the expense of psychosocial and behavioral factors (Teachman et al., 2014).

The thesis that mental disorders are brain disorders represents a philosophical position not fully grounded by the available empirical evidence, since brain-behavior correlations have, at best, only partially supported this putative identification. The paradigm (and indeed, only)

case of a possible identification of a mental disorder with a brain disorder—namely, the discovery that general paresis is caused by advanced syphilitic infection of the brain (Ghaemi, 2013)—has not generalized. As yet, no brain disorder has been established as identical with any mental disorder defined in the *Diagnostic and Statistical Manual of Mental Disorders* (5th edition; *DSM-5*; American Psychiatric Association, 2013). Moreover, no means currently exist for reliably diagnosing mental disorders by genetic or neuroscientific tests (Kapur, Phillips, & Insel, 2012), and no unambiguous biomarkers exist for paradigmatic mental disorders such as schizophrenia (Lakhan & Kramer, 2009). Consequently, the thesis that mental disorders are brain disorders is best understood as an ontological hypothesis regarding the nature of mental disorders that awaits empirical validation.

In our view, the claim that mental disorders are brain disorders represents a powerful overarching theoretical principle wielded with the intent of dominating the practice, research, and politics of all mental health disciplines, not merely psychiatry. The sobering lack of evidence supporting this claim has been insufficient to stem the tide of enthusiasm for exclusionary neuroscientific approaches in psychiatry. Because the claim that mental disorders are brain disorders arguably diverts dollars away from research into psychosocial, behavioral, and cultural aspects of mental disorders, we believe that psychologists disagreeing with this view should vocally resist “brain disease” rhetoric by emphasizing the psychological, social, and cultural constitution of mental disorders. If we are correct that the identification of mental disorders with brain disorders represents an ontological hypothesis, then psychologists should have conceptual arguments ready to hand to refute or problematize this hypothesis.

In this paper we offer a conceptual argument against the hypothesis that mental disorders are brain disorders by adapting conceptual tools developed by philosophers of mind. Specifically, we turn to the concept of supervenience with respect to mental disorders: a mental disorder supervenes upon a brain disorder if there could be no change in the mental disorder without a change in the brain disorder. Although supervenience has been recognized as a relevant

concept in psychiatry (Glannon, 2002; Radden, 1996; Woodward, 2008), its application to specific mental disorders has not been explored in adequate depth, and the possibility of supervenience failures has been underexplored; for example, supervenience is sometimes mentioned in passing as if it were simply a truism that all of mental life supervenes upon brain processes (e.g., Brülde & Radovic, 2006). In contrast, we argue that careful attention to supervenience can be used to undermine the claim that mental disorders are brain disorders. Our argument, in brief, is as follows:

1. If (all) mental disorders are brain disorders, then (all) mental disorders supervene upon brain disorders.
2. Not all mental disorders supervene upon brain disorders (i.e., some fail to so supervene).
3. Therefore, it is false that (all) mental disorders are brain disorders.

In the first section of the paper we establish our terminology and background assumptions. We then proceed to define what is meant by supervenience and to substantiate the first premise above: that the thesis that mental disorders are brain disorders logically entails that mental disorders supervene upon brain disorders. Following this, we explore general considerations regarding supervenience, the logical consequences of supervenience failure, and the application of supervenience to mental disorders, generally speaking. At this point we turn to an analysis of specific mental disorders, marshalling thought experiments attending to the criteria by which mental disorders are individuated (i.e., identified and differentially diagnosed) to show that some mental disorders fail to supervene upon brain disorders, substantiating the second premise above. Finally, we consider some implications of this argument.

In essence, then, we argue that a basic heterogeneity exists in psychiatry's subject matter such that not all mental disorders can be identified with brain disorders. In our view, this heterogeneity arises because at least some mental disorders constitutively involve psychological experiences or sociocultural relationships to the external environment that cannot be identified with or reduced to brain states or functioning. We contend that establishing supervenience

failures represents one way—perhaps a central way—to identify more genuinely mental (as opposed to brain) disorders, and thus to delineate the proper logical boundary where pure neuroscience research might be successful (for cases of supervenience) and where psychological, social, and culture considerations would be ineliminable (for cases of supervenience failure).

Assumptions and Terminology

Before delving into issues of supervenience, we offer some background discussion of terminology and our assumptions. We assume that psychological states and psychological types exist and have explanatory value and causal efficacy, and that mental disorders involve such states and types. We stake no claim, however, as to the ultimate nature of psychological states/types, and we neither affirm nor deny that psychological states/types are identical with physical states/types. Some positions—namely, eliminative materialism (Churchland, 1981)—deny outright that psychological states and types exist. Eliminative materialists cannot therefore endorse the claim that mental disorders are brain disorders for the reason that they deny that mental states and types exist. We do not further address this position.

Theses about supervenience, identity, and reduction in the philosophy of mind are often discussed with respect to mental and/or brain states or properties. We acknowledge, however, that mental disorders possess greater complexity than mental states or individual mental properties, and so we adapt philosophy of mind considerations to a higher level of abstraction that concerns sets of (rather than individual) mental and brain states and properties. In general, mental disorders comprise many symptoms, and current diagnostic criteria allow that the same disorder may manifest distinct, even divergent symptom presentations (Olbert, Gala, & Tupler, 2014). What's more, individual symptoms may possess greater complexity than simpler mental states. Fatigue in the context of depression, for example, involves the temporal unfolding of a complex experience that may include subjective feelings of tiredness, lack of motivation, muscle weakness, and so forth. For the purposes of our discussion, we represent symptoms as sets of properties and states, and

for the purposes of this paper we understand mental disorders to be sets of symptoms, which are, in turn, sets of properties and states.

By the term “brain” we understand a living collection of neurons (and relevant supporting tissue and chemicals, e.g., glial cells and neurotransmitters) along with the interconnections between those neurons connected up via a spinal cord to a living human body. Analogous to mental disorders, brain disorders represent complex amalgams of brain properties spanning a variety of different cases ranging from gross anatomical lesions to neural connectivity deficits to more subtle neurotransmitter hypo- or hyperregulatory processes. As with mental disorders above, we construe brain disorders as sets of neurophysiological properties and states for the purposes of adapting philosophical claims to this domain.

Third, and centrally, we assume that the thesis that mental disorders are brain disorders represents a robust claim about the nature of mental disorders. If this thesis did not represent just such a robust claim, it could not plausibly motivate paradigm shifts in mental health research and funding priorities. As we previously suggested, this thesis is not rhetorical, but rather a substantive theoretical position. For example, the neuroscientist Eric Kandel states that “all mental processes, even the most complex psychological processes, derive from operations of the brain” (Kandel, 1998, p. 460). “Derive” here represents a strong claim, for Kandel goes on to state as a “corollary” of this principle: “behavioral disorders that characterize psychiatric illness are disturbances of brain function, even in those cases where the causes of the disturbances are clearly environmental in origin” (p. 460, emphasis added). We return later to considering the details behind the claim that mental disorders are brain disorders.

The Concept of Supervenience

The concept of supervenience bears directly upon the issue of whether mental disorders are brain disorders. Indeed, we argue that any robust thesis to the effect that mental disorders are brain disorders logically entails that mental disorders supervene upon brain disorders. Before making this argument, we first explore supervenience in more detail. “Supervenience” names a relation between sets of lower- and higher-order

properties: a set of M-properties supervenes upon a set of B-properties when there could not be a difference in M-properties without there being a difference in B-properties. In terms of mental disorders, the thesis that mental disorders supervene upon brain disorders amounts to the claim that two individuals¹ who have the same brain disorder(s) could not have a difference in their mental disorder(s), and that two individuals could not manifest a difference in mental disorder without also manifesting some difference in brain disorder.

Supervenience by itself weakly captures the intuition that mental disorders are strongly dependent on processes internal to the individual: although the supervenience of mental disorders upon brain disorders allows that there might be mental aspects of disorders that are not intrinsically identical to the physical aspects of those disorders, supervenience also entails that those mental aspects necessarily correlate with or depend upon the physical aspects.² Supervenience thus preserves the clinical neuroscientist’s intuition that brain disorders constitute the physical basis of (and perhaps the explanatory locus for) mental disorders. Supervenience also captures the intuition that two individuals might manifest the same clinical syndrome yet suffer from distinct brain disorders (in other words, supervenience accommodates multiple realizability) while allowing for some autonomy of the mental. Unlike stricter identity theses, supervenience does not require that mental disorders and brain disorders must share all properties in common. In short, supervenience highlights the neural, physical basis of mental disorders and renders psychological, social, and cultural aspects of mental disorders secondary without necessarily entirely obviating them.

Supervenience theses come in many forms. For the purposes of this discussion we understand supervenience to refer to *local supervenience*, in the sense that mental disorders are construed as supervenening upon disorders of the three-pound lump of meat within a person’s

¹ Or the same individual at different times or across counterfactual situations.

² Strictly speaking, supervenience by itself does not suffice for explanatory or logical dependence in the manner implied, although such dependence follows on some widely held assumptions about the priority of physical over mental properties (Kim, 2005; van Cleve, 1990).

skull (plus the spinal column). One could also construct a thesis about *global supervenience* (McLaughlin, 1995), holding that mental disorders supervene upon the omnibus physical state of the world. Because we understand the clinical neuroscientist's identification of mental disorders with brain disorders as arising from the project of attempting to conceptualize, diagnose, and treat mental disorders via neuroscientific interventions, we understand supervenience here to refer to a relationship of mental disorders to local brain disorders rather than to the global physical state of the world.

Local supervenience comes in stronger and weaker varieties. Strong supervenience (Kim, 1984) can be formulated as follows, recalling that we construe mental and brain disorders as sets of properties: a mental disorder (*M*) *strongly supervenes* upon a brain disorder (*B*) if and only if, necessarily, for each person *X* who has *M* and *B* and for each property *F* in *M*, if *X* has *F*, then then there is a property *G* in *B* such that *X* has *G*, and necessarily if any person *Y* has *G*, she has *F*. *Weak supervenience* is weaker with respect to its modal force, and is formulated analogously to strong supervenience omitting the second 'necessarily.' Using the language of possible world semantics, weak supervenience says that within a given possible world (e.g., our actual world) there are no individuals who are indiscernible with respect to brain disorder but discernible with respect to mental disorder; strong supervenience enhances this claim to apply across all individuals across all possible worlds.

We hereafter focus on the concept of weak supervenience. For one thing, from a practical, empirical standpoint, only the facts about disorders in our actual world (rather than in all other possible worlds) bear upon research and treatment for disorders in the actual world. Clinicians and bench scientists are rightly more concerned with actualities than philosophical abstractions. More importantly, however, we believe that weak supervenience represents the "canary in the coal mine" for the thesis that mental disorders are brain disorders. Specifically, we argue that regardless of how one elaborates the thesis that mental disorders are brain disorders, adopting this thesis logically commits the clinical neuroscientist at a minimum to the view that mental disorders weakly supervene upon brain disorders.

First, we note that the thesis that mental disorders are brain disorders admits of various elaborations. One method for rigorously posing the thesis involves strictly identifying either types of mental disorders or tokens (instances) of mental disorders with types or tokens of brain disorders. The putative identity of mental and brain states has been extensively debated in the philosophy of mind literature (Armstrong, 1968; Chalmers, 1996; Kripke, 1980; Smart, 1959), although both type and token identity theses have been criticized; for example, type identity does not attest to the possibility that mental disorders may be *multiply realized* by an array of distinct physiological states of the brain (Putnam, 1967). Other possible ways to cash out the claim that mental disorders are brain disorders would be to state that mental disorders are reducible to, depend on, are determined by, or are sufficiently explained by brain disorders.

Although the thesis that mental disorders are brain disorders is ambiguous between various more specific positions, they share a common conceptual core, namely, weak supervenience. Consider the following: reduction of the mental to the physical straightforwardly entails the supervenience of the mental upon the physical (McLaughlin, 2006). Identity theses also entail supervenience as follows. Token identity (the thesis that each instance of a mental disorder is identical to some brain disorder instance) logically entails supervenience (but not vice versa). This can be shown as follows: if a mental disorder instance (*m*) and a brain disorder instance (*b*) are identical, then any change to *m* entails a change in *b* since identity implies that *m* and *b* share all properties in common. Type identity (the thesis that each type or class of mental disorder is identical to some type of brain disorder) logically entails token identity (Fodor, 1974), and therefore type identity transitively entails supervenience. Positions such as Kandel's (1998, quoted above) have also been argued to require a version of nonreductive physicalism involving supervenience of the mental upon the physical (Van Oudenhove & Cuypers, 2010), and it is generally recognized that physicalist approaches to the mind and mental states require that the mental supervenes upon the physical (Kim, 2005).

More generally, examining the implications of supervenience relations failing to hold illuminates the connection between supervenience

and the thesis that mental disorders are brain disorders. If mental disorders did not supervene upon brain disorders, then there could be two individuals with identical brain disorders who nonetheless have different mental disorders; alternatively, one individual with a given brain disorder could have a corresponding mental disorder at one time and then achieve remission of that mental disorder without the brain disorder changing or remitting. Such scenarios would render any putative links between mental disorders and neuroscience purely contingent and nonnomological, and would seriously undermine the aspirational claim that mental disorders should, as a rule, be treated by neuroscientific interventions. For these reasons, we conclude that irrespective of whether the clinical neuroscientist intends identity, reduction, or nonreductive neuroscientific physicalism with respect to mental disorders, the view that mental disorders are brain disorders logically entails that mental disorders supervene upon brain disorders. Any specific view might entail either a strong or weak version of the supervenience thesis, but because strong supervenience logically implies weak supervenience (Kim, 1984), the view that mental disorders are brain disorders logically requires weak supervenience irrespective of the specific entailment. We focus on weak supervenience because, as we have argued, it represents a shared logical commitment across different possible ways of elaborating the thesis that mental disorders are brain disorders.

Supervenience and Psychiatric Disorders: General Considerations

To summarize, we have argued that the thesis that mental disorders are brain disorders entails the thesis that mental disorders supervene (at least weakly and locally) upon brain disorders. Supervenience therefore represents a theoretical keystone: if the logically weaker thesis of supervenience is false, then the logically stronger position that mental disorders are brain disorders must also be false. In subsequent sections, we shall argue that, in fact, some mental disorders fail to supervene upon brain disorders for the reason that their psychological, social, and cultural aspects represent primary, constitutive factors in the individuation of those mental disorders. Before making this argument, we pres-

ently explore an example far afield from mental disorders in which supervenience fails in order to clarify the dialectical stakes. We then turn to some general considerations that call into question the blanket claim that mental disorders supervene upon brain disorders.

What Would a Failure of Supervenience Show?

M-Properties supervene upon B-properties when there could not be a difference in M-properties without a difference in B-properties. Hence, if scenarios can be constructed in which two individuals could be equivalent with respect to B-properties yet distinct with respect to M-properties, this would constitute evidence that M-properties fail to supervene upon B-properties.

Consider money. Suppose that we have two individual hundred dollar bills that have equivalent microphysical properties: each is made of the same type of paper, marked with the same patterns of green ink, printed with the same serial number, and so forth. One bill was issued by the U.S. Federal Reserve; the second bill, however, was counterfeited by a forgery ring tied to international terrorism. Despite having equivalent microphysical properties, the first bill is legal tender for all debts private and public, whereas spending the second bill might get you brought in for police questioning. The two bills thus have equivalent microphysical properties but distinct properties writ large, especially with respect to their historical and causal properties. Specifically, the properties “being legal tender” and “being a forgery”—and, therefore, the causal properties of bills qua money generally speaking—fail to (locally) supervene upon the microphysical properties of printed bills. If the properties of money did, in fact, supervene upon microphysical properties, no principled distinction would exist between genuine and counterfeit bills. Although it would not be possible to detect differences between the two bills given mere observation at a single time, this does not threaten the conclusion that supervenience fails. Supervenience is an ontological fact, while our inability to visually inspect differences between two objects is an epistemic fact. Despite our ignorance, it would remain a (physical, causal, historical) fact that

the U.S. Federal Reserve issued one bill and not the other.

In general, entities individuated on the basis of their relational properties or causal histories may fail to supervene upon the local, microphysical properties of those entities. The money example highlights an important point for discussions about mental disorders. Note that the genuine versus counterfeit status of each bill depends solely upon entirely mundane physical, causal, and historical facts. Supervenience fails in the foregoing case because some of the properties of money are relational historical properties that involve facts about the world beyond the local microphysical properties of bills. Drawing upon the distinction between local and global supervenience explained above (McLaughlin, 1995), we emphasize that although the relational properties of money fail to locally supervene upon the microphysical properties of bills, the properties of money, generally speaking, do, in fact, globally supervene upon the physical state of the world. This distinction illustrates that the failure of (local) supervenience does not necessitate positing mysterious, nonphysical, unscientific properties. Likewise, even if mental disorders fail to (locally) supervene upon brain disorders, we need not forswear mundane, naturalistic, scientific approaches to understanding mental disorders, for mental disorders may very well globally supervene upon the physical state of the world.

Supervenience and Mental Disorders

In addressing whether mental disorders supervene upon brain disorders we focus on mental disorders as defined in the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013). Although the field may eventually move away from the *DSM-5* in favor of new classifications stemming from the RDoC initiative, no such classification system currently exists, and the methods and criteria used in practice to individuate mental disorders represent the best available guide to the nature of those disorders. At present, mental disorders are individuated based on the criteria defined in the *DSM-5*. Although the *DSM-5* recognizes that “current diagnostic criteria for any single disorder will not necessarily identify a homogenous group of patients” (p. 20) and that “diagnostic criteria

sets do not constitute comprehensive definitions of underlying disorders” (p. 19), these diagnostic criteria continue to be used in clinical diagnostic practice and to select cases for etiological research (see, e.g., Bosker et al., 2011; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013).

In practice with the *DSM-5*, clinicians diagnose an individual with major depressive disorder (as opposed to generalized anxiety disorder or schizophrenia or no disorder at all) on the basis of his or her having symptoms that meet the criteria for major depressive disorder and not those of other disorders. One strategy for demonstrating that a given mental disorder fails to supervene upon brain disorders consists in constructing scenarios in which two individuals (or the same individual at different times or in counterfactual environments) could be equivalent with respect to a brain disorder even though only one meets *DSM-5* diagnostic criteria for the mental disorder while the other does not. Two overarching principles that guide practical individuation of mental disorders might appear to suggest a general strategy. First, mental disorders tend to be individuated on the basis of contextual factors such as cultural standards; the definition of “mental disorder” in *DSM-5* states that “an expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder” (p. 20). Second, most mental disorders contain an exclusion criterion such as “the [disorder or episode] is not attributable to the physiological effects of a substance or to another medical condition” (p. 161). Such exclusion criteria highlight that whether a particular syndrome constitutes a mental disorder depends on the syndrome’s causal history.

Nevertheless, the criteria and exclusionary clauses for individual mental disorders are myriad, and we doubt that a general argument for or against supervenience exists at the level of the broad definition of “mental disorder.” Indeed, no such definition existed prior to the fourth edition of the *DSM* (American Psychiatric Association, 2000), and the definition of ‘mental disorder’ was a contested issue in the creation of *DSM-5* (Stein et al., 2010). Changes to the definition of “mental disorder” might render such an argument moot, and relying solely upon common features of mental disorders may inadvertently beg crucial questions by illegitimately

homogenizing the potentially heterogeneous category of mental disorders. We turn instead to examining individual mental disorders.

Analyses of Specific Mental Disorders

We take on two assumptions for the sake of argument. First, we provisionally adopt realism about mental disorder (i.e., that psychiatric diagnoses reflect genuine disorders that exist independent of anyone's beliefs about those disorders). Second, we provisionally accept that *DSM-5* diagnostic criteria suffice to individuate the disorder described by those criteria: we assume that *DSM-5* symptoms as listed are indicators of a given disorder and likely at least partly constitutive of that disorder, even if they do not constitute an exhaustive definition of that disorder.

Our argument focuses on specific diagnostic criteria, an approach which draws two *prima facie* challenges. First, the clinical neuroscientist might object that future neuroscience will supplant the symptoms currently associated with specific disorders. This objection, however, would beg the question with respect to a crucial issue to which we later turn, namely, the relation between mental disorders and their symptoms. Moreover, as previously stated, current methods for individuating disorders represent our best guide at present to the nature of those disorders.

Second, it might be thought that the heterogeneity inherent in the *DSM-5*'s polythetic diagnostic criteria sets (defined by multiple symptoms, not all of which are necessary to meet diagnostic thresholds) render mental disorder criteria too imprecise or inconsistent to serve as candidates for a supervenience relationship. Although we agree that polythetic heterogeneity presents challenges for psychiatric research (Olbert et al., 2014), appeal to the notion of alternative supervenience bases sidesteps this issue. Mental disorders might be construable as supervening upon their symptoms, such that people have mental disorders because or in virtue of their symptoms; the mental disorder symptoms could, in turn, be argued (by the clinical neuroscientist) to supervene upon brain disorders, thereby creating a logical chain of supervenience relationships. Kim (1984) notes that, in general, "a supervenient property will have *alternative supervenience bases* – base properties

that are each sufficient for the supervening property" (p. 165, emphasis original), and so heterogeneity poses no necessary constraint on the possibility of supervenience relationships. Moreover, supervenience only requires consistency of judgments (Kim, 1984) to the effect that two individuals who have the same symptoms (i.e., properties in the supervenience base) must have the same disorder, not that two individuals who have the same disorder must have the same symptoms. Irrespective of these issues, however, one disorder upon which we focus (specific phobia) does not rely upon polythetic criteria and so does not evince the sort of imprecision at issue, and the sort of criteria on which we focus in the case of personality disorders are so pervasive in this disorder category as to render minor any concerns about heterogeneity.

In any case, we turn later to addressing important issues regarding the distinction between disorders and their symptoms. We presently argue that one specific disorder type (specific phobia) and one class of disorders (personality disorders) fail to supervene upon brain disorders.

Specific Phobia

In this section we employ thought experiments to show that a coherent conceptualization of specific phobia constitutively requires certain (types of) relations between individuals and their physical and/or social environments, and that brain states alone cannot provide a general adequate account for the nature of phobias. Specific phobia requires "marked fear or anxiety about a specific object or situation" (*DSM-5*; American Psychiatric Association, 2013, p. 197). Two criteria for specific phobia carry supervenience implications.

Criterion D: Normative standards. Specific phobia requires fear or anxiety that "is out of proportion to the actual danger posed by the specific object or situation and to the socio-cultural context" (Criterion D; American Psychiatric Association, 2013, p. 197). The words "out of proportion" hint at the normative standards clearly invoked in the fourth edition of the *DSM*, which required that "the person recognizes that the fear is excessive or unreasonable" (*DSM-IV-TR*; American Psychiatric Association, 2000, p. 449). The implications for super-

venience are more vivid in the *DSM-IV-TR* criteria. "Recognize" is a success-verb. A person cannot successfully recognize that a fear is excessive or unreasonable without that fear actually being excessive or unreasonable. Whether a fear is unreasonable, however, is not a fact about a person's brain. Whether a fear is unreasonable depends upon facts about the social, cultural, and linguistic community in which a person happens to be embedded. Reasonableness does not plausibly depend only upon the brain.

Similar issues can be raised for the *DSM-5* definition. It seems unlikely that a person's fear being "out of proportion" to actual danger solely concerns facts about that person's brain. A more cogent interpretation would make reference to the social, cultural, and linguistic community in which he or she is embedded, and define "out of proportion" relative to the norms of that community. The following thought-experiment spells out this argument in greater vividity.

Thought-experiment. Consider two individuals: Ana and Anansi. Suppose, for the sake of argument, that each woman has relevantly equivalent brain states and properties corresponding to an intense fear of spiders. Ana lives in a culture where spiders are considered normal household pests, and are alternately tolerated, ushered calmly out of the house, casually killed, or targeted for extermination, depending on individual preference. Anansi, however, lives in a culture where local religion enshrines a Lucifer-like evil deity figure that takes the form of spiders in order to spy upon, poison, and terrorize the wicked. In this culture, encountering a spider is normatively understood as an ill omen, a sign of one's own wickedness. The normative standards of Ana's community render her fear grossly out of proportion, whereas Anansi's community perceives her fear as a sign of righteousness and considers her frantic efforts to avoid spiders laudable. Although the brain states and cued behavior of each of Ana and Anansi in the presence of spiders are in all relevant respects equivalent, differences in the reasonableness of their respective fears track differences in cultural norms. Thus, there could be a difference in mental disorder without a corresponding difference in brain states/properties. Further, since brain disorders represent special cases of (sets of) brain states/

properties, specific phobias do not supervene upon brain disorders.

Objections. One objection to this sort of thought-experiment consists in noting that it is, for all intents and purposes, impossible for two individuals to have precisely equivalent brain states. We grant that this is, empirically speaking, an implausible presupposition. We raise the issue in this way to clarify the conceptual implications of the claim that mental disorders supervene upon brain disorders. A case in which two individuals have precisely the same brain disorder represents the best-case scenario for the clinical neuroscientist who wishes to claim that mental disorders are brain disorders. If it could be shown that mental disorders are brain disorders, surely this could be shown with the greatest ease when the brain disorders in question are precisely equivalent. If it cannot be demonstrated in this case, it cannot logically be demonstrated at all. Hence, although scientists may balk at the empirical improbability of the thought-experiment as outlined, our strategy has been to argue for our position by granting the clinical neuroscientist's logically best-case position. Although this methodology may be unfamiliar in psychology and psychiatry, it is par for the course in philosophy, where the goal is to clarify the logical structure of arguments and the conceptual implications of theories.

The clinical neuroscientist might also object that "out of proportion" represents a placeholder, in the sense that the *DSM-5* only instrumentally relies on normative standards and that sufficiently advanced clinical neuroscience will one day be able to define phobic fear solely in terms of physiologically excessive limbic system activity. This objection may fall prey to a version of Moore's (1903) "open question" argument. We admit the possibility of calibrating limbic system activity to the subjective sense of fear or distress upon being exposed to phobic loci and thereby determining general correlations between limbic activity and proportionality of the fear. However, a threshold beyond which limbic activity is deemed out of proportion must be chosen by reference to community standards, and it would always be an open question whether a given person with a specific level of limbic system activity genuinely experiences fear out of proportion to those standards. Individual differences in neural physiology may render it possible for an individual to exceed a

defined threshold yet not experience distress sufficient to warrant the label “unreasonable.” We grant that it remains an empirical question whether such individual differences truly exist.

The argument might be pressed further. The clinical neuroscientist might maintain that there exists a threshold of limbic system activity beyond which no human being could continue functioning normally even in principle, just as there surely exist degrees of physiological trauma that preclude, in principle, certain movements or functions. Just as severing the spinal column precludes volitional leg movements, flooding the limbic system with chemical or electrical stimuli may preclude functional behavior. It might be argued that, evolutionarily speaking, phobic fear arises from objective malfunctioning of the limbic system (Wakefield, 1992, 1995). We do not believe this argument succeeds. As Murphy and Woolfolk (2000) have pointed out, disorder may exist absent gross malfunction. For example, there exist temperature thresholds above which human organ systems simply cannot function; nevertheless, heat stroke typically occurs at temperatures lower than threshold temperatures that cause objective organ mechanism malfunctions. Heat stroke per se thus cannot be adequately characterized by reference to thresholds of objective malfunction. Similarly, we see no reason, barring extensive empirical evidence to the contrary, to conflate maladaptive fear with malfunctioning limbic system activity.

Another possible response for the clinical neuroscientist consists in denying that the reasonableness of behaviors or beliefs depends upon cultural norms. Put another way, one could claim that a fact of the matter exists as to whether someone’s behaviors or beliefs are reasonable irrespective of how those behaviors and beliefs are regarded by any given community. This objection, however, entails two troubling commitments. First, this objection entails that one may consistently believe that mental disorders are brain disorders only if one has a prior commitment to realism about, and the objective status of, norms. Second, insisting that reasonableness is not sensitive to social context appears to undermine the justification for including cultural exclusions or culture-bound syndromes as genuine factors in mental disorders. We now proceed to briefly examining a

similar thought-experiment involving a different criterion of specific phobia.

Criterion B: Context and fear cues. The requirement that “the phobic object or situation almost always provokes immediate fear or anxiety” points to the importance of physical context. Consider two individuals, DJ and Doje, with equivalent affective and perceptual neural circuitry. DJ grew up in the rural American Midwest, while Doje grew up in Tibet. Despite drastically different environments, DJ and Doje’s neural circuits are so configured that they would each experience intense and irrational fear upon being exposed to heights. A constellation of brain structures, neurotransmitter deficiencies, and neural networks corresponding to fear activates whenever Doje sees or approaches the rims of Tibetan plateaus. DJ, however, never experiences such fear so long as he remains ensconced in flat farmland.

In this example, Doje warrants a diagnosis of specific phobia, while DJ does not because he does not satisfy Criterion B: because the events that would trigger his alleged phobia never occur, he never experiences phobic fear. At most, we can consider DJ biologically prepared to experience such fear should he ever encounter sufficient heights. Now, in our understanding, the principle motivation behind insisting that mental disorders are brain disorders consists in the perception that physiological substrata of disorders are more important, basic, necessary, or causally relevant than social, psychological, or contextual aspects of disorders. Indeed, the clinical neuroscientist seems committed to the view that the psychological and social aspects of disorders possess only instrumental importance as signifiers of underlying dysfunction (a point considered in more detail below). In the present case, any underlying neurobiological dysfunction would have to be explained dispositionally, such that the very identity of certain neurobiological states consists in their propensity to trigger immediate experiences of fear or anxiety were an individual to encounter a certain type of phobic focus.

In principle, we do not take issue with understanding—even individuating—brains and brain states on the basis of their dispositional properties or on the basis of their relational properties. Indeed, there exists philosophical precedent for doing just that (Burge, 1986). However, as specified earlier, we understand

brains for the purposes of this argument as neurons and the interconnections between them, such that brains and their properties exist within skulls (plus spinal cords), and can be individuated and described without reference to the environment beyond the skull. To deny that equivalent neurophysiological configurations across different environments (or at different times) entail brain-equivalence is already to grant that mental disorders cannot be individuated only with reference to neurophysiological configurations, for under this conceptualization of brains, not even brains themselves can be individuated only with reference to neurophysiological configurations.

In short, holding that specific phobia constitutively requires encounters between an individual and an aspect of her environment amounts to a concession that brain-environment relationships—and not brain-states per se—constitute the ontological core of specific phobia. If one grants the stronger thesis that understanding brain disorders (or brain functioning generally) requires understanding how an individual is embedded in his or her environment, then the relevant point has already been ceded that psychiatry cannot be fruitfully (or even coherently) pursued only at the level of local neurophysiological properties.

Personality Disorders

Given what we have said so far, certain personality disorders (PDs) are very unlikely to supervene upon brain disorders. Because the fundamental reasons why features of PDs do not supervene upon brain disorders are essentially similar to arguments for specific phobia, in the interests of brevity we shall simply briefly outline two reasons why PDs, in general, fail to supervene upon brain disorders. We elaborate these considerations for PDs in order to show that the considerations we have raised above for specific phobia are not unique to specific phobia. Although an encyclopedic survey of mental disorders cannot be undertaken here, exploring specific examples helps to demonstrate certain characteristic features of mental disorder criteria (e.g., references to physical and social environments or to social norms) that signal likely supervenience failures.

First, some PDs make explicit reference to the physical and social environment outside of

the individual. One criterion for schizotypal PD, for example, requires “lack of close friends or confidants other than first-degree relatives” (American Psychiatric Association, 2013, p. 656). “Not having close friends or confidants” does not specify a plausible candidate property for neural-level description, although perhaps this trait could be explained dispositionally as in specific phobia above. Antisocial PD involves “repeated failure to sustain consistent work behavior or honor financial obligations” (American Psychiatric Association, 2013, p. 659); yet how someone attends to their job or financial obligations could not be a property of their brain, for the simple reason that facts about one’s job or financial obligations obtain independently of one’s brain state. While it may be possible to infer from brain states that someone believes that she has a job, the truth-value of that belief remains an open question absent further information about the state of the world. Avoidant PD references the avoidance of “occupational activities,” “restraint within interpersonal relationships,” and inhibition “in new interpersonal situations” (American Psychiatric Association, 2013, p. 673). One might object that these symptoms only have the status of examples of how certain internal (neural) deficits might manifest in certain environments. We later return to this consideration.

Second, references to social norms—even, perhaps, moral norms (Zachar & Potter, 2010)—pervade PD criteria. Schizotypal PD involves “odd beliefs or magical thinking that influences behavior and is *inconsistent with subcultural norms*” (American Psychiatric Association, 2013, p. 655, emphasis added here and elsewhere in this paragraph); paranoid PD requires preoccupation with “*unjustified* doubts about . . . friends or associates,” “*unwarranted* fear,” and doubts about partner fidelity that are “*without justification*” (American Psychiatric Association, 2013, p. 649); antisocial PD explicitly references “failure to *conform to social norms*” (American Psychiatric Association, 2013, p. 659); histrionic PD references “*inappropriate* sexually seductive or provocative behavior” (American Psychiatric Association, 2013, p. 667); narcissistic PD mentions “*unreasonable* expectations”; dependent PD involves “difficulty making everyday decisions without an *excessive* amount of . . . reassurance” as well as “*exaggerated* fears of being unable to care”

for oneself (American Psychiatric Association, 2013, p. 675); obsessive-compulsive PD involves being “*overconscientious* . . . about matters of morality, ethics, or values” (American Psychiatric Association, 2013, p. 678). As we saw with specific phobia, such normative language poses a problem for supervenience theses for the reason that the same brain state can only be considered normative or deviant relative to the environment in which the brain state occurs. Although perhaps one can coherently formulate PD criteria while avoiding references to the external social and physical environment or to normative community standards, we suspect such references are ineliminable, not least because a *prima facie* difficulty exists for any account of personality that fails to refer to other people.

DSM Diagnosis: Coda

In this section, we scrutinized psychiatric diagnostic criteria for specific phobia and personality disorders, arguing that these mental disorders are individuated in actual practice on the basis of criteria representing poor candidate properties for supervenience upon brain disorders. In particular, criteria involving elaboration or specification of physical, social, or cultural conditions beyond the individual herself present challenges for supervenience theses, which suggests that disorders involving such criteria are (in part) constituted by psychosocial and cultural factors. This provides *prima facie* evidence that at least some mental disorders do not supervene upon brain disorders. We consider it possible—perhaps likely—that other disorders (major neurocognitive disorder due to Alzheimer’s disease, perhaps) may represent better candidates for successful supervenience theses. The existence, however, of some disorders whose individuation conditions render supervenience theses implausible suggests the possibility that there exists heterogeneity within the basic ontology of the class of entities referred to as “mental disorders.”

Beyond DSM-5: Signs, Symptoms, and Neurobiology

The clinical neuroscientist might object to our reliance on DSM symptom criteria. Although clinicians individuate mental disorders

on the basis of such criteria, one might insist that these symptoms do not constitute the disorder, but rather possess only pragmatic utility, serving as convenient but ultimately eliminable shorthand for identifying more basic disturbances in neural circuitry that give rise to the characteristic behaviors and syndromes identified in psychiatric diagnostic categories. Just as coughing up blood was once taken to be a reliable symptom of tuberculosis but was supplanted by assays to detect mycobacterium infection, one might claim that better understanding of neurobiology will obviate symptom criteria sets such as those in *DSM-5*, revealing them to be purely instrumental. As mentioned in the introduction, the RDoC initiative embodies this hope (Insel et al., 2010). If this conceptualization were correct, then our argument that some mental disorders fail to supervene upon brain disorders premised upon an analysis of current methods of individuating mental disorders (i.e., *DSM-5* diagnostic criteria) would appear to miss the point.

The view that neurophysiological indicators could supplant mental disorder symptoms requires a model in which latent pathology underlies those symptoms (Borsboom, 2008). This disease model underlying RDoC takes its precedent from medical science, in which at least for some disorders pathophysiology may exist and be detected independently of the disorder’s symptoms. For disorders where such essential pathophysiology is both known and observable, observation of that pathophysiology would represent an essential sign of the disorder. An essential sign represents the “thisness” of a disease; in one sense, an essential sign is the disease. An example helps illustrate this concept.

Consider prostate cancer. Certain signs and symptoms indicate prostate cancer, such as nocturia, dysuria, and elevated prostate-specific antigen (PSA). These indicators are not essential signs for prostate cancer, which may be present with or without elevated PSA, nocturia, and so forth. Adenocarcinoma on the prostate (identified by a pathologist), by contrast, represents an essential sign of prostate cancer. Having prostate cancer means the presence of adenocarcinoma in the prostate. Elevated PSA represents a sign of prostate cancer by virtue of probabilistic nomological links between elevated PSA and the presence of an adenocarcinoma of the prostate. Medicine relies on such links between in-

dicators and pathophysiological processes. The existence of such indicators—and hence our ability to make reliable inferences on the basis of their identification—necessarily depends on such indicators being statistically correlated with but ontologically distinct from the disease they indicate. By contrast, no ontological distinction exists between essential signs and the diseases they indicate. One may have elevated PSA yet not have prostate cancer, and one may have prostate cancer yet not have elevated PSA. One may not, however, have prostate cancer without also having an adenocarcinoma of the prostate, and one may not have an adenocarcinoma of the prostate without also having prostate cancer.

For mental disorders, the crucial point concerns whether the symptoms relied upon to individuate mental disorders merely constitute instrumental indicators which could potentially be supplanted once (putative) essential signs of mental disorders are discovered. If so, then it would be coherent to suppose that someone could manifest the (presumably neurobiological) essential signs of a mental disorder without manifesting symptoms associated with that mental disorder, in the same way that someone could manifest the essential signs of prostate cancer (i.e., adenocarcinoma) without manifesting symptoms of prostate cancer (e.g., elevated PSA, nocturia, dysuria). In this context, a doctor can coherently state that a patient has prostate cancer yet has never displayed any (contingent, nonessential) signs or symptoms of prostate cancer. It seems considerably less coherent, however, for a doctor to state that a patient has major depressive disorder yet has never felt depressed, anhedonic, or suicidal; functions normally in her job and relationships; and so forth.

Imagine the following scenario: Jane falls into a coma due to head trauma following a car accident. Doctors perform an X-ray and discover a tumor on Jane's lung. While Jane remains comatose, the doctors operate, resect the tumor, and declare Jane cancer-free. Jane's husband John was also involved in the accident, and likewise falls into a coma due to head trauma. Now suppose, counterfactually and for the sake of argument, that a biomarker (a surrogate endpoint; *Biomarkers Definitions Working Group, 2001*) exists for major depressive disorder that can be detected via spinal tap (e.g.,

the presence of certain depressogenic chemicals beyond a certain threshold). If such a biomarker existed, we might imagine the doctors testing John's spinal fluid and discovering that his depressogenic chemical scores exceed the diagnostic threshold. While John remains comatose, the doctors might prescribe an antidepressant, and after a month, John's depressogenic chemical scores might fall below the diagnostic threshold. In the scenario as sketched, Jane and John have each been in a coma for a month. It is unambiguous that Jane had cancer and was cured of cancer while comatose. It is not unambiguous—and arguably simply false—that John had and was cured of major depressive disorder while comatose. Because he was comatose, by hypothesis he would not have experienced depression in any colloquial or symptomatic sense: arguably, he would not have experienced anything whatsoever. To insist that John had and was cured of major depressive disorder in this thought-experiment is to entirely divorce the scientific category “major depressive disorder” from its natural context of meaning and eliminate its connection to the psychological.

Although we do not place much evidentiary weight on this example, it serves to highlight the conceptual tension inherent in RDoC's aspirations. The position of the clinical neuroscientist requires that symptoms of mental disorders can be separable from their putative underlying pathology, such that John had and was cured of depression. It bears note on this point that conceptualizations of mental disorders exist that do not require latent traits that underlie symptoms (Borsboom & Cramer, 2013). Although we acknowledge that mental disorders involve neurophysiological elements, we find it hard to imagine how symptoms of mental disorders could be separable from their putative underlying pathology. It is hard to imagine meaningfully diagnosing antisocial personality disorder in the absence of antisocial behavior, diagnosing specific phobia absent fear, diagnosing panic disorder absent panic attacks, or diagnosing major depressive disorder absent depression or anhedonia—it is hard, in short, to imagine coherently conceptualizing mental disorders without the mental. These scenarios radiate an aura of absurdity not because our scientific knowledge remains insufficiently advanced, but because these scenarios represent conceptual solecisms. Symptoms appear to play

a distinct role in physical disorders in contrast to many mental disorders.

We posit that perhaps what makes many mental disorders mental is precisely that they involve inherently psychological, social, or culturally based symptoms that *constitute* rather than merely *indicate* the disorder. In other words, at least some mental disorders constitutively involve experiences and relationships to the external environment that cannot be identified with or reduced to brain states or functioning. In our view, for these sorts of mental disorders, their first-personal and intersubjective features ought to be considered not as instrumental indicators or eliminable aspects of the disorder, but rather as the conceptual loci of the disorder itself.

Supervenience and the Heterogeneity of Psychiatry

We examined one individual mental disorder (specific phobia) and one class of mental disorders (personality disorders), arguing that in these instances, it seems unlikely that mental disorders supervene upon brain disorders due to central, possibly ineliminable, reference to physical or social conditions external to the disordered individual's body and brain, such as normative standards of the community in which she is embedded. We consider it an open question whether the same is true for other disorders; in all likelihood, some disorders are more amenable to supervenience theses than others.

The failure of some mental disorders to supervene upon brain disorders obviates the blanket assertion that mental disorders *are* brain disorders. Supervenience represents a minimal, necessary condition for equating mental disorders and brain disorders (i.e., any robust interpretation of this claim entails that mental disorders supervene upon brain disorders). Some mental disorders are, we have claimed, constituted (at least in part) by psychological and/or sociocultural properties that extend well beyond what can plausibly be characterized by reference to brain states and properties alone. Such robustly psychological and/or socioculturally constituted disorders fail to supervene upon brain disorders, and thus we claim that mental disorders fail to supervene upon brain disorders in at least some cases. By contraposition, it is false that such disorders are brain disorders.

If, as we have argued, some mental disorders do not supervene upon brain disorders, it does not follow that neurobiological research into such mental disorders has no value. Far from it: understanding the neurobiology involved in mental disorders has the potential to inform and improve treatment and prevention. Our claim merely suggests that neuroscientific facts do not exhaust all of the facts about mental disorders. The failure of supervenience would show that that neurobiological understanding of mental disorders cannot replace wholesale the psychological, social, and cultural aspects of mental disorders, nor render them merely instrumental or pragmatic.

Furthermore, the psychological and sociocultural constitution of mental disorders that fail to supervene upon brain disorders undermines the current efforts to privilege neuroscientific research and intervention over psychosocial and behavioral interventions (Teachman et al., 2014). More than that, however, cases of supervenience failure may help to delineate the probable value of specific approaches to research and treatment. Because psychosocial and cultural relations to the external environment pose problems for supervenience theses, we propose that identifying cases of supervenience failure may represent a (sufficient, but perhaps not necessary) method for differentiating more robustly mental (as opposed to brain) disorders. For mental disorders where supervenience failure cannot be shown, more purely neurobiological approaches have a substantially higher likelihood of furthering research aimed at clarifying etiology and treatment approaches for those disorders. For mental disorders where claims for supervenience failure can be substantiated, it can be presumed that approaches involving psychological states, behavior, and sociocultural considerations will be necessary to develop adequate accounts of those disorders.

The upshot of this argument is that the only coherent way to retain the thesis that mental disorders are brain disorders is to eliminate all references to relations to the social and physical environments in which individuals are situated and to remove all normative language from characterizations of and diagnostic procedures for mental disorders. In other words, an adequate characterization of neural dysfunction in which individuals may manifest such dysfunction yet manifest no psychological symptoms of

that dysfunction seems to require a complete redescription of psychopathology that excludes *in principle* the experience and phenomenology of the individual. Under such a redescription, such notions as “sadness,” “loss of interest,” and “worry” would be rendered quaint hold-overs from folk psychology with no place in scientific psychiatry. Although those inclined toward reductionism or eliminative materialism may welcome such a conclusion, we regard this position as fundamentally misguided for the reason that, as we have already stated, psychological experiences and sociocultural relationships (at least) sometimes constitute rather than merely indicate mental disorder.

Consider an analogy: imagine a world not so different from ours where science is just getting under way. In this world, some proto-scientists study oceans and some study rivers. These scientists study, among other phenomena, tidal bores. Tidal bores occur when the leading edge of a wave traverses a river against the river current’s direction. Ocean-scientists insist that tidal bores are ocean-wave phenomena, while river-scientists insist that tidal bores are river-current phenomena. In actuality, each possesses half the story and none of the understanding, for a tidal bore constitutively requires the turbulent interplay between an ocean wave front and a river current. Analogously, we view it as at best only partially correct to say that mental disorders are brain disorders. Although mental disorders involve brain (mal-)functioning, mental disorders that fail to supervene upon brain disorders logically cannot reduce to or be identified with brain disorders, in the same way that tidal bores cannot reduce to or be identified with river currents. Although the matter deserves more extensive and careful treatment, we offer the suggestion that this sort of intertwining (in cases of supervenience failure) of individual-level neural configurations and the physical and social environment beyond the individual could provide ontological grounds for a détente between neuroscientific and humanistic perspectives on mental disorders.

If the argument of this paper is correct, then a dilemma faces those who would rebrand psychiatry as clinical neuroscience. Either the subject matter of psychiatry and related mental health disciplines is heterogeneous (in that at least some disorders cannot be characterized by neuroscience alone), or the subject matter of

psychiatry must be redescribed in such a way as to eliminate references to the psychological, social, and cultural aspects of mental disorders that render implausible the logical independence of mental disorders and their symptoms. We have argued against the latter redescription project, at least for some mental disorders.

We believe that a unitary, one-size-fits-all approach to psychiatry and psychiatric research ignores critical distinctions within the basic ontology of the discipline’s subject matter. Although some mental disorders may intimately relate to brain dysfunction, other mental disorders are more intrinsically mental, involving relational properties that hold between an individual and her physical and/or sociocultural environment. These latter, more robustly, mental disorders do not reside within the brain, in much the same way that Beethoven’s Moonlight Sonata does not reside within pianos, numbers do not reside within calculators, and laws do not reside within statute books. What we mean by “within” must be carefully delineated. With certain exceptions, perhaps, mental disorders reside not within us, but between us.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev. ed.). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Armstrong, D. M. (1968). *A materialist theory of the mind*. New York, NY: Routledge.
- Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, 69, 89–95. <http://dx.doi.org/10.1067/mcp.2001.113989>
- Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, 64, 1089–1108. <http://dx.doi.org/10.1002/jclp.20503>
- Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. <http://dx.doi.org/10.1146/annurev-clinpsy-050212-185608>
- Bosker, F. J., Hartman, C. A., Nolte, I. M., Prins, B. P., Terpstra, P., Posthuma, D., . . . Nolen, W. A. (2011). Poor replication of candidate genes for major depressive disorder using genome-wide as-

- sociation data. *Molecular Psychiatry*, 16, 516–532. <http://dx.doi.org/10.1038/mp.2010.38>
- Brülde, B., & Radovic, F. (2006). What is mental about mental disorder? *Philosophy, Psychiatry, & Psychology*, 13(2), 99–116. <http://dx.doi.org/10.1353/ppp.2007.0001>
- Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95, 3–45. <http://dx.doi.org/10.2307/2185131>
- Chalmers, D. (1996). *The conscious mind*. Oxford, UK: Oxford University Press.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78, 67–90. <http://dx.doi.org/10.2307/2025900>
- Cooper, R. (2004). What is wrong with the DSM? *History of Psychiatry*, 15, 5–25. <http://dx.doi.org/10.1177/0957154X04039343>
- Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet*, 381, 1371–1379. [http://dx.doi.org/10.1016/S0140-6736\(12\)62129-1](http://dx.doi.org/10.1016/S0140-6736(12)62129-1)
- Deacon, B. J. (2013). The biomedical model of mental disorder: A critical analysis of its validity, utility, and effects on psychotherapy research. *Clinical Psychology Review*, 33, 846–861. <http://dx.doi.org/10.1016/j.cpr.2012.09.007>
- Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28, 97–115. <http://dx.doi.org/10.1007/BF00485230>
- Ghaemi, S. N. (2013). Taking disease seriously in DSM. *World Psychiatry; Official Journal of the World Psychiatric Association (WPA)*, 12, 210–212. <http://dx.doi.org/10.1002/wps.20082>
- Glannon, W. (2002). Depression as a mind-body problem. *Philosophy, Psychiatry, & Psychology*, 9, 243–254. <http://dx.doi.org/10.1353/ppp.2003.0041>
- Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? Precisely. *Science*, 348, 499–500. <http://dx.doi.org/10.1126/science.aab2358>
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, 167, 748–751. <http://dx.doi.org/10.1176/appi.ajp.2010.09091379>
- Insel, T. R., & Quirion, R. (2005). Psychiatry as a clinical neuroscience discipline. *JAMA: Journal of the American Medical Association*, 294, 2221–2224. <http://dx.doi.org/10.1001/jama.294.17.2221>
- Kandel, E. R. (1998). A new intellectual framework for psychiatry. *American Journal of Psychiatry*, 155(4), 457–469. [http://dx.doi.org/10.1016/S0014-3855\(02\)00104-4](http://dx.doi.org/10.1016/S0014-3855(02)00104-4)
- Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17, 1174–1179. <http://dx.doi.org/10.1038/mp.2012.105>
- Kim, J. (1984). Concepts of supervenience. *Philosophy & Phenomenological Research*, 45, 153–176. <http://dx.doi.org/10.2307/2107423>
- Kim, J. (2005). *Physicalism, or something near enough*. Princeton, NJ: Princeton University Press.
- Klerman, G. L. (1978). The evolution of a scientific nosology. In J. C. Shershow (Ed.), *Schizophrenia: Science and practice* (pp. 99–121). Cambridge, MA: Harvard University Press.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Lakhan, S. E., & Kramer, A. (2009). Schizophrenia genomics and proteomics: Are we any closer to biomarker discovery? *Behavioral and Brain Functions*, 5, 2. <http://dx.doi.org/10.1186/1744-9081-5-2>
- McLaughlin, B. (1995). *Varieties of supervenience*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511663857.002>
- McLaughlin, B. P. (2006). Supervenience. *The Stanford encyclopedia of philosophy*. Retrieved from <http://stanford.library.usyd.edu.au/entries/supervenience/>
- Moore, G. E. (1903). *Principia ethica*. Cambridge, UK: Cambridge University Press.
- Murphy, D. (2006). *Psychiatry in the scientific image*. Cambridge, MA: MIT Press.
- Murphy, D., & Woolfolk, R. L. (2000). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, 7, 241–252.
- Nelson-Gray, R. O. (1991). DSM-IV: Empirical guidelines from psychometrics. *Journal of Abnormal Psychology*, 100, 308–315. <http://dx.doi.org/10.1037/0021-843X.100.3.308>
- Olbert, C. M., Gala, G. J., & Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: Theoretical framework and empirical application. *Journal of Abnormal Psychology*, 123, 452–462. <http://dx.doi.org/10.1037/a0036068>
- Patil, T., & Giordano, J. (2010). On the ontological assumptions of the medical model of psychiatry: Philosophical considerations and pragmatic tasks. *Philosophy, Ethics, and Humanities in Medicine*, 5, 3. <http://dx.doi.org/10.1186/1747-5341-5-3>
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind, and religion* (pp. 37–48). Pittsburgh, PA: Pittsburgh University Press.
- Radden, J. (1996). Philosophy of mind and mental concepts. *Current Opinion in Psychiatry*, 9,

- 364–367. <http://dx.doi.org/10.1097/00001504-199609000-00013>
- Redei, E. E., Andrus, B. M., Kwasny, M. J., Seok, J., Cai, X., Ho, J., & Mohr, D. C. (2014). Blood transcriptomic biomarkers in adult primary care patients with major depressive disorder undergoing cognitive behavioral therapy. *Translational Psychiatry*, 4, e442. <http://dx.doi.org/10.1038/tp.2014.66>
- Reynolds, C. F., III, Lewis, D. A., Detre, T., Schatzberg, A. F., & Kupfer, D. J. (2009). The future of psychiatry as clinical neuroscience. *Academic Medicine*, 84, 446–450. <http://dx.doi.org/10.1097/ACM.0b013e31819a8052>
- Smart, J. J. C. (1959). Sensations and brain processes. *The Philosophical Review*, 68, 141–156. <http://dx.doi.org/10.2307/2182164>
- Stein, D. J., Phillips, K. A., Bolton, D., Fulford, K. W., Sadler, J. Z., & Kendler, K. S. (2010). What is a mental/psychiatric disorder? From *DSM-IV* to *DSM-V*. *Psychological Medicine*, 40, 1759–1765. <http://dx.doi.org/10.1017/S0033291709992261>
- Teachman, B., Prinstein, M., McKay, D., Abramowitz, J., Kraut, A., DeRubeis, R., & Chambless, D. (2014). Society for a Science of Clinical Psychology Response to NIMH Strategic Plan. Retrieved June 4, 2015, from www.sscpweb.org/RecentAnnouncements/3167957
- Van Cleve, J. (1990). Supervenience and closure. *Philosophical Studies*, 58, 225–238. <http://dx.doi.org/10.1007/BF00368283>
- Van Oudenhove, L., & Cuypers, S. E. (2010). The philosophical “mind-body problem” and its relevance for the relationship between psychiatry and the neurosciences. *Perspectives in Biology and Medicine*, 53, 545–557. <http://dx.doi.org/10.1353/pbm.2010.0012>
- Wakefield, J. C. (1992). Disorder as harmful dysfunction: A conceptual critique of *DSM-III-R*'s definition of mental disorder. *Psychological Review*, 99, 232–247. <http://dx.doi.org/10.1037/0033-295X.99.2.232>
- Wakefield, J. C. (1995). Dysfunction as a value-free concept: A reply to Sadler and Agich. *Philosophy, Psychiatry, & Psychology*, 2, 233–246.
- Woodward, J. (2008). Cause and explanation in psychiatry. In K. S. Kendler & J. Parnas (Eds.), *Philosophical issues in psychiatry. Explanation, phenomenology, and nosology* (pp. 132–184). Baltimore, MD: John Hopkins University Press.
- Zachar, P., & Potter, N. N. (2010). Personality disorders: Moral or medical kinds—or both? *Philosophy, Psychiatry, & Psychology*, 17, 101–117. <http://dx.doi.org/10.1353/ppp.0.0290>

Received March 5, 2015

Revision received June 18, 2015

Accepted June 19, 2015 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!