Natural kinds, psychiatric classification, and the history of the DSM

Jonathan Y. Tsou

Department of Philosophy and Religious Studies, Iowa State University

ABSTRACT: This paper addresses philosophical issues concerning whether mental disorders are natural kinds and how the DSM should classify mental disorders. I argue that some mental disorders (e.g., schizophrenia, depression) are natural kinds in the sense that they are natural classes constituted by a set of stable biological mechanisms. I subsequently argue that a theoretical and causal approach to classification would provide a superior method for classifying natural kinds than the purely descriptive approach adopted by the DSM since DSM-III. My argument suggests that the DSM should classify natural kinds in order to provide predictively useful (i.e., projectable) diagnostic categories and that a causal approach to classification would provide a more promising method for formulating valid diagnostic categories.

Introduction

The first two editions of the *Diagnostic and Statistical Manual of Mental Disorders* (henceforth, DSM) adopted a theoretical and etiological approach to psychiatric classification insofar as mental disorders were classified according to psychoanalytic etiological theories (see APA, 1952, 1968). The publication of the third edition of the DSM (APA, 1980) marked American psychiatry's 'revolution' in psychiatric classification insofar as the authors of DSM-III self-consciously replaced the theoretical and etiological system of classification adopted in DSM-I and DSM-II with an atheoretical and purely descriptive ('neo-Kraepelinian') approach to classification, which has remained the favored methodology in subsequent editions of the DSM. While DSM-III has been widely heralded as a success in its provision of reliable diagnostic categories and its introduction of diagnostic criteria for defining disorders, a persistent problem with the DSM is its failure to provide valid diagnostic categories, which correspond to 'real' mental disorders in nature.

In this paper, I argue that the DSM's inability to classify naturally occurring disorders (i.e., natural kinds) is directly related to the purely descriptive approach for classifying mental disorders employed since DSM-III. Moreover, I argue that a theoretical and causal approach to classification based on biological psychiatry would offer more effective method for formulating valid diagnostic categories. The paper proceeds in two sections. In the first section, I argue that some mental disorders are natural kinds (i.e., mechanistic property cluster kinds) in the sense that members of a kind share a similar biological causal structure (i.e., members of a kind are constituted by a similar set of biological mechanisms). I present this view against Ian Hacking's argument that the objects of classification in the human sciences are inherently unstable because of the social feedback effects ('looping effects') of human science classifications. In the second section, I argue that the DSM should classify disorders that are natural kinds and that a theoretical and causal approach to classification would offer a superior approach for meeting this goal compared to the purely descriptive approach currently favored by the DSM. My argument emphasizes that diagnostic categories that individuate natural kinds will be projectable in the sense that they yield reliable inductive inferences about members of a kind. I suggest that the DSM's neo-Kraepelinian, purely descriptive approach to classification has been an impediment to formulating valid diagnostic categories that classify natural kinds because it has failed to provide a testable and revisable system of classification. On the contrary, diagnostic categories since DSM-III have tended to become further entrenched (and reified) over successive editions of the DSM, and diagnostic categories have guided research in a dogmatic manner, rather than standing in dynamic relationship with research on psychopathology. The theoretical approach that I advocate integrates the DSM's descriptive diagnostic categories with information about the causes of mental disorders. As an ideal, I propose that the operational criteria that define mental disorders should be associated with biological mechanisms that underwrite these signs. This would provide a more promising method for testing the validity of diagnostic categories.

Natural kinds in psychiatry

In the following section, I argue that some mental disorders classified by the DSM (e.g., schizophrenia, depression, bipolar disorder, obsessive compulsive disorder) are natural kinds insofar as members of a kind share a similar biological causal structure. In particular, some

mental disorders are *mechanistic property cluster kinds:* classes of abnormal behaviour that are constituted by a cluster of stable biological mechanisms at multiple levels (e.g., molecular, developmental, neurobiological) that interact to produce the stereotypical features of the kind.

Natural kinds, human kinds, and looping effects

In philosophy of science, the distinction between natural kinds and artificial kinds refers to distinction between natural classes that are discovered and conventional classes that are invented respectively. In this framework, natural kinds are *naturally-occurring classes of objects, properties, or processes that are discovered and exist independent of classifiers* (e.g., electrons, H_2O, fish). By contrast, artificial kinds are *conventional classes that are imposed on nature by and invented by classifiers* (e.g., triangles, money, vehicles). As such, natural kinds are assumed to be a natural (or 'real') feature of the world and the proper objects of scientific investigation and laws. For this reason, scientific classifications that individuate natural kinds are thought to 'carve nature at the joints'. Conversely, artificial kinds are understood as arbitrary classes that serve some useful social function for classifiers, but whose kindhood is not constituted by a natural basis.

While there are clear-cut cases of natural kinds in sciences such as physics and chemistry (e.g., fundamental physical particles and chemical elements), the existence of natural kinds in biology and the human sciences is much more controversial. A large reason for this difference concerns the relative *stability* of the kinds of entities classified in the biological and human sciences (Bird and Tobin, 2015). For instance, biological species, which were once considered to be paradigm cases of natural kinds, are subject to change over time due to evolutionary processes. The kinds of entities classified in the human sciences, such as psychology or sociology, are arguably even less stable since the objects of classification in these sciences (i.e., kinds of people) are aware of how they are classified and may *change* in response to how they are classified. Hence, compared to the immutability of the kinds studied in the natural sciences, the objects of classification in the human sciences appear to be far less stable and the particulars that are classified under a general classification are subject to much more variance.

With respect to the human sciences, Ian Hacking (1995b) has articulated an influential argument for thinking that there is a fundamental difference between the kinds of entities studied in the natural sciences and human sciences, which is premised on his account of the 'looping effects' of human science classifications. Looping effects are social feedback effects, wherein the meaning of a human science classification (e.g., 'depression', 'homosexuality', 'multiple personality') changes the experiences and behaviour of people who are so classified (e.g., a person diagnosed with major depression acts in accordance to the stereotypes associated with the depression label).[1] For Hacking, an important feature of classifications in the human sciences that distinguishes them from classifications in the natural sciences is that former possess looping effects whereas the latter do not. In one of his most developed accounts of this distinction, Hacking (1999) explicates this distinction in the language of 'interactive kinds' and 'indifferent kinds':

---

[1] For a more detailed exploration of how DSM categories can affect diagnosed individuals' understanding of themselves, see Tekin (2011, 2014).

(1) *Interactive kinds* (human kinds): Human science classifications (e.g., 'homosexuality', 'depression', 'autism') that possess looping effects and interact with their objects of classification.

(2) *Indifferent kinds* (natural kinds): Natural science classifications (e.g., 'lemon', 'quark', 'sulfur') that do not possess looping effects and do not interact with what they classify.

It is important to note that Hacking's distinction between interactive and indifferent kinds is intended to distinguish between two types of *classifications*, rather than two types of *objects of classification*. Moreover, Hacking aims to draws this distinction solely in terms of the presence of looping effects: interactive kinds are classifications that possess looping effects, while indifferent kinds are classifications that do not.[2]

Hacking has many interesting things to say about interactive kinds in connection with psychiatry and the human sciences. For my purposes, I want to highlight his claim that—*because of looping effects*—the objects (or 'targets') of human science classifications are inherently unstable. Hacking (1999: 108) writes:

[A] cardinal difference between the traditional natural and social sciences is that the classifications employed in the natural sciences are indifferent kinds, while those employed in the social sciences are mostly interactive kinds. *The targets of the natural sciences are stationary. Because of looping effects, the targets of the social sciences are on the move.* (emphasis added)

More recently, Hacking (2007: 293) has articulated this idea as follows:

We think of . . . kinds of people [e.g., obese people or impoverished people] . . . as definite classes defined by definite properties. . . . But it is not quite like that. *They are moving targets because our investigations interact with the targets themselves, and change them.* And since they are changed, they are not quite the same kind of people as before. The target has moved. That is the looping effect. (emphasis added)

In these passages, Hacking draws the radical conclusion that—*because of looping effects*—the objects classified by interactive kinds (i.e., kinds of people) are perpetually unstable entities. Because interactive kinds causally affect the kinds of people that they classify, the people so classified will *constantly be changing* in response to how they are classified and the associated classifications will need to be *constantly revised* in order to accommodate such changes. This is the sense in which Hacking contends that looping effects render their objects of classification *moving targets*.

In this paper, I argue against Hacking's conclusion that the objects of classification in the human sciences are inherently unstable because of looping effects. Hacking's inference from the

---

[2] While the initial motivation of Hacking's account of human kinds was to articulate a crisp demarcation criterion for distinguishing the natural sciences from the social sciences (see Martínez, 2009), he now regards his earlier efforts as wrongheaded, and eschews the terminology of 'human kinds', 'interactive kinds', and 'indifferent kinds' (Hacking, 2007: 291-294). Despite Hacking's rejection of his earlier views, a critical examination of his work is useful for clarifying some distinctions important for this paper.

presence of looping effects to the instability of objects of classification is a hasty generalization that fails to distinguish different possible consequences of looping effects (Tsou, 2007). Looping effects are a ubiquitous feature of almost all human science classifications: so long as classified people are *aware* of how they are being classified, their experiences and behaviours will change in response to looping effects. It is important to recognize, however, that the mere presence of looping effects will not necessarily render objects of human science classifications *unstable*. While an individual who is classified as 'depressed' will be invariably affected by how she is classified, the ways in which looping effects change the experiences and behaviour of classified individuals will not necessarily require revisions in the classification 'depression', or what it means to be 'clinically depressed'. For a revision of this kind to occur, something specific needs to happen: looping effects must *uniformly* change the experiences and behaviour of individuals who fall under a classification. Moreover, a *sufficient number* of individuals need to be uniformly changed by looping effects such that *the meaning of the classification* or *the criteria that constitute membership for that classification* must be revised. An example of this phenomenon can be found in Hacking's (1995a) fascinating historical study of multiple personality disorder. When this classification was introduced in the nineteenth century, the disorder was understood as a special case of hysteria. After a number of reported cases in the 1970s that drew causal connections between multiple personality disorder and childhood sexual abuse, this led to transformations in the classification. Hacking (1995a: 256) maintains that changes in the multiple personality classification were directly related to changes in the way that individuals classified as multiples understood themselves:

> The intimate relationship connecting recovery of traumatic memories of child abuse with multiple personality is no accident. . . . Dissociation is explained as coping mechanism. The multiple comes to understand that she is as she is now because of the way she deployed coping mechanisms in the past.

According to Hacking, by the mid-1980s, the classification 'multiple personality disorder' was *transformed* such that repeated child abuse was understood to be the primary cause of multiple personality disorder. In this case, looping effects changed the experiences and behaviours of individuals, which subsequently lead to revisions in the meaning of the classification.

While Hacking suggests that looping effects render objects of human science classifications unstable, he cannot draw this conclusion unless he can show that the consequences of looping are like those in the multiple personality example. Hacking fails to establish this point, and his analysis conceals the fact that the consequences of looping are often less drastic. In this paper, I argue that some mental disorders (e.g., schizophrenia, depression, bipolar disorder) are *stable objects of classification* in spite of the looping effects associated with the classifications of these kinds. These disorders can be identified in terms of a stereotypical cluster of psychological states and behaviours that are underwritten by a set of biological mechanisms. Insofar as the signs that define these disorders and biological mechanisms that underwrite them remain stable despite the presence of looping effects, they are natural kinds.

Some mental disorders are natural kinds: mechanisms, stability, and projectability

In articulating a theoretical account of mental disorders as natural kinds, I assume that the distinction between natural kinds and artificial kinds is best understood as a distinction of degree,

rather than a sharp distinction in kind (LaPorte, 2004: 21-27). The classes individuated by psychiatric classifications—like all scientific classifications—are inevitably theory-laden (and hence conventional) insofar as they are formulated relative to various human values, interests, and standpoints (Szasz, 1960; Scheff, 1963; Laing, 1967). This theory-ladeness, however, does not vitiate the fact that some classifications will pick out classes that have a *more* natural basis than others. John Dupré (1981: 82) has defended a similar view in his pluralistic account of natural kinds, which he calls 'promiscuous realism': 'The realism derives from the fact that there are many sameness relations that serve to distinguish classes … in ways that are relevant to various concerns; the promiscuity derives from the fact that none of these relations is privileged'. Dupré's view suggests that all natural kinds will include conventional aspects, but members of these kinds will be relevantly similar to one another with respect to some natural (or 'real') properties. For human classifications, some categories (e.g., 'introversion', 'breast cancer', 'Down syndrome') refer to natural kinds insofar as members of these classes can be individuated in terms of a distinctive set of natural properties. By contrast, other human categories (e.g., 'widows', 'liberals', 'plumbers') refer to artificial kinds insofar as they are purely conventional classes that are not associated with any distinctive natural basis. In the case of mental disorders that are natural kinds, my analysis emphasizes the biological mechanisms that are shared by members of a kind. On this view, classes of abnormal behaviour that fall on the 'more natural' end of the natural-artificial spectrum can—to a large extent—be individuated with reference to a set of natural (biological) properties.[3]

Some mental disorders (e.g., schizophrenia, depression, bipolar disorder) are natural kinds insofar as members of these classes share a similar biological causal structure in the sense that they are constituted by a similar set of biological mechanisms. This perspective draws upon Richard Boyd's (1991, 1999) homeostatic property cluster (HPC) theory of natural kinds. The HPC theory is a non-essentialist theory of natural kinds, in which members of a kind share a cluster of similar properties, but no property is essential for membership in a kind. According to Boyd (1999: 143-144), the key characteristics of HPC kinds are the following:

(1) There is a family (F) of properties that are contingently clustered in nature.
(2) Their co-occurrence is the result of what may be described as homeostasis: either the presence of some properties tends to favor the presence of others, or there are underlying mechanisms that tend to maintain the properties in F, or both.
(3) There is a kind term (t) that is applied to things in which the homeostatic clustering of most of the properties in F occurs.

Boyd's HPC theory offers a useful account of how members of psychiatric kinds share a similar biological causal structure.[4] Within this framework, the characteristic signs of a mental disorder can be understood as the cluster of properties (F), while its biological causes are the mechanisms referred to in (2) that determine these properties.

---

[3] For a more comprehensive discussion of natural kinds in psychiatry, see Haslam (1998, 2000, 2002), Zachar (2000a, 2000b, 2014), Cooper (2005, 2007), Samuels (2009), Beebee and Sabbarton-Leary (2010), Godman (2013), Tsou (2013), and Kincaid and Sullivan (2014).
[4] For discussion of Boyd's HPC theory in the context of psychopathology, see Cooper (2005, ch. 2; 2007, ch. 4), Murphy (2006, ch. 9), Craver (2009), Beebee and Sabbarton-Leary (2010), Kendler, Zachar, and Craver (2011), and Tsou (2013).

Drawing on Boyd's HPC theory, Kendler, Zachar, and Craver (2011) argue that natural kinds in psychiatry are best understood as *mechanistic property cluster* (MPC) *kinds*: kinds that are constituted by networks of biological mechanisms at multiple levels (e.g., molecular, developmental, neurobiological) that interact to produce the key features of the kind.[5] An important feature of the MPC account of natural kinds is that it can explain the *projectability* of natural kind terms (*t*)—i.e., it can account for how successful inferences about members of a kind can be made on the basis of induction—which is an important feature of any satisfactory account of natural kinds (see Goodman, 1955/1983; Quine, 1969; Boyd, 1985, 2010).[6] In the account of MPC kinds discussed herein, it is the shared (law-like) causal structure captured by a natural kind term (*t*) that allows for successful inductive inferences about members of a kind. In the case of mental disorders, this importantly includes inferences regarding the prognosis of a particular disorder and predictions concerning how an individual diagnosed with a disorder will respond to specific treatment interventions. Moreover, the shared biological causal structure of MPC kinds can account for the relative stability of such kinds over time. Because the characteristic signs of MPC kinds are constituted by stable ('homeostatic') networks of biological mechanisms, MPC kinds are *stable objects of classification*.

Some mental disorders are good examples of MPC kinds insofar as there is compelling evidence that their characteristic signs are underwritten by a set of stable biological mechanisms.[7] I assume that evidence is compelling when multiple lines of independent research support a common hypothesis (see Wimsatt, 1981, 2007, ch. 4; Franklin and Howson, 1984; Culp 1994, 1995; Staley, 2004; Wylie, 2002, ch. 14; Chang, 2004, ch. 1; Stegenga, 2009). According to this view, a causal hypothesis or theory is 'robust' (i.e., well-confirmed) when multiple lines of (at least partially) independent research all support a common result. For example, there is good evidence that the symptoms of schizophrenia are underwritten by stable neurobiological mechanisms. According to the 'dopamine hypothesis', excessive dopamine activity in the mesolimbic pathway—which projects from the midbrain in the ventral tegmental area to the nucleus accumbens—causes the 'positive symptoms' (viz., delusions, hallucinations) of schizophrenia (Kring et al., 2007: 363-365; Carlson, 2008: 460-461). Evidence for the dopamine hypothesis comes from a variety of independent pharmacological findings. The positive symptoms of schizophrenia can be reliably alleviated by anti-psychotic drugs, which reduce dopamine activity in the mesolimbic pathway. Moreover, in sufficiently large doses, drugs with opposite pharmacological effects (i.e., stimulant drugs), which increase dopamine activity in the same pathway, can *induce* the positive symptoms of schizophrenia, and this 'amphetamine psychosis' can be treated with anti-psychotics (McKim, 2007: 249-250). While this story is oversimplified and glosses over some complexities (see Tsou, 2012), there is strong

---

[5] In this article, mechanisms are understood as complex systems of entities and activities that are organized in a way to produce regular changes (see Bechtel and Richardson, 1993; Glennan, 1996, 2002; Machamer, Darden, and Craver, 2000; Craver and Darden, 2001; Machamer, 2004).

[6] For a more comprehensive discussion of the projectability of psychological and psychiatric categories, see Fodor (1974), Kim (1992), Block (1997), Griffiths (1997, 2004), Millikan (1999), Antony (2003), Zachar (2008), Godman (2013), Khalidi (2013), and Tsou (2013).

[7] While I discuss the signs of (more) natural kinds of mental disorders being associated with a set of biological mechanisms in this paper (see note 5), if one regards mechanisms as serving some evolutionary function (e.g., see Garson, 2013), it would be more accurate to say that these disorders are associated with predictable *disruptions* of biological mechanisms (cf. Craver, 2001, 2013). For general accounts of mental disorder that invoke dysfunction in naturally selected mechanisms, see Boorse (1976, 1977) and Wakefield (1992a, 1992b).

support for the hypothesis that some symptoms of schizophrenia are caused by excessive dopamine activity in the mesolimbic pathway (cf. Kendler and Schaffner, 2011). Similarly, multiple lines of research indicate that the core signs of depression are underwritten by stable neurobiological mechanisms. According to the 'monoamine hypothesis', depression is caused by deficient activity of monoamine neurotransmitters, especially serotonin and norepinephrine. Evidence for the monoamine hypothesis comes from research on pharmacological interventions with depressed patients. Antidepressant drugs (e.g., MAO inhibitors, tricyclic antidepressants, specific serotonin reuptake inhibitors) that can relieve the signs of depression are all monoamine agonists that increase the activity of monoamine neurotransmitters (McKim, 2007: 303-306). This hypothesis also receives support from the finding that monoamine antagonist drugs with opposite pharmacological effects as antidepressants (e.g., reserpine) can *induce* depressive symptoms in non-depressed individuals (Carlson, 2008: 476-477). Research indicates that antidepressants exert their influence primarily on serotonin and norepinephrine systems, with serotonin specifically implicated.[8]

Mental disorders that are MPC kinds, such as schizophrenia and depression, possess a distinctive and stable biological causal structure. These (more) natural kinds of disorders will be stable objects of classification *in spite of looping effects* precisely because they are constituted by a set of stable biological mechanisms (e.g., neurobiological mechanisms). The existence of MPC kinds in psychiatry is significant because classifications of these kinds will be projectable in the sense that they yield successful inductive inferences about members of a kind (e.g., individuals who exhibit psychotic symptoms will likely benefit from treatment with antipsychotic drugs). By contrast, other mental disorders (e.g., drapetomania, hysteria, narcissistic personality disorder) are artificial kinds insofar as the stereotypical signs that define membership of these classes are determined primarily by social mechanisms (e.g., social norms, the effects of labeling, role-adoption). Although social mechanisms can sometimes yield predictions about members of a kind (e.g., drapetomanics will attempt to flee from their slave-owners), they will not yield the kinds of *stable and robust predictions* that natural kinds underwritten by biological mechanisms will yield (cf. Steel, 2004). For this reason, we would not expect classifications of (more) artificial kinds to provide *reliable predictions in the long-term*. In the second half of this paper, I argue that the projectability of classifications of MPC kinds provides the rationale for constructing a causal system of classification that individuates natural (rather than artificial) kinds.

DSM and the classification of mental disorders

In the following section, I argue that the DSM ought to classify mental disorders that are natural kinds. This argument is premised on the principle that psychiatric diagnostic categories should be *projectable* if they are to play a substantial role in guiding treatment and facilitating research on

---

[8] For a more comprehensive discussion of the neurobiological basis of schizophrenia and depression, see Tsou (2012, 2013). Besides pharmacological research, I assume that other lines of research (e.g., epidemiology, genetics, cognitive science) support the hypothesis that schizophrenia and depression are natural kinds underwritten by distinctive biological mechanisms. For example, cross-cultural research demonstrating that there is a core cluster of symptoms exhibited by most patients of schizophrenia (WHO, 1973) and depression (Sartorious et at., 1983) across cultures is particularly relevant. Another valuable resource for identifying natural kinds are historical studies examining the stability of symptoms of mental disorders over long time periods (e.g., see Berrios, 1996).

mental disorders. I present my argument with reference to the historical development of the DSM in the second half of the twentieth century and, in particular, the debate concerning whether the DSM should adopt a purely descriptive or causal approach to psychiatric classification. Compared to purely descriptive approaches, I argue that a theoretical and causal approach to psychiatric classification would provide a more effective method for formulating valid diagnostic categories and offer a more readily testable system of classification.

The history of the DSM and the validity problem

Since the publication of DSM-III (APA, 1980), the DSM has explicitly endorsed a purely descriptive approach to psychiatric classification. This 'revolution' in the history of psychiatric classification resulted from disenchantment with the failure of DSM-I (APA, 1952) and DSM-II (APA, 1968) to provide clear and reliable definitions of mental disorders, which was an important goal in the context of the late 1970s for purposes of providing reimbursement for individuals seeking mental health treatment in the United States and for properly diagnosing patients who could benefit from emerging pharmacological drugs that could treat the symptoms of more severe mental disorders (Wilson, 1993; Mayes and Horwitz, 2005; Tsou, 2011). In response to these issues, the DSM-III taskforce self-consciously replaced the theoretical (i.e., psychoanalytic) and etiological approach to classification implicitly adopted in DSM-I and DSM-II with an 'atheoretical' and purely-descriptive approach to classification (see Tsou, 2011).

From a historical perspective, the shift to a purely descriptive approach to classification in DSM-III was in large part due to the shared neo-Kraepelinian outlook of the DSM-III taskforce chaired by Robert L. Spitzer. The 'neo-Kraepelinian' label was originally coined by Gerald Klerman (1978), who reported a Kraepelinian revival in psychiatry, which was accompanying the decline of Meyerian and Freudian psychoanalysis.[9] This movement followed the tradition of psychiatry championed by Emil Kraepelin, who articulated a biological approach to psychiatry that competed with the psychoanalytic approach that dominated American psychiatry in the early twentieth-century (Blashfield, 1984, ch. 1). In the psychoanalytic tradition, it was assumed that psychological mechanisms play a central role in the emergence of mental illness and the boundary between normality and abnormality is continuous (ranging from mild neuroses to severe psychoses). By contrast, the neo-Kraepelinians assumed that mental illnesses are *discrete biological diseases* and there is a clear boundary between normality and abnormality (Klerman, 1978). Given this assumption, the neo-Krapelinians maintained that it was crucial to distinguish the mentally ill from the well with *reliable and valid diagnostic categories*, which in psychoanalysis was secondary to the goal of elucidating the meaning of individuals' observed symptoms (Wilson, 1993). This desire for reliable and valid diagnostic categories motivated the neo-Kraepelinian belief that mental disorders should be classified on the basis of their observable symptoms, rather than speculative inferences regarding their causes (Mayes and Horwitz, 2005).

---

[9] For a more comprehensive discussion of the neo-Kraepelinian outlook of DSM-III, see Klerman (1978), Blashfield (1984), Wilson (1993), Compton and Guze (1995), Mayes and Horwitz (2005), Decker (2007, 2013), and Tsou (2011).

In the introduction to DSM-III, the neo-Kraepelinian, purely descriptive approach to classification was justified in terms of promoting widespread usage of the manual by clinicians of diverse theoretical orientations:

> The approach taken in DSM-III is atheoretical with regard to etiology . . . except for those disorders for which this is well established . . . *The major justification for the generally atheoretical approach . . . is that the inclusion of etiological theories would be an obstacle to use of the manual by clinicians of varying theoretical orientations* . . . Because DSM-III is generally atheoretical with regard to etiology, it attempts to describe comprehensively what the manifestations of the mental disorders are, and only rarely attempts to account for *how* the disturbances come about . . . This approach can be said to be 'descriptive' in that the definitions of the disorders generally consist of descriptions of the clinical features of the disorders. . . . at the lowest order of inference necessary to describe the characteristic features of the disorder. (APA, 1980: 7, emphasis added)

As indicated herein, one of the main motivations for shifting to a purely descriptive approach was pragmatic: to ensure that the DSM can be used by many mental health professionals who approach problems in psychopathology from different theoretical perspectives. This intended pragmatic goal of DSM-III was largely achieved insofar as the DSM only became a widely used—and increasingly entrenched—psychiatric manual after the publication of DSM-III. In terms of the problems motivating the shift to a purely descriptive approach, DSM-III was also successful in providing a manual with clear and reliable diagnostic categories in its introduction of the now familiar 'diagnostic criteria', which define mental disorders in terms of operationally-defined necessary and sufficient criteria (mainly behavioural or otherwise observable criteria). For these reasons, DSM-III has been heralded as a great success in psychiatry, especially in its provision of reliable diagnostic categories.

While the DSM-III and its successors were successful in ensuring that its diagnostic categories possess high inter-rater reliability, a fundamental problem with the DSM is its inability to ensure that its diagnostic categories are valid (Kendell, 1989; Kendell and Jablensky, 2003). While there is no agreed upon concept of validity in psychiatry, valid diagnostic categories are generally understood as classifications that individuate real natural phenomena, i.e., valid categories 'carve nature at the joints.'[10] While some theorists have argued that validity is best understood in terms of utility, I assume that these concepts are distinct, although it is important to recognize that valid diagnostic categories will be predictively useful (i.e., projectable), but not necessarily vice versa. Given the importance of making reliable predictions in psychiatry—and the difficulty in evaluating more general ideals of validity—my discussion of validity will be limited to predictive validity, viz., the extent that a diagnostic category can make reliable predictions (see Schaffner, 2012). From this perspective, the purely descriptive approach to classification favored by the DSM is disadvantageous insofar as its definitions of mental disorders do not yield reliable predictions, e.g., regarding the prognosis of a disorder or how a disorder should be treated.

---

[10] For a more comprehensive discussion of various proposed definitions of validity (e.g., construct, content, predictive), see Cronbach and Meehl (1955), Robins and Guze (1970), Kendler (1990), Kendell and Jablensky (2003), First et al. (2004), and Jablensky (2012).

The DSM's failure to formulate (predictively) valid diagnostic categories is directly related to the purely descriptive methodology favored by the DSM since DSM-III. While the neo-Kraepelinian, purely descriptive approach to classification has been useful for providing reliable diagnostic categories, this approach has been an impediment for individuating valid diagnostic categories. Because mental disorders are defined at a behavioural level, without consideration of the causes of disorders, there is no assurance that the disorders being classified are natural kinds, rather than artificial kinds that are constituted primarily by social mechanisms. A theoretical and causal approach to classification, by contrast, could individuate disorders by etiology (i.e., MPC kinds that are constituted by a distinctive set of biological mechanisms), which would provide a more effective and direct method for identifying valid diagnostic categories (Murphy, 2006). Significantly, the purely descriptive approach championed by the DSM does not lend itself to the *continual testing* of diagnostic categories across successive editions of the DSM. On the contrary, the DSM's purely descriptive approach has functioned in practice to *protect* its diagnostic categories from revision and to potentially *reify* artificial kinds with no natural basis (Hyman, 2010). Currently, the DSM guides research on psychopathology in a top-down manner by providing operational definitions of mental disorders, which researchers employ to select homogenous populations of patients to study. While this ensures that researchers working in different locales are studying disorders in a uniform manner, this approach offers a recipe for further entrenching current definitions of disorders, as opposed to identifying valid diagnostic categories. A theoretical and causal approach to classification, by contrast, would demand that the definitions of mental disorders included in the DSM have a natural basis and that the particular symptoms that define a disorder are constituted by natural (biological) mechanisms.

The gains in diagnostic reliability achieved by the DSM's purely descriptive approach to classification (viz., providing definitions of disorders that can be used in a widespread and uniform manner) do not outweigh the costs of failing to provide valid diagnostic categories. While the shift to an atheoretical and descriptive approach to classification in the context of the late 1970s was justified because it helped to remove suspect psychoanalytic etiological theories from the DSM, in our current situation, the DSM's atheoretical stance has served to isolate the manual from some of the most promising scientific theories on psychopathology, including genetics, neurobiology, and the cognitive sciences (Murphy, 2006, ch. 9; Hyman, 2007; Insel et al., 2010; Sanislow et al., 2010). This circumstance is particularly problematic given that one of the intended roles of the DSM is to reflect the best current scientific findings on psychopathology. Prior to the publication of DSM-5 (APA, 2013), members of the DSM-5 taskforce—chaired by David J. Kupfer—indicated that DSM-5 would address this problem by moving away from the DSM's purely descriptive approach to classification towards a theoretical approach informed by sciences such as genetics and the neurosciences:

> [L]imitations in the current [neo-Kraepelinian] diagnostic paradigm suggest that research exclusively focused on refining the DSM-defined syndromes may never be successful in uncovering their underlying etiologies. For that to happen, an as yet unknown paradigm shift may need to occur. . . [An] important goal is to transcend the limitations of the current DSM paradigm and to encourage a research agenda that goes beyond our current ways of thinking to attempt to integrate information from a wide variety of sources and technologies. (Kupfer, First, and Regier, 2002: xix).

The subsequent publication of DSM-5 (APA, 2013) would be disappointing for those hoping for a revolutionary paradigm shift away from the neo-Kraepelinian, purely descriptive approach to classification associated with DSM-III and its successors (the main discernible difference between DSM-5 and its predecessors is the greater use of dimensional measures). In the following section, I argue that a stronger shift towards a theoretical and casual approach would provide a more effective means for formulating valid diagnostic categories.

An argument for a causal approach to classification

While the published editions of DSM have been ambiguous on whether the DSM intends to classify natural kinds, there are reasons for thinking that it does. Prior to the publication of DSM-III, the authors of DSM-III planned to include a statement in the manual indicating that the mental disorders listed in the DSM are 'a subset of medical conditions' (see Spitzer, Sheehy and Endicott, 1977; Spitzer and Endicott, 1978), which would suggest that the DSM does aim to classify natural kinds. This statement, however, was ultimately not included in DSM-III due to protests by psychologists, social workers, and counsellors, who regarded this statement as a declaration that psychiatrists with medical training were solely responsible for the treatment of mental disorders (Mayes and Horwitz, 2005). The general definition of mental disorder offered in the DSM also indicates that the DSM aims to classify natural kinds: 'A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior *that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning*' (APA, 2013: 20, emphasis added; cf. APA, 2000: xxxi). This definition indicates that the mental disorders classified by the DSM are natural kinds insofar as these disorders are *caused by a dysfunction in some naturalistic (i.e., psychological, biological, or developmental) process*. Significantly, classes of behaviour that merely deviate from social norms—as opposed to naturalistic norms—are explicitly excluded from the DSM's definition of mental disorder (APA, 2013: 20). Despite this understanding, there are reasons for thinking that some of the disorders included in the DSM (e.g., narcissistic personality disorder, histrionic personality disorder) are artificial kinds, rather than natural kinds.

I submit that a theoretical and causal approach to classification would provide a more effective method for classifying natural kinds and formulating valid diagnostic categories than the atheoretical and purely descriptive approach to classification favored by the DSM (cf. Tabb, 2015). Compared to purely descriptive approaches, the distinguishing feature of the approach that I advocate is that diagnostic categories would *incorporate information about the causes of disorders*. In this regard, it is important to note that the theoretical and causal approach that I defend would not jettison the DSM's descriptive categories. Rather, it would integrate theoretical information about the causes of mental disorders into the DSM's descriptive diagnostic categories.[11] This hybrid approach to classification would aim to classify mental disorders with a

_____

[11] In his classic analysis of psychiatric classification, Carl Hempel (1965) argued that taxonomic systems ought to advance from initially descriptive stages to increasingly theoretical stages. For Hempel, moving beyond a purely descriptive stage is crucial in the development of a taxonomic system since the introduction of general laws or theoretical principles provides the basis for *explanation and prediction*. For the purposes of this paper, Hempel's analysis suggests that in order for the DSM to provide a *testable* system of classification, it needs to integrate theoretical information (e.g., regarding etiology) into its descriptive categories.

clear natural (i.e., biological) basis and ensure that the diagnostic criteria that define mental disorders have identifiable (or hypothesized) biological causes. Compared to purely descriptive approaches, the main advantages of this approach is that it would: (1) provide a clear methodological basis for formulating valid and projectable diagnostic categories, (2) offer an explicit basis for deciding when a disorder should be included in the DSM, and (3) engage the DSM more closely with a large body of scientific research on psychopathology (Tsou, 2015). Considering psychiatric classification as a long-term endeavour and an iterative process, a theoretical and causal approach to psychiatric classification would also offer a methodology that would facilitate the ongoing testing and revision of diagnostic categories across successive editions of the DSM. By contrast, the DSM's purely descriptive approach to classification has arguably functioned to entrench and reify its diagnostic categories and protect DSM categories from revision or removal from the manual.

The theoretical approach that I defend would not necessarily *include* the causes of mental disorders in its diagnostic categories (cf. Hyman, 2007); rather, it would *incorporate* information about causes by requiring—*as an ideal*—that the operational symptoms that constitute the diagnostic criteria of a disorder are caused—in whole or in part—by identifiable biological mechanisms. According to this ideal, in order to be included in the DSM, there needs to be evidence that there is a *stable and distinctive biological causal structure for classified disorders.* This stance can accommodate the fact that scientists often do not have knowledge of the biological causes of some disorders (e.g., bipolar disorder), although they do have evidence that these disorders are associated with a distinctive biological causal structure (e.g., dysregulation of serotonin systems). However, the DSM's diagnostic categories ought to be revised and refined over time such that the characteristic signs that are employed to define mental disorders are associated with identifiable biological mechanisms.

As an example of a DSM diagnostic category that possesses some (predictive) validity, consider the DSM's definition of schizophrenia, which requires that two or more of the following symptoms be present over a one-month period (APA, 2013: 99-100):

1. Delusions.
2. Hallucinations.
3. Disorganized speech (e.g., frequent derailment or incoherence).
4. Grossly disorganized or catatonic behavior.
5. Negative symptoms (i.e., diminished emotional expression or avolition).

At least one of the symptoms must be (1), (2), or (3).

The diagnostic criteria for the DSM's schizophrenia classification, while not ideal, are useful precisely because there is robust evidence that these particular symptoms are associated with identifiable biological causes. For example, neurobiological research indicates that delusions (1) and hallucinations (2)—the 'positive symptoms' of schizophrenia—are caused by excessive dopamine activity in the mesolimbic dopamine pathway, while 'negative symptoms' (5) are caused by deficient dopamine activity in the mesocortical pathway (see Tsou, 2012). For this reason, the DSM's definition of schizophrenia possesses validity inasmuch as the symptoms that

constitute this classification are associated with dysfunctions in naturalistic (i.e., neurobiological) processes.[12]

As an example of a DSM diagnostic category that lacks validity, consider the DSM's definition of histrionic personality disorder, which is understood as a pervasive pattern of excessive emotionality and attention seeking. Diagnosis of histrionic personality disorder requires five of the following criteria to be met (APA, 2013: 667):

1. Is uncomfortable in situations in which he or she is not the center of attention.
2. Interaction with others is often characterized by inappropriate sexually seductive or provocative behaviour.
3. Displays rapidly shifting and shallow expression of emotions.
4. Consistently uses physical appearance to draw attention to self.
5. Has a style of speech that is excessively impressionistic and lacking in detail.
6. Shows self-dramatization, theatricality, and exaggerated expression of emotion.
7. Is suggestible (i.e., easily influenced by others or circumstances).
8. Considers relationships to be more intimate than they actually are.

This diagnostic category is a paradigm instance of an artificial kind in the sense that it is a classification that individuates a class of abnormal behaviour that lacks a natural basis (Gould, 2011). We do not have good reasons for thinking that any of these symptoms—with the possible exception of (3)—are constituted by biological mechanisms. On the contrary, we have reasons to believe that the cluster of signs that constitute the histrionic personality disorder classification are entirely determined by social mechanisms (e.g., imitation of cultural stereotypes, adoption of a social role). Given this state-of-affairs, we should not expect this classification to yield *stable* inductive inferences about members of this class.

My analysis suggests that mental disorders should be classified at a level of generality such that the characteristic signs of disorders are associated with stable biological mechanisms. In this regard, it is important to notice that the characteristic signs that define histrionic personality disorder (e.g., uncomfortable when not the center of attention, using physical appearance to gain attention) are formulated at a level that is too specific to be associated with biological mechanisms. By contrast, the signs that constitute schizophrenia (e.g., delusions, hallucinations) are specified at a sufficiently general level such that the symptoms that make up the classification are associated with certain dysfunctional biological processes. The theoretical and causal approach to classification that I articulate recommends that mental disorders should be defined at *a level of generality that captures common biological mechanisms* that underwrite a particular disorder. For this reason, my analysis suggests that DSM disorders that only appear in specific cultures (e.g., anorexia nervosa appears primarily in Western cultures) are classified at a level that is too specific to individuate a natural kind. It would be more fruitful to classify such disorders at a more general level (e.g., as a specific instantiation of an anxiety disorder) such that classifications of disorders capture a common cluster of symptoms that appear across cultures.[13]

---

[12] The DSM's diagnostic criteria for schizophrenia could be improved by including 'cognitive impairments' (e.g., deficits in attention, memory, and executive functioning), which are theorised to be caused by dopamine dysfunction in the prefrontal cortex (Hyman and Fenton, 2003).

[13] With respect to the generality of DSM definitions, I have elsewhere argued (Tsou, 2015)—drawing on Lara Keuck's research (see Kutschenko, 2011)—that the DSM should be reconceived as an 'epistemic hub', i.e., a

My analysis also highlights the limited kind of *stability* that we can expect to find associated with natural kinds in psychiatry. The (more) natural kinds of disorders that I have argued ought to be classified by the DSM are MPC kinds whose characteristic signs are constituted by networks of stable biological mechanisms. What ensures that disorders remain stable—and that the classifications for these disorders are projectable—over time (and across cultures) is that there is a common cluster of characteristic signs associated with these disorders (identified at an appropriately general level of description) that are underwritten by a stable set of biological mechanisms. However, we should expect a certain range of variation in how these signs are expressed, both within and across cultures, given the causal effects of social mechanisms (cf. Marková and Berrios, 2015). For instance, cross-cultural research has indicated that there is a core cluster of symptoms that are observed in over three quarters of depressed patients across cultures: sadness, joylessness, anxiety, tension, lack of energy, decreased interest and concentration, and feelings of inadequacy and worthlessness (Sartorious et al., 1983: 92). There are also variations in how depression is expressed in certain cultures, e.g., depression is expressed more in terms of somatic complaints in non-Western cultures (see Kleinman, 1988, ch. 3; cf. Kirmayer, 2001). On the theoretical approach to classification outlined in this paper, mental disorders should be defined at a level of generality that captures the common cluster of signs that appear across cultures, which are assumed to reflect common biological mechanisms. According to this picture, there are two potential sources of variation in symptomatology for disorders that are natural kinds: (1) variations due to changes at the biological level, and (2) variations due to changes at the social level. The theoretical approach that I advocate is intended to account for (2), but not necessarily (1). While the biological mechanisms that constitute MPC kinds might themselves be subject to variation over long time-scales due to various evolutionary processes, for purposes of psychiatric classification, the biological mechanisms will be *sufficiently stable* over time to ensure projectable inferences about members of a kind (cf. Berrios, 1999).

Compared to the atheoretical and purely descriptive approach currently adopted by the DSM, the theoretical and causal approach to classification that I have articulated would provide a more effective means for formulating valid diagnostic categories and continually testing these categories. The requirement that the DSM classify MPC kinds amounts to the demand that particular disorders included in the DSM have a similar biological basis in the sense that the cluster of signs that define these disorders are constituted by networks of biological mechanisms. As suggested above, the symptoms that make up the diagnostic criteria for a disorder should be described at a level of generality such that these symptoms are associated (or hypothesized to be associated) with identifiable biological mechanisms. Moreover, the DSM should aim to classify disorders at a level of generality that will capture cross-cultural similarities in the symptomatology of disorders. This would help to transform the DSM's purely descriptive project of classification into a more empirical endeavour of attempting to identify the appropriate level of description for defining the symptoms of mental disorders and validating such symptoms by identifying their biological bases. This approach would provide a more concrete and transparent criterion (other than prevalence rates) for deciding when a disorder should be included in the

---

common reference point that allows for the exchange of information and integration of explanations among the various mental health professionals who use the manual. Because epistemic hubs are intended for widespread usage in different contexts, the DSM ought to provide definitions at a level of generality that allows them to be coordinated with alternative and more specific definitions of disorders.

DSM and how its diagnostic criteria should be revised. Rather than encouraging the further entrenchment of existing diagnostic categories, a theoretical and causal approach would facilitate the ongoing revision of both diagnostic categories and the diagnostic criteria that are used to define mental disorders based on the best available scientific theories of psychopathology.

Conclusion

Seen from a historical perspective, the DSM has taken the specific shape it currently has because of contingent historical factors. When the DSM-III revolutionized psychiatric classification in its adoption of a purely descriptive approach, it did so largely in response to the declining status of psychoanalysis in the historical context of late 1970s American psychiatry and in response to emerging societal needs in the United States. While this paradigm shift in the history of psychiatric classification was undoubtedly an instance of scientific progress in the DSM's provision of reliable diagnostic categories, in our current historical context, the failure of the DSM to move beyond this purely descriptive stage is a sign of scientific stagnation.

In this paper, I argued that the DSM should shift towards a theoretical approach to classification that incorporates information about the causes of mental disorders and classifies (more) natural kinds of disorders. Compared to the purely descriptive approach currently favored by the DSM, a theoretical and casual approach would provide a more promising method for formulating valid diagnostic categories, a more readily testable system of classification, and a more scientifically transparent approach for classifying mental disorders. Some authors have urged that the DSM is not prepared for a paradigm shift wherein the DSM's purely descriptive approach is replaced with an etiological approach (Kendler and First, 2010), while others have stressed the importance of conservatism in revising future editions of the DSM (Frances and Widiger, 2012). I am sympathetic to these concerns, and the ideal of psychiatric classification articulated in this paper can accommodate these worries. Because the theoretical approach that I have defended retains the DSM's descriptive approach (i.e., the use of diagnostic criteria to define disorders), theoretical information about the causes of mental disorders could be incorporated into existing diagnostic categories in a gradual and piecemeal fashion. When there is uncertainty concerning the validity or causes of a disorder, my analysis recommends retaining current definitions until there are compelling scientific reasons for making revisions. Regardless of whether a shift towards the theoretical and causal approach to classification that I advocate would constitute a paradigm-shifting revolution in the classification of mental disorders, it would undoubtedly mark a significant point of scientific progress by supplanting the DSM's antiquated neo-Kraepelinian approach.

References

Antony L (2003) Who's afraid of disjunctive properties? *Noûs* 37 (Supplement: *Philosophical Issues* 13): 1–21.

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

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

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

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

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

Bechtel W and Richardson RC (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

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

Berrios GE (1996) *The History of Mental Symptoms: Descriptive Psychopathology since the Nineteenth Century*. Cambridge: Cambridge University Press.

Berrios GE (1999) Classifications in psychiatry: a conceptual history. *Australian & New Zealand Journal of Psychiatry* 33: 145–160.

Bird A and Tobin E (2015) Natural kinds. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy* (Spring 2015 ed.); accessed (2015) at: http://plato.stanford.edu/archives/spr2015/entries/natural-kinds/

Blashfield RK (1984) *The Classification of Psychopathology: Neo-Kraepelinian and Quantitative Approaches*. New York: Plenum Press.

Block, N (1997) Anti-reductionism slaps back. *Noûs* 31 (Supplement: *Philosophical Perspectives* 11): 107–132.

Boorse C (1976) What a theory of mental health should be. *Journal for the Theory of Social Behavior* 6: 61–84.

Boorse C (1977) Health as a theoretical concept. *Philosophy of Science* 44: 542–573.

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

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

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

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

Carlson NR (2008) *Foundations of Physiological Psychology*, 7th edn. Boston: Allyn and Bacon.

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

Compton WM and Guze SB (1995) The neo-Kraepelinian revolution in psychiatric diagnosis. *European Archives of Psychiatric and Clinical Neuroscience* 245: 196–201.

Cooper R (2005) *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*. Dordrecht: Springer.

Cooper R (2007) *Psychiatry and Philosophy of Science*. Montreal and Kingston: McGill-Queen's University Press.

Craver CF (2001) Role functions, mechanism, and hierarchy. *Philosophy of Science* 68: 53–74.

Craver CF (2009) Mechanisms and natural kinds. *Philosophical Psychology* 22: 575–594.

Craver CF (2013) Functions and mechanisms: a perspectivalist view. In: Huneman P (ed) *Functions: Selection and Mechanisms*. Dordrecht: Springer, 133–158.

Craver CF and Darden L (2001) Discovering mechanisms in neurobiology: the case of spatial memory. In: Machamer PK, Grush R and McGlaughlin (eds) *Theory and Method in the Neurosciences*. Pittsburgh, PA: University of Pittsburgh Press, 112–137.

Cronbach LJ and Meehl PE (1955) Construct validity in psychological tests. *Psychological Bulletin* 52: 281–302.

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

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

Decker HS (2007) How Kraepelinian was Kraepelin? How Kraepelinian are the neo-Kraepelinians?—from Emil Kraepelin to the DSM-III. *History of Psychiatry* 18: 337–360.

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

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

First MB, Pincus HA, Levine JB, Williams JB, Ustun B and Peele R (2004) Clinical utility as a criterion for revising psychiatric diagnoses. *American Journal of Psychiatry* 161: 946–954.

Fodor JA (1974) Special sciences (or: The disunity of science as a working hypothesis). *Synthese* 28: 97–115.

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

Garson J (2013) The functional sense of mechanism. *Philosophy of Science* 80: 317–333.

Glennan S (1996) Mechanisms and the nature of causation. *Erkenntnis* 44: 49–71.

Glennan S (2002) Rethinking mechanistic explanation. *Philosophy of Science* 69: S342–S353.

Godman M (2013) Psychiatric disorders qua natural kinds: the case of the 'apathetic children'. *Biological Theory* 7: 144–152.

Goodman N (1955/1983) *Fact, Fiction, and Forecast*, 4th edn. Cambridge, MA: Harvard University Press.

Gould CS (2011) Why the histrionic personality disorder should not be in the DSM: a new taxonomic and moral analysis. *International Journal of Feminist Approaches to Bioethics* 4: 26–40.

Griffiths PE (1997) *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Griffiths PE (2004) Emotions as natural and normative kinds. *Philosophy of Science* 71 (Proceedings): 901–911.

Hacking I (1995a) *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton, NJ: Princeton University Press.

Hacking I (1995b) The looping effects of human kinds. In: Sperber D, Premack D, and Premack AJ (eds) *Causal Cognition: A Multidisciplinary Debate.* Oxford: Clarendon Press, 351–383.

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

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

Haslam N (1998) Natural kinds, human kinds, and essentialism. *Social Research* 65: 291–314.

Haslam N (2000) Psychiatric categories as natural kinds: essentialist thinking about mental disorder. *Social Research* 67: 1031–1058

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

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

Hyman SE (2007) Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience* 8: 725–732.

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

Hyman SE and Fenton WS (2003) What are the right targets for psychopharmacology? *Science* 299: 350–351.

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C and Wang P (2010) Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry* 167: 748–751.

Jablensky A (2012) The nosological entity in psychiatry: a historical illusion or a moving target? In: KS Kendler and J Parnas (eds) *Philosophical Issues in Psychiatry II: Nosology*. Oxford: Oxford University Press, 77–94.

Kendell RE (1989) Clinical validity. *Psychological Medicine* 19: 45–55.

Kendell R and Jablensky A (2003) Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry* 160: 4–12.

Kendler KS (1990) Toward a scientific psychiatric nosology: strengths and limitations. *Archives of General Psychiatry* 47: 969–973.

Kendler KS and Schaffner KF (2011) The dopamine hypothesis of schizophrenia: an historical and philosophical analysis. *Philosophy, Psychiatry, & Psychology* 18: 41–63.

Kendler KS, Zachar P and Craver, C (2011) What kinds of things are psychiatric disorders? *Psychological Medicine* 41: 1143–1150.

Khalidi MA (2013) *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press.

Kim J (1992) Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52: 1–26.

Kincaid H and Sullivan J (eds) (2014) *Classifying Psychopathology: Mental Kinds and Natural Kinds*. Cambridge, MA: MIT Press.

Kirmayer LJ (2001) Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of Clinical Psychiatry* 62 (Supplement 13): 22–28.

Kleinman A (1988) *Rethinking Psychiatry: From Cultural Category to Personal Experience*. New York: The Free Press.

Klerman GL (1978) The evolution of scientific nosology. In: Shewshow JC (ed) *Schizophrenia: Science and Practice.* Cambridge, MA: Harvard University Press, 99–121.

Kring AM, Davison G, Neal, JM and Johnson SL (2007) *Abnormal Psychology*, 10th edn. Hoboken, NJ: John Wiley & Sons.

Kupfer DJ, First MB and Regier DA (2002) Introduction. In: Kupfer DJ, First MB and Regier MA (eds) *A Research Agenda for DSM-V*. Washington, DC: American Psychiatric Association, xv–xxiii.

Kutschenko LK (2011) How to make sense of broadly applied medical classification systems: introducing epistemic hubs. *History and Philosophy of the Life Sciences* 33: 583–602.

Laing RD (1967) *The politics of experience*. New York: Ballantine.

LaPorte J (2004) *Natural Kinds and Conceptual Change*. Cambridge: University of Cambridge Press.

Machamer P (2004) Activities and causation: the metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science* 18: 27–39.

Machamer P, Darden L, and Craver CF (2000) Thinking about mechanisms. *Philosophy of Science* 67: 1–25.

Marková IS and Berrios GE (2015) Neuroimaging in psychiatry: epistemological considerations. In: Zachar P, Stoyanov D, Aragona M and Jablensky A (eds), *Alternative Perspetives on Psychiatric Validation: DSM, ICD, RDoC and Beyond*. Oxford: Oxford University Press, 112–127.

Martínez ML (2009) Ian Hacking's proposal for the distinction between natural and social sciences. *Philosophy of the Social Sciences* 39: 212–234.

Mayes R and Horwitz AV (2005) DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences* 41: 249–267.

McKim WA (2007) *Drugs and Behavior: An Introduction to Behavioral Pharmacology*, 6th edn. Upper Saddle River, NJ: Pearson Prentice Hall.

Millikan RG (1999) Historical kinds and the 'special sciences'. *Philosophical Studies* 95: 45–65.

Murphy D (2006) *Psychiatry in the Scientific Image*. Cambridge, MA: MIT Press.

Quine WV (1969) Natural kinds. In Quine WV, *Ontological Relativity and Other Esssays*. New York: Columbia University Press, 114–138.

Robins E and Guze SB (1970) Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry* 126: 983–987.

Samuels, R (2009) Delusion as a natural kind. In: Broome M and Bortolotti (eds) *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford: Oxford University Press, 49–82.

Sanislow CA, Pine DS, Quinn KJ, Kozak MJ, Garvey MA, Heinssen RK, Wang PS and Cuthbert BN (2010) Developing constructs for psychopathology research: Research Domain Criteria. *Journal of Abnormal Psychology* 119: 631–639.

Sartorius N, Davidian H, Ernberg G, Fenton FR, Fujii I, Gastpar M, Gulbinat W, Jablensky, A, Kielhotz P, Lehmann HE, Naraghi M, Shimizu M, Shinfuku N and Takahashi R (1983) *Depressive Disorders in Different Cultures: Report on the WHO Collaborative Study on Standardized Assessment of Depressive Disorders*. Geneva: World Health Organization.

Schaffner KF (2012) A philosophical overview of the problems of validity for psychiatric disorders. In: KS Kendler and J Parnas (eds) *Philosophical Issues in Psychiatry II: Nosology*. Oxford: Oxford University Press, 169–189.

Scheff TJ (1963) The role of the mentally ill and the dynamics of disorder: a research framework. *Sociometry* 26: 436–453.

Spitzer RL and Endicott J (1978) Medical and mental disorder: proposed definition and criteria. In: Spitzer RL and Klein DF (eds) *Critical Issues in Psychiatric Diganosis*. New York: Raven Press, 15–39.

Spitzer RL, Sheehy M and Endicott J (1977) DSM-III: guiding principles. In: Rakoff VM, Kedward HB (eds) *Psychiatric Diagnoses*. New York: Brunner/ Mazel, 1–24.

Staley KW (2004) Robust evidence and secure evidence claims. *Philosophy of Science* 71: 467–488.

Steel D (2004) Social mechanisms and causal inference. *Philosophy of the Social Sciences* 34: 55–78.

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

Szasz TS (1960) The myth of mental illness. *American Psychologist* 15: 113–115.

Tabb K (2015) Psychiatric progress and the assumption of diagnostic discrimination. *Philosophy of Science* 82.

Tekin Ş (2011) Self-concept through the diagnostic looking glass: narratives and mental disorder. *Philosophical Psychology* 24: 357–380.

Tekin Ş (2014) Self insight in the time of mood disorders: after the diagnosis, beyond the treatment. *Philosophy, Psychiatry, & Psychology* 21: 139–155.

Tsou JY (2007) Hacking on the looping effects of psychiatric classifications: what is an interactive and indifferent kind? *International Studies in the Philosophy of Science* 21: 329–344.

Tsou JY (2011) The importance of history for philosophy of psychiatry: the case of the DSM and psychiatric classification. *Journal of the Philosophy of History* 5: 446–470.

Tsou JY (2012) Intervention, causal reasoning, and the neurobiology of mental disorders: pharmacological drugs as experimental instruments. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 542–551.

Tsou JY (2013) Depression and suicide are natural kinds: implications for physician-assisted suicide. *International Journal of Law and Psychiatry* 36: 461–470.

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

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

Wakefield JC (1992b) The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47: 373–388.

WHO (1973) *The International Pilot Studies of Schizophrenia*. Geneva: World Health Organization.

Wilson, M (1993) DSM-III and the transformation of American psychiatry: a history. *American Journal of Psychiatry* 150: 399–410.

Wimsatt WC (1981) Robustness, reliability, and overdetermination. In: Brewer MB and Collins, BE (eds) *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass, 124–163.

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

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

Zachar P (2000a) Psychiatric disorders are not natural kinds. *Philosophy, Psychiatry, & Psychology* 7: 167–182.

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

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

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