

Philosophy of Psychiatry

First published Wed Jul 28, 2010; substantive revision Thu Jan 22, 2015

Philosophical discussions of mental illness fall into three families. First, there are topics that arise when we treat psychiatry as a special science and deal with it using the methods and concepts of philosophy of science. This includes discussion of such issues as explanation, reduction and classification. Second, there are conceptual issues that arise when we try to understand the very idea of mental illness and its ethical and experiential dimensions. Third, there are interactions between psychopathology and the philosophy of mind; philosophers have used clinical phenomena to illuminate issues in the philosophy of mind, and philosophical findings to try to understand mental illness. This entry will discuss issues in the philosophy of science and philosophy of mind that pertain to psychiatry.

- [1. Introduction](#)
 - [2. The Medical Model and its Implications](#)
 - [2.1 The Minimal Interpretation](#)
 - [2.2 The DSM Conception of Mental Illness And Its Critics.](#)
 - [2.3 The Strong Interpretation of the Medical Model](#)
 - [3. Explanation](#)
 - [3.1 Levels of Explanation](#)
 - [3.2 Disorders as Statistical Networks](#)
 - [3.3 From Histories to Causes](#)
 - [3.4 Causal Explanatory Strategies](#)
 - [3.5 Cross-Cultural Issues](#)
 - [3.6 Clinical Application](#)
 - [4. Philosophical Psychopathology](#)
 - [4.1 Varieties of Philosophical Psychopathology](#)
 - [4.2 Thought Insertion](#)
 - [4.3 Psychopaths and Moral Psychology](#)
 - [4.4 Multiple Personality](#)
 - [4.5 Delusions](#)
 - [Bibliography](#)
-

1. Introduction

Scholars and textbooks alike agree (though they might not like it) that psychiatry now adheres to the “medical model”, which advocates “the consistent application, in psychiatry, of modern medical thinking and methods” (Black 2005, 3) because psychopathology “represents the manifestations of disturbed function within a part of the body” (Guze 1992, 44) to wit, the brain. But what does it mean to adopt this view of psychiatry, and what difference does it make?

One might think that the medical model merely commits us to a brain-based view of mental illness with few implications for science, and it is true that clinical or scientific differences across practitioners seldom seem to have much to do with divergent interpretations of the medical model. However, many theorists have argued that our current diagnostic categories, as compiled into DSM-IV-T-R (American Psychiatric Association 2000: I refer to different editions of this work by its abbreviated title), are faulty because they are derived from observable variables rather than underlying physical pathologies. These theorists are sceptical about many existing psychiatric diagnoses, and not just on empirical grounds; they see DSM diagnoses as collections of symptoms rather seeing them as medicine understands diseases—in terms of destructive processes realized in bodily tissues. Genuine mental illnesses, on this view, are not just sets of co-occurring symptoms but destructive processes taking place in biological systems. Following Murphy (2009), this may be termed the strong interpretation of the medical model. In contrast, a minimal interpretation of the medical model thinks of mental disorders as collections of symptoms that occur together and unfold in characteristic ways, but it makes no commitments about the underlying causes of mental illness.

These two interpretations correspond to different ways of cashing out the medical model. They are set out in section 2. Explanation is discussed in section 3. The strong interpretation naturally suggests that psychiatry should embrace the practices of medical explanation. If disease is a pathological process in bodily systems, then there must be a way of understanding how such processes occur in the brain, and how they explain the clinically observable facts about mental illness. But it seems that the logic of the medical model does not force us to privilege any one level of explanation—notably, it does not commit us to explanations restricted to the resources of molecular biology. Psychiatry, then, is a multi-level science. But we often know very little about the mental illnesses that psychiatrists study, and much explanation in psychiatry involves case studies or narrative accounts that cite the characteristics of a disorder, rather than underlying systems. Theorists disagree over whether these should be seen as different kinds of explanation, or as rudimentary forms of a full causal explanation. Section 3 also discusses the difficulties raised by cross-cultural issues. Section 4 discusses some examples of specific mental illnesses that have been held to teach philosophical lessons.

2. The Medical Model and its Implications

The idea that psychiatry is a branch of medicine is not universal, but even if we stick to professed believers in the medical model with the same broad view of the subject, we find disagreements about how its core commitments should be understood. It may be helpful to distinguish *minimal* and *strong* interpretations. A minimal interpretation makes no commitments about the underlying physical structure that causes mental illness. The *strong* interpretation of the medical model, in contrast, dissents on just this issue. It says that the proper medical understanding of disease is in terms of morbid anatomy. It is committed to specific causal hypotheses in terms of abnormalities in underlying neurobiological systems. Minimalists treat diagnostic labels as useful heuristics rather than natural kind terms, whereas a strong interpretation commits psychiatry to a view of mental illness as a medical disease in the strongest sense, that of a pathogenic process unfolding in bodily systems.

2.1 The Minimal Interpretation

Minimalists employ a concept of disease as the observable, regular unfolding of a suite of symptoms. Kraepelin applied it to psychiatry as the basis for differential diagnosis, for example between dementia praecox (schizophrenia) and other forms of insanity (1899, 173–175). It is a familiar idea that the DSM's syndrome based conception of mental illnesses stands in the tradition of Kraepelin, who argued that “*only the overall picture of a medical case from the beginning to the end of its development can provide justification for its being linked with other observations of the same kind*” (1899, 3). This familiar neo-Kraepelinian picture is that mental illnesses are regularly co-occurring clusters of signs and symptoms that doubtless depend on physical processes but are not defined or classified in terms of those physical processes.

Guze (1992), for example, is a minimalist. When he advocates seeing psychiatry as a part of medicine what he means is that we can classify disorders in terms of their characteristic symptoms and courses. McHugh and Slavney (1998) seem to agree. They say that a disease “is a construct that conceptualizes a constellation of signs and symptoms as due to an underlying biological pathology, mechanism and cause (302).” To diagnose a patient as suffering from a mental disorder, for McHugh and Slavney (48), is to label them in a way that we judge helpful as a starting point for investigating physical processes.

McHugh and Slavney deny that a disease is a physical process—“its essence is conceptual and inferential” (48). As a precursor of this view, they cite Thomas Sydenham, the great seventeenth century English physician who distinguished varieties of pox based on their characteristic courses and outcomes. This way of thinking about diseases prescind from theorizing about underlying causes of a disease entity in favor of a concentration on observable phenomena, not hidden causes. However, it is diseases understood as destructive processes that make up the taxa of medicine. Similarly, the successive editions

of the DSM have sought to classify mental illnesses based on course and symptoms but not specific causes.

However, Sydenham has been dead for over three hundred years, and medicine moves on. If psychiatry adheres to a conception of disease that looks like Sydenham's it may depart from the rest of medicine. Many scholars have argued (e.g. Bynum 1994, Carter 2003, Whitbeck 1977) that in the nineteenth century medicine changed in many ways, including a shift in thinking about disease. Instead of seeing disease as the identification of symptoms, it was the linkage of symptoms to distinct pathologies, understood as departures from normal function, that came to define disease. Heinrichs (2001, 271), for example, insists that psychiatry needs to employ this basic medical outlook, and therefore needs a developed theory of how the mind/brain works that can be used to identify psychological abnormalities and explain how they arise. Andreasen (2001, 172–76) embraces the strong interpretation. She argues that psychiatry now emerging as a form of cognitive neuroscience, is for the first time able to identify the specific pathophysologies that underpin the symptoms of mental illness. This development fits the pattern suggested by Hempel (1965), who expected psychiatry, as it matured, to develop a theory-driven classification scheme based on an understanding of the objective nature of mental illness rather than just pragmatic factors (Hempel 1965; See Fulford et. al 2006: 324–341: Thornton 2008: 169–74).

2.2 The DSM Conception of Mental Illness And Its Critics.

The DSM treats mental disorders as syndromes along the lines of the minimal medical model. Individuals who share a DSM diagnosis have a subset of symptoms in common, often drawn from a larger list so that although some people may have all their symptoms in common, there may be no overlap at all in other cases. These collections of symptoms are also supposed to unfold over time in more or less the same way, once we make due allowances for some individual variation. They are also expected to respond to treatment in the same way. The previous version of the DSM assumed that each diagnosis represented malfunction in some mental, physical or behavioural trait or capacity (DSM-IV-TR, xxi). However, the diagnoses were listed without worrying about what that underlying malfunction might be, and in most cases there was (and remains) no agreement about what causes what. DSM-5 (American Psychiatric Association 2013) defines mental disorders (p. 20) as syndromes comprising clinically significant disturbances of cognition, emotion or behaviour that reflect underlying dysfunctions. These are normally associated with distress or social disability, but cannot be diagnosed if the behaviour is culturally normal or merely socially deviant, unless it reflects a dysfunction.

DSM diagnoses are usually taken to reflect a conventional concept of disease, despite the talk of underlying malfunction, because of their emphasis on syndromes and their lack of any causal hypotheses. In the manner of the minimal medical model, they let us make use of descriptive and statistical reasoning and offer the hope of accurate prediction and

effective control. In the opinion of some philosophers, that is all science ever aspires to. I began by citing a textbook declaration that psychiatry uses normal medical methods, and there is certainly nothing within McHugh and Slavney's discussion of mental illness that makes scientific methods inapplicable.

Other philosophers, though, will tell you that the job of science is to discover the causal structure of the world. Furthermore, there are plenty of students of psychopathology who argue that the neglect of causal structure in psychopathology is getting in the way of science (Poland, Von Eckardt and Spaulding 1994, Murphy 2006, Gerrans 2014). They worry that we are lumping together different groups of people based on behavioral evidence alone, when in fact their observable similarities mask important underlying differences, including differences between normal people who are troubled but basically healthy, and their pathological counterparts. From this philosophical-cum-clinical perspective it looks like dereliction of duty when any scientific discipline, but especially psychiatry, decides to isolate itself from inquiry into causes. Suppose we ask why the conventional view of diseases is successful in organizing prediction and control (what psychiatrists sometimes call “predictively valid”). The obvious answer is that successful predictive categories are actually tracking the causal structure in the world that accounts for the shared nature of the different instances, so we should aim for a stronger validation of the diagnosis than mere predictive validity. Whilst the reliability of a diagnosis is a gauge of agreement across measurements and observers, validity is supposed to be about what is really there. It looks obvious that even the most expert observers could all agree but nonetheless be wrong, whereas proper validation reassures us that we are measuring something that is really there. (Psychiatrists employ the concept of validity in several ways; for help in sorting them out, see Fulford et al. (2006: 318–19), Schaffner (2012) and the essays in Zachar et al. 2014.)

Someone committed to a strong version of the medical model sees validating a diagnosis as understanding its underlying causal structure: a diagnosis is valid if it rests on a biological process that can be identified by experiment and observation using the methods of the biological and cognitive sciences. Any approach to psychiatry that looks to science to validate its categories in this way must meet at least two conceptual challenges. First, there is a metaphysical challenge, which is that a concept of validity tied to the uncovering of neurobiological processes seemingly adopts a realist philosophical position, thinking that science can tell us how the world is really put together. Many philosophers think that realism is a wholly unnecessary and unwarranted metaphysical commitment, and that all science can really tell us about is a set of relationships among the data. These relationships let us make predictions and exert some control over nature, but do not tell us what is really out there. Philosophical disputes over realism have often centred on the notion of the unobservable. Electrons are a paradigm example, but other posits of fundamental physics play the same role. In psychiatry the situation is different; theories and constructs employ latent variables rather than reference to unobservables. Schizophrenia, as a construct, is not a further type of stuff that lies behind the appearances and explains them: it is constituted by visible phenomena. We are not looking for an unobservable, but asking which observable phenomena (neurological, genetic, developmental, social) explain the symptoms. If the strong conception of the medical model is philosophically realist, it is so

in light of its commitment to causes, not to unobservables. We may look for a hidden variable that explains the observable phenomena, or for some other hypothetical underlying cause. Gold and Gold (2014, p.163-66) for example, suggest that many delusions and other psychotic symptoms may reflect a malfunctioning “Suspicion System” that has evolved to detect and monitor social threats. If such a system exists it is not directly observable, but its components should be open to observation and manipulation via the standard methods of the cognitive sciences. Psychiatry, like many sciences, deals in hidden variables or hypothetical causal structures, but it is not helpful to think of these as unobservable in the traditional sense.

The second conceptual challenge to the realist interpretation of validity is normative. The challenge is that there is an important sense in which diagnoses cannot be validated at all, if by “validation” we mean “shown to be a real disorder”. All validation can do is show that a pattern of behaviour deemed to be clinically significant depends on a physical process. Whether that pattern of behaviour is really pathological - rather than immoral or harmlessly odd- is another matter. The second challenge probably is very hard to meet. It requires that judgements of pathology be like findings of positive charge, i.e. scientifically grounded, rather than judgements of ugliness, i.e human responses. If so, there has to be some natural fact of the matter about whether some physical system is dysfunctional. If this cannot be done, then predictions about physical states can be validated, but disorders cannot be.

I noted above that the DSM approach is often called “neo-Kraepelinian”. But Kraepelin did not adopt it as a first choice. He saw classification by clinical description as an interim measure. It satisfies the practical requirements of physicians and establishes a productive heuristic for subsequent pathological and etiological inquiry, since “the value of every diagnosis is thus rated essentially by the extent to which it opens up reliable prospects for the future” (1899, 4). Kraepelin's preferred basis for classification and inquiry actually rested on his less well-remembered belief that “pathological anatomy promises to provide the safest foundation” for classification of mental illness in a mature psychiatry (1899, 2). He considered the correct taxonomy would be one in which clinical description, etiology and pathophysiology coincided: “cases arising from the same causes would always have to present the same symptoms and the same post-mortem result” (3).

There is a substantial difference between thinking of clinically-based, syndromic classification in this way and thinking of it as the DSM does. The DSM classification has been designed (DSM-IV-TR, xxiii) to improve communication between psychiatrists and across disciplines and provide a basis for education. But it is not advertised as the jumping-off point for a mature system of causally organised classification and practice. This reflects a minimal interpretation of the medical model; it can guide empirical research but not uncover causal structure.

DSM leaves itself open to the charge that a classification which groups together observable phenomena can sort entities into heterogeneous classes, with underlying diversity masked by surface similarities. We already know from other areas of medicine that what looks like the same phenomenon—a cough, say, or a sore throat, or chest pain—can reflect different biological processes and hence may really signify different conditions. A taxonomy that

relies content with surface features risks lumping different conditions together and keeping related ones apart.

Here are some examples of this charge in the recent literature (for others, see Murphy 2006 (chapters 1–3) and McNally, forthcoming). Ross et al. (2008, 33–34, 200–01) suggest some revisions to the current diagnostic category of Pathological Gambling (PG) (see DSM-IV-TR, 671–74). The category of pathological gambling is designed to operationalize a concept whose content is, roughly, gambling that causes problems for the gambler. But Ross et al. reason as follows: the existing category “maps relatively smoothly onto the class of people who exhibit classic patterns of addictive behaviour as a result of a specific kind of dysfunction in their dopaminergic reward system and consequent neuroadaptation impairing frontal control circuits. We therefore believe that clarity would be served ... by replacing the behaviourally derived concept of PG with the neuroscientifically anchored concept of addictive gambling (200)”.

A judgment that the DSM-5 categories have insufficient validity has led the National Institute of Mental Health to introduce the Research Domain Criteria (RDoC) in grant proposals (Insel et al. 2010). The originators of the RDoC acknowledge that the system entrenched by previous versions of the DSM has increased diagnostic reliability. But they worry that it is too detached from the nature of mental illness, which they conceive of as disorders of brain circuits: though the domains of inquiry themselves are to be identified in psychological terms. For example, rather than existing categories such as depression, the RDoC envisages a domain of negative valence which will include symptoms of anxiety or depression that might come to be seen as expressions of an underlying disorder of specific brain circuits (Sanislow et al 2010) . These disorders could be studied at many levels and need not be identified with simple lesions. But future models of mental illness are expected, on this vision, to draw on psychological, neurological, and genetic mechanisms, as well as information about the wider cultural context. It is expected that current research and future diagnosis will have to be based on these models in order to be properly warranted by the underlying facts about the causal structure of mental illness, rather than on clinical signs and symptoms. Clearly, in this case, validating a diagnosis is thought of as understanding its underlying causal structure: a diagnosis is valid if it rests on a biological process that can be identified by experiment and observation using the methods of the biological and cognitive sciences.

Ross and his colleagues argued along similar lines in their discussion of gambling, which may be seen as an example of the sort of work the RDoC foresees. They think there are two benefits to the conceptual change that they advocate. First, it would provide a clearer and principled demarcation of the addicted gambler from the “problem gambler” which is currently somewhat arbitrary (201). They also show the way to a better understanding of the nature of the addictive disorders more generally via the use of their preferred approach, which combines neuroscience and behavioural economics. They draw on prior research (ch. 6.2) to suggest that addictive gambling resembles dependence on stimulants (like cocaine) more than it does alcoholism, and hence enlarges our understanding of addiction more fully than purely behavioural criteria would do. The worry is that a behavioural

approach misses the similarities and differences between forms of addiction by treating all as more or less the same, based on shared behavioural and phenomenological effects.

To make these arguments we have to distinguish the truly pathological gambler, whose life is dominated (and typically wrecked) by gambling, from merely habitual gamblers. Habitual gamblers might engage in gambling more than they feel they should, or periodically lose more than they can afford, but there is a qualitative difference between this population and the population of genuinely addicted gamblers. This qualitative difference is detected by observing behaviour and rating the results of questionnaires. However, if we classify gamblers according to quantitative measures alone, we miss the reason why different groups behave differently. The reason is a causal process. The brains of the real addicts don't work like the brains of the merely troubled; thinking of addiction merely in terms of habitual behaviours will not let us make the discriminations among gamblers that neuroscientific evidence detects.

A stronger but similar argument was made by Horwitz and Wakefield (2007). They claim that only a hypothesis about causes distinguishes a normal population from a disordered one. They begin with the DSM criteria for diagnosing major depression. To receive a diagnosis of Major Depressive Episode, a subject must suffer from five of the following nine symptoms in a two week period (including either or both of depressed mood or diminished interest or pleasure in almost all activities): depressed mood; diminished interest; weight gain or loss (without dieting) or change in appetite; insomnia or excessive sleep; observable psychomotor agitation or retardation; fatigue or loss of energy; feeling worthless or excessively guilty; diminished ability to think or concentrate, or indecisiveness; recurrent thoughts of death or suicide or a planned or attempted suicide.

Obviously many ordinary episodes in a person's life can cause behaviors or feelings from this list, such as losing one's job. The traditional concept of melancholy or depression, which has a continuous history going back to classical antiquity (Radden 2000), allowed that life's tragedies cause melancholy with perfect reason. That traditional view of melancholy sees pathology only where one's circumstances do not justify a normal melancholy response: "pathological depression is an exaggerated form of a normal human emotional response" (Horwitz and Wakefield 2007, 71).

The DSM, however, ignores this tradition. Anyone who fits the syndrome receives the diagnosis regardless of how their life is going. The sole exception stipulates that grief following bereavement does not count towards diagnosis of Major Depressive Disorder until two months have elapsed, after which one may diagnose major depression. But not even that qualification is made if you show depressive symptoms after your intended jilts you at the altar or you fail all your exams or get your legs blown off. In subsequent work Horwitz and Wakefield (2012) have extended their analysis to the anxiety disorders. They argue both that normal human anxieties are being pathologized and that the diagnostic criteria for anxiety and depression have been much too sharply distinguished from those for depression, leading to an inability to explain why depression and anxiety so often co-occur.

During the run-up to DSM-5 the Mood Disorders Work Group proposed eliminating the bereavement exclusion. This met considerable professional opposition and attracted much adverse public comment (see Zachar 2014, ch. 10 for a summary of the debate and a discussion of the issues). From the standpoint of Horwitz and Wakefield, the proposal appeared to be pathologizing a normal human reaction. However, it was never suggested that grieving itself was pathological. Rather, the proposal was based on the idea that stressful life events often precipitate episodes that meet the criteria for a major depressive episode, and there was no reason to exempt bereavement. Many people, after the death of a loved one, exhibit symptoms very like those of major depression, and are at risk for subsequent episodes. From this perspective, the elimination of the bereavement exclusion is justified by science, regardless of common sense (Kendler et al 2007; Lamb et al 2010). Why should bereavement alone be exempted from the general claim that depression is caused by stressful life events?

Horwitz and Wakefield suggest that genuine depression is caused by a dysfunction on the part of a loss-response system that evolved to respond to losses that threaten to deprive us of reproductive resources. In depression the system misfires. It causes a loss-response in situations where it should not. Whether or not that is convincing, the argument that Horwitz and Wakefield make about the conflation of depression and sadness stands on its own merits. They make the case that the DSM conception of depression respects neither folk psychological common sense nor previous psychiatric consensus about depression. They conclude that the concept of depression defined by this diagnostic syndrome represents a major conceptual break. Our use of the new concept, they suggest, means that many people who are just normally miserable for understandable and ordinary reasons are being diagnosed as depressed. This leads to unnecessary alarm about an epidemic of depression as well as carrying unfortunate consequences for many people who are wrongly diagnosed. However, it is certainly open to those theorists opposed to the bereavement exclusion to insist that if depression is the result of an underlying system that misfires, it could misfire in response to bereavement. The proposal is not that everyone who grieves should be called depressed, but that there are reactions to bereavement that are severe enough to meet the criteria for major depression and should be classed as such. (For a more detailed guide to the debate see Zachar 2014, Ch. 10.)

The DSM concept of depression, on the face of it, lumps together different psychological and behavioral types in the same category. It does this because it neglects the diverse etiologies that may produce similar signs and symptoms. There is no way to answer this charge, or to arrive at a satisfactory taxonomy that mirrors that of general medicine, unless we adopt a causal foundation for nosology as suggested by psychiatrists who urge the adoption of what I have called the strong medical model. Indeed, without an account of underlying causal structure, it becomes difficult to give a principled answer to the question why some conditions should be seen as pathological to begin with. Merely saying that members of a class have certain properties does not tell us whether, let alone why, they are mentally ill. To answer that question we need some basis for making the judgment. The DSM, as we have seen, states that we have such a basis; mental disorder is an outward sign of dysfunction. But to make that argument, we need a commitment to finding dysfunction

somewhere on the basis of the outward, observable phenomena. No amount of observation alone will uncover the dysfunction without a commitment to the idea that there is a destructive neuropsychological process at work that causes an underlying dysfunction. In other words, the minimal model can only answer questions about the relationship between membership in a taxon and some other variable: it will not explain why the taxon has the nature that it has, or why it should be seen as a kind of disorder.

Doing that involves not just the empirical study of mental disorder, but a strong interpretation of the medical model, with its commitment to a view of disease as not just a syndrome but a destructive process. It is certainly possible to collect data using the DSM concept of depression, especially information about epidemiology and natural history. In that sense normal medical reasoning can indeed be employed, but there much medical reasoning is directed at trying to understand causes. It is here that we come to the strong interpretation, with its commitment to the idea that a mental disorder, like any other disease, is not just a construct to guide inquiry, but a genuine part of the causal structure of the biological world—a destructive entity with a distinctive pathophysiology which explains the observable phenomena that the minimal model is content with.

2.3 The Strong Interpretation of the Medical Model

Proponents of the strong interpretation, then, can agree with Guze that there are distinct mental illnesses with characteristic symptoms and natural histories. But they go further. If psychiatry is really a branch of medicine, they suspect, we should see the specific causal hypotheses emerge about mechanisms that cause the symptoms of mental illness. Both Andreasen and Guze assume that to find the specific causes of disease we need a background theory of normal function. Psychopathology is to be identified as the departure of a psychological system from its proper state. (We should note what counts as a malfunction is often contested in psychology and neuroscience. Furthermore, the presence of a malfunction does not suffice for mental illness. Most scholars agree that our thoughts about the nature of disease are sensitive to biological abnormalities but are also guided by judgments that such abnormalities impede human flourishing (See Cooper 2007, Murphy 2006, Sadler 2004, Wakefield 1992). The medical model also sees abnormalities or dysfunctions as necessary but not sufficient for mental illness. The broader issues involved will be put aside here. Our question is: how far can psychiatry adopt medical reasoning about pathogenic neurobiological processes?)

The strong interpretation seeks explanations that cite pathogenic processes in brain systems, just as bodily diseases are explained by processes in other organs. The process at issue need not be destructive, in the sense that it makes the system collapse. Bolton and Hill (2004, 252) note that many mental illnesses seem to be the outcome of systems in a poorly regulated state that is stable, albeit suboptimal. But, as they say, the same is true of hypertension and Cushing's disease; the idea of a specific pathogenic process in medicine includes dysregulation.

Heinrich and Andreasen assume that since the brain is an information processing organ the obvious background theory is cognitive neuroscience. Other thinkers have argued that the correct background theory for psychiatry is molecular biology (Kandel 2005, ch.2).

Bolton and Hill worry that whatever background theory the medical model employs, it will require explanations in terms of biological rather than intentional phenomena. They conclude that biological psychiatry can only accommodate disorders in which “intentionality has run out” (2004, 256). If true, this rules out using the medical model to explain the many psychopathologies that appear to involve irreducibly intentional or meaningful processes (cf. Thornton 2007, ch 4). The worry is founded on the assumption that there is an unbridgeable distinction between understanding a phenomenon in intentional terms and providing a causal explanation of it. But the brain is a cognitive organ—and, indeed, a social one. Many disciplines study the effects of healthy cognition on behaviour, and there is no reason to expect that cognition will suddenly become irrelevant to the explanation of behaviour when we study mental illness. For instance, Kendler and Prescott (2006, 148ff) found that “stressful life events” are among the chief causes of depression, that they are especially depressogenic if they involve experiences of humiliation (160), and that humiliation cannot be given a molecular reduction (350). Kendler and Prescott call Major Depression one of the oldest diagnoses in medicine (52) and they are committed to finding specific etiologies. Yet they are happy to employ “classically mental constructs” (350) as parts of their causal models of mental illness.

As presented so far, the strong interpretation of the medical model includes a demand for understanding biological systems in terms of their normal function if we are to understand why they become abnormal. Kincaid (2008, 375) has argued that a demand like that is unreasonable, because we can conduct research on depression (his example), as we do medical research more generally, based on “partial and unsystematic” causal understanding. Kincaid identifies the demand for a background theory with the search for a “complete wiring diagram of the organism from fertilization to maturity” (377) and he identifies that search with a conception of science as a search for laws of nature and natural kinds. But these seem to be separable: a background theory could comprise a number of partial wiring diagrams arrived at by piecemeal investigation, nor does it have to be driven by a desire for explanation in terms of kinds participating in laws. The questions are, what sort of general orientation toward mental phenomena does the medical model permit, and what sorts of explanations can it encompass? We do not need a complete theory of how biological systems work in order to conduct fruitful research. The issue is whether, at a minimum, there need to be some rough theoretical frameworks and shared empirical claims within which otherwise abnormalities can be identified and outcomes assessed.

First, nothing in the strong interpretation of the medical model rules out explanations that cite cognitive processes in brains: Andreasen (1997) argues that schizophrenia and depression are cognitive pathologies and sees psychiatry as both a form of medicine and “the discipline within cognitive neuroscience that integrates information from all these related disciplines in order to develop models that explain the cognitive dysfunctions of psychiatric patients based on knowledge of normal mind/brain function” (1586). This

conception of psychiatry is fully biological but not committed to a particular metaphysical interpretation of what biological explanation must look like. It sees mental illness as the result of pathogenic processes taking place in brain systems but does not force us to choose reductive explanations as a matter of logic. So Bolton and Hill's worry about the compatibility of the medical with their "neural encoding" theory (2005, ch.2) seems overstated. They are right to stress that the brain can be seen as a socially situated, cognitive organ, but there are advocates of a strong medical model who are equally attuned to this conception of neuroscience. Existing medical and biological practice already cites processes that cross levels of explanation (Schaffner 1993, 1994) and incorporate environmental information about stress or other risk factors. There is in principle no objection to adding an intentional level to the mix. Any conception of normal neurological functioning must take information-processing into account, because processing information is what brains are for.

Most mental illnesses are caused by diverse environmental and genetic factors. We have little reason to think that all the relevant causes can be given a reductive analysis, for as Schaffner dryly points out, to specify and measure environmental variables in purely molecular terms "would be a very long-term project" (1994, 287). But one might still think that molecular processes are the root causes of mental illness even if many other contributory factors exist and define the context within which genes work. Kandel (2005) offers a very stringent version of this view, which assigns all explanatory significance to proximal genetic causes on the grounds that "all of 'nurture' is ultimately expressed as 'nature' (39)" He means that behavioral and cultural factors influence gene expression, but he also seems to think that this implies that psychiatric explanations need only mention gene expression, which will screen off everything else.

Kandel's genetic reductionism is a metaphysical position aimed at identifying a privileged level in nature and a general explanatory strategy that unifies the phenomena and offers lawlike explanations. However, in medicine the goal of causal explanation is instrumental. We seek the factors that make a difference in a given case, not a metaphysically privileged level or a general unifying account. There is little reason so far to believe that the reductonistic approach Kandel urges has been more fruitful.

In studying eating disorders, for example, we find that social factors may explain particular epidemiological patterns, like different levels of eating disorder across populations. But social factors don't tell us why only one girl in a family gets bulimia. To explain that we can look at a particular brain chemistry that puts her at risk (Steiger et al 2001). But that does not establish that neurobiology really is fundamental. Rather, nothing is fundamental. Culture or trauma may require biological mediation to have its pathological effect, and vice-versa. Neither one should be considered fundamental, but seen rather as the best way to answer a specific question about what makes a difference in a context.

The identification of a cause of a disease often depends on what we want to control. The reason for this, as Whitbeck (1977) and Kincaid (2008) have convincingly argued, is that physicians care about causes for instrumental reasons. The goal is to understand why a biological system has departed from a functional state and identify components and

processes within the system that can be manipulated so as to understand the abnormal character of the system, and, ideally, mitigate it. In looking for the relevant factors we do not seek a final description of the system or look for the ultimate sources of an abnormal or disvalued outcome. Rather, we hope to identify the variables that made a difference to the outcome.

That is why alternative etiologies can complement, rather than compete with, each other, as in Meehl's (1977) example of an elderly man, genetically somewhat at risk for depression, who develops it after his spouse dies. It seems wrong to call this a genetically caused episode of depression: the genes needed an unusual environment, so that in most other circumstances the man would not be depressed. But he would not have developed clinical depression had his genes been different. Both factors are causally relevant.

Because of this stress on figuring out what makes a difference in a context, medicine is a natural home to manipulationist accounts of causation, as Whitbeck also saw (1977, 627–28). The basic idea behind the manipulationist approach is that it puts us in a position to begin explaining a phenomenon “when we have identified factors or conditions such that manipulations or changes in those factors or conditions will produce changes in the outcome being explained” (Woodward 2003, 10). We may have a correlation—say that between smoking and heart disease. We get from there to a causal explanation if we can intervene so as to manipulate the value of the causal variable by itself—by stopping one group from smoking and letting an otherwise identical group smoke, for example. If smoking causes heart disease, we will see an effect on the cancer variable that reflects our meddling with the smokers. If smoking doesn't cause heart disease then there will not be a correlation between the two variables. (This view is compatible with many proposals about the underlying metaphysics. There will be more on this account of causation in section 3.1. For the technical details see Spirtes et al 2000, Pearl 2009).

Kandel's avowed molecular reductionism dissents from the goal-driven, manipulationist view. Kandel bets that molecular explanation picks out nature's own preferred level. He assumes that as we learn more we will be able to supply fundamental explanations at such a level. Kandel invites us to give up a fruitful approach in favor of a promissory note, on which we have scant reason to expect payment. Molecular explanations do not appear to be flourishing alone among all possible approaches to mental illness. (Schaffner (1998) points out how hard it is to supply reductive explanations even for simple model organisms.)

Things may turn out as Kandel expects, but it is not clear that they will, based on the current fortunes of different kinds of psychiatric explanation. In any event, it seems that the medical model does not require any particular level of explanation. Rather, it seems to be marked, in its strong interpretation, by a search for explanations of a particular form that can involve many different levels. These are explanations that show how symptoms of mental illness systematically depend on disruptions to the normal functioning of biological and cognitive processes. The strong interpretation of the medical model bets on such explanations being available and—which is a further claim—being clinically useful when they are finally uncovered. Whether the ensuing explanations privilege one form of explanation over others

is an empirical question to be answered case by case. However, psychiatry currently recognizes many different causal factors and we should expect that to continue. One way of cashing out this idea is to say that psychiatry must pay attention to different levels of explanation. The next section, on explanation, will begin with a discussion of this point.

3. Explanation

3.1 Levels of Explanation

I begin with a picture that is familiar to philosophers of psychology, Marr's (1982, 24–5) distinction between three levels of explanation in cognitive science. The highest specifies the computational task accomplished by the system of interest. It tells us what the goal of the system is, specified in terms of what it computes. The middle level describes the actual representations and algorithms that carry out the goal. The lowest level tells us how brain tissue or other material substrate, such as the parts of a machine, can implement the algorithm. Note that the levels of description are different perspectives on the same mental process - Marr thinks of vision as the construction of a 3D representation from a 2D input, and the three levels are three different ways to understand that construction. Other cognitive phenomena can also be understood in these three ways.

Marr sought to understand psychological capacities in the abstract without worrying about biology: although he certainly assumed that capacities could be realised in a physical set up, he thought that the details of the particular set up would not make a difference to the way the psychology was understood. But as the sciences of the mind have become more interested in the brain we are learning that this idea may need to be amended. Cognitive neuroscience is learning to look for physical structures in the brain that carry out information-processing jobs, or what Glymour (1992) calls “cognitive parts”. Our understanding of biological realization feeds back into and constrains our understanding of the abstract processes of cognition (Bechtel and Richardson 1993). Rather than relating abstract psychological capacities to each other, higher level descriptions of the mind/brain in cognitive neuroscience relate capacities understood as functional descriptions of fairly coarse-grained brain areas, not purely abstract computational task specifications. Theories of the normal system specify the nature and interrelation of cognitive parts. Armed with these assumptions about cognitive parts, psychiatrists can see a disorder as a breakdown in normal functioning within and between cognitive parts. Marr's first level might seem to be useful as a source of background theory for doing psychiatry, because it asks “what is it for?” and much of psychiatry is about finding out psychological systems fail to do what they are for.

However, we still face the question of how psychological deficits are related to other levels of explanation, and here Marr's picture comes under real stress. Marr's three levels are different representations of the same process. But in psychiatry different levels typically represent distinct causal processes. For example, the causes of many mental illnesses include a mix of genes and environmental factors, as in Meehl's example, cited above, of

genetics and bereavement conspiring to cause depression. The genes and environmental effects are different kinds of processes, not different levels of one process. If we were dealing with one process, describable in different ways, then we could anticipate a reductive account, in which higher-level variables are mapped on to lower-level ones. But even though it is hard enough to imagine a molecular or neurological reduction of a psychological construct like humiliation (which, we saw above, is depressogenic), it is even harder to imagine a reductive analysis of socio-cultural factors like unemployment or childhood sexual abuse. They have brain effects, but the brain effects vary across classes of individuals in ways that depend on other environmental and genetic contexts (see Kendler and Prescott 2006 for a comprehensive review.)

Schaffner (1993, 1994) suggests this structure is the norm throughout biology, and sees it as relating causes and effects at different levels of explanation. Appealing to levels of explanation is unobjectionable if it just involves a reminder that we need to relate variables of many sorts. But it is not clear that we have any principled grounds for sorting phenomena into levels. Marr did, because he imagined levels as descriptions of the same process (the construction of a 3D image from 2D retinal impacts) couched in the vocabularies of different sciences. But when we move outside the skull and begin introducing environmental factors and other kinds of cause, the Marrian picture looks less plausible.

Another way in which levels talk has featured in recent philosophy of psychiatry is as a constraint on permissible kinds of explanation. Chris Frith (1992, 26) for example, imagines what we would say if we found a strong correlation between a spike in dopamine levels and delusions of thought insertion (in which, for example, one has the experience of another person's thoughts in one's own mind). He argues that such a correlation is “simply not admissible” as an explanation because we would remain in the dark about why a sudden upsurge in dopamine causes thought insertion rather than something else. To show why thought insertion comes about, Frith thinks, we need a cognitive story which shows us how it is caused by failures of normal subpersonal processing.

Frith's idea is that a causal explanation of a psychotic phenomenon, or some other mental symptom, must cite properly cognitive variables to be explanatory. That is, explanations should relate variables from the same level. The relationship between dopamine and thought is just opaque to reason, and we want variables that are related in ways that are transparent to reason. There is a constraint here; the idea that explanations of psychological phenomena should relate cause and effect in ways that we find intelligible. Campbell (2008) in attacking this picture, notes that talk of levels of explanation is often wheeled in at this point, as it is by Frith, to supply relations that are transparent to reason in the right way. The idea is that only processes at a level let us grasp explananda at that level. Explanantia at a different level are not intelligible in the right way, so although lower-level processes may realise the higher ones they cannot explain them. This amounts to a constraint on acceptable explanation. Should we accept it?

The debate here quickly gets into technical arguments about the concept of causation. Campbell argues against Frith in favour of the Humean idea that we simply cannot tell in advance of inquiry what causal relations obtain in nature. Anything can cause anything and

it is just as unwarranted to suppose that the relevant explanatory variables for psychotic or affective phenomena must be cognitive as it is to adopt a simple reductionism and insist on their being biological. Decades of research have revealed that mental illness is caused by, and causes in turn, many different physical, psychological and social factors. We simply have to take our causal relations where we find them, including interlevel ones. Craver and Bechtel (2007, 554) on the other hand, argue that it stretches the concept of causation to breaking point to admit interlevel causes: they say that “to accept interlevel relationships as causal violates many of the central ideas associated with the concept of causation”. Craver and Bechtel argue from the perspective of mechanistic explanation. Mechanisms explain effects in terms of interlocking parts, and the relation across levels, they affirm, is one of constitution, not causation; causation can only be intra-level. Interlevel causation, on this view, amounts to something causing itself, because different levels are different ways of talking about the same thing. However, talk of levels in psychiatry often involves relating different parts of the world, as in Meehl's example: genetics and bereavement may be seen as requiring different levels, but they are not different ways of describing the same process. Causal models must interrelate phenomena of very different types.

In part because of these issues, that levels in psychiatry are not Marrian, Campbell (2006) argues for an interventionist approach to causation. This is the view that when we say *X* is a cause of *Y* we are saying that intervening on *X* is a way of intervening on *Y* (Woodward and Hitchcock 2003, Woodward 2003, Pearl 2000): manipulating one variable makes a difference to another. This is not a reductive analysis of causation, since it makes use of causal ideas—it just states that questions about whether *X* causes *Y* are questions about what would happen to *Y* if there were an intervention on *X*. Kendler and Campbell (2009) have argued that an interventionist model provides a rigorous way of articulating the idea that any combination of variables might characterize the causes of a disorder, whilst at the same time providing a clear test of what variables are actually involved, thus avoiding a simple-minded holism that just says that lots of things are relevant. They further note (2009, 997) that interventionism “permits the clear separation of causal effects from the mechanistic instantiations of those effects”, thus directly confuting the approach favored by Bechtel and Craver.

Appealing to levels of explanation in psychiatry, then, might just be a reminder that we need to relate variables of many sorts in explaining the causes of disorder. But it may also be connected with particular perspectives on reduction, explanation and causation. We won't go into these debates here; rather, we will move on to discuss a further problem. As well as great diversity in causal factors, psychiatry also confronts the problem of diverse symptoms; many mental illnesses take a different form in different people. Although two people might both be diagnosed with major depression, they need share few symptoms. Both are very likely to feel miserable and lose interest in activities they normally enjoy. However, one might lose weight, sleep less and becomes physically agitated while the other sleeps more and becomes lethargic and tormented by feelings of guilt and thought of suicide. The philosophical problem arises from the confrontation between the great variation in clinical reality and the need to simplify in order to render that reality scientifically tractable. In summary form, the problem is; what are we trying to explain

when we explain a mental illness? Is it a diverse set of real-world phenomena, or some simplified representation of those phenomena?

3.2 Disorders as Statistical Networks

One way to start thinking about what we have to explain in psychiatry is to borrow from Thagard's (1999, 114–5) account of diseases as networks of “statistically based causal relations” which are discovered using epidemiological and experimental methods. This fits the picture that Kendler and Prescott present for depression quite neatly (although their models are designed to incorporate both causal relations between variables and correlations). Thagard's causal-statistical networks, like Kendler and Prescott's path models, provide “a kind of narrative explanation of why a person becomes sick” (p15). They incorporate information about the typical course of a disease as it unfolds over time, including information about typical risk factors for the disease—such as the finding that heavy use of aspirin increases acid secretion which makes a duodenal ulcer more likely. The causal network does not specify how each causal factor produces its effects, though. It is really a descriptive model that lets us ask the question; what facts make it true that people get sick in these ways?

In effect we have a set of exception-prone generalizations about the pathways a disorder takes: people in this situation are likely to become depressed unless such and such intervenes. And when they are depressed they will probably have the following experiences, unless they have these other ones. Thagard's idea of a narrative is helpful here; path models represent typical stories about characteristic ways of getting sick. But a narrative by itself might not explain anything; if it is explanatory, it is because earlier events mentioned in the narrative pick out the causes of later events.

Rachel Cooper (2007) presents a way of thinking about psychiatric explanation that fits well with the idea that a diagnosis lets us tell a story. We often find ourselves with narratives that tell us how someone gets sick, and it might be that a useful narrative is one which isolates the temporally prior variables and lets us see how later phenomena depend on them. One way to think about this is that a good narrative is one that talks about a natural kind. If we have a natural kind, with some causal structure underpinning it, the systematic dependence of later states on earlier ones is easy to grasp. Several theorists have defended a view of psychiatric disorders as natural kinds. Influential recent philosophical treatments have rejected the idea that mental disorders have essences, but have defended the idea that mental disorder are natural kinds in Boyd's sense of homeostatic property clusters (Beebe and Sabbarton-Leary 2010; Kendler et al. 2010; Parnas et al. 2010; Samuels 2009, Zachar 2014), in which properties occur together often enough, and reliably enough, to support ampliative inferences, even if the mechanisms responsible for the kind do not make every member of it the same as every other member. When inductive reasoning works, thinks Boyd, it works because we have latched onto these underlying causal mechanisms that bring about the clustering. Successful inductions track the observable manifestations of the causal mechanisms responsible for the characteristics of the kind. The mechanisms,

we might say, leave an identifiable causal signature in the world. The HPC account seems to suit mental disorders. They often vary across cases, with characteristic symptoms sometimes absent even when the patient presents a typical cluster of signs and symptoms. So it is understandable that theorists who are sympathetic to the medical model in psychiatry should be drawn to the homeostatic property cluster idea.

3.3 From Histories to Causes

Cooper distinguishes natural-history based explanations from case studies. She presents them as distinct kinds of psychiatric explanation, but she doesn't go into detail about what the other types might be, so her overall position is hard to grasp. But one idea is very clear; at least some mental illnesses are natural kinds, and natural history explanations work by “invoking natural kinds” (2007, 47). Once we know which kind an object belongs to, we can explain and predict its behaviour based on its kind membership: we can say why a substance has expanded upon being heated by invoking the fact that it's a bit of metal, and metals expand when heated. Or we can predict that if Rex is in the kitchen then leaving the roast beef there unattended is a bad idea, since Rex is a dog and dog-meat interactions take a predictable form. And, it is hoped, we can explain why Laura hears voices by appealing to the fact that she has schizophrenia. And our confidence that she is schizophrenic warrants (pessimistic) predictions about her future.

Cooper also talks about case histories and argues that they explain another person's unique situation by drawing on our capacities for simulation of other people's states of mind. This claim involves commitments in the explanation of theory of mind that would take this essay too far astray. The relation of folk psychology to case histories is unexplored, as is the nature of reasoning from cases more generally: we lack a good general account of what a case history explains or is otherwise good for, in psychiatry or in any other field. (Forrester (1996) points out that reasoning with cases seems to be a distinctive and ubiquitous style of thought in science and elsewhere, and we have no good account of it.) So for now, we must put case histories aside.

Cooper argues that a natural history explanation needs to cite natural kinds if it is to be helpful. But in order to work the explanations do not need to cite the actual causal mechanisms responsible for the behaviour of the members of the kind in question. We know that Rex will devour the roast if we leave him in the kitchen with it, and we can safely make this prediction despite total ignorance of canine physiology. The history of psychiatry offers a number of instances in which pharmacology has led to the recognition of new kinds even in the absence of detailed causal knowledge. For example Cade's (1949) discovery that lithium greatly inhibited the startle reflex and other forms of anxiety in guinea pigs was useful even without good information about their underlying cognitive neurobiology, since it led him to predict that whatever was going on might also apply in humans.

Invoking kinds give us power over the world, by making it easier to predict, control or mitigate the outcomes we care about. One philosophical question we can ask, of course, is

whether science wants anything more. If you think science just aims for empirical adequacy, then it may seem that sorting patients into kinds is enough, provided that we can understand the course of the disease.

In particular, natural history explanations fit a conception of diseases as syndromes unfolding over time that is enshrined in both DSM-IV-TR and the minimal construal of the medical model. It does not deny that there are causal processes, but sees mental illnesses as collections of signs and symptoms with characteristic histories. They doubtless depend on physical processes but are not defined or classified in physical terms. This picture of disease naturally goes along with the employment of natural history explanations, since they too invoke natural kinds without worrying about the mechanisms that explain the behaviour of members of the kind. It's enough to be able to identify kind membership, thereby giving us a degree of explanatory and predictive control.

Natural history explanations fit Thagard's disease network nicely; if a disease network provides a narrative account of why people get ill, and predicts some outcomes, we can see it as a graphical representation of the historical pathway typical of a kind. Much epidemiological work in psychiatry fits the pattern here; we try to isolate the causal factors that define a disease history that is met with regularly. These histories display relations between variables that show how later phenomena, like the symptoms of depression or substance abuse, depend on earlier phenomena, like the genes you were born with, or the parental abuse you endured.

Unlike some of the more homespun examples, the natural histories uncovered by systematic longitudinal research can be quite surprising and genuinely explanatory. But (as Cooper and Thagard both suggest) they leave us with a lot to do; whatever the philosophers of science may say about how empirical adequacy is good enough, physicians have spent a century and a half looking for causal explanations of pathological phenomena (Carter 2003).

3.4 Causal Explanatory Strategies

Cooper suggests (174n2) that natural history explanations provide us with what Murphy (2006) calls causal discrimination, as opposed to causal understanding. We can know two kinds are causally different even if we don't why they are different, because the details of the underlying causal structure evade us. Different types of plant may need to be put in the ground at different times, or in different seasons, in order to maximize crop yield, for instance, and we may be able to predict accurately that different patients will respond to different drugs even if the basis for these differences remain unknown. But we assume that these differences rest on an underlying causal structure. In psychiatry, differential diagnosis follows a similar logic, but sceptics wonder if the degree of variation makes the predictions too unreliable for this approach to really work.

An explanation of this type places entities in kinds, thereby tracking the as yet unknown causal factors that regularly make the same kind of difference in patients of the same type.

Rather than identifying the causal factors, then, the natural history approach tracks the operation of hitherto unknown causes, by sorting patients into kinds with a presumed common nature, based on the natural history of their conditions, or on other grounds that I have not talked about, like differential responses to drugs.

Our complete explanatory job is to explain the observed causal-statistical network by identifying the mechanisms and other forces that give rise to it, viz. the causal structure of disorders. The problem, of course, is the sheer variety of the phenomena, which does not give us the degree of epistemic security we get when we sort phenomena into kinds in other sciences. One way to make this variety tractable, suggested by the natural history approach, is to look for some stable phenomena that do not vary. These are not likely to be found at the level of diagnoses, because in many cases two people with the same diagnosis will exhibit different behaviours and other symptoms depending on their personal lives, cultural background, other mental illnesses, and so on. There is too much variety at the level of diagnosis. So we must search for smaller units of explanation.

We can think of this first approach as zooming-in on one small portion of the overall network, looking for a unit that has enough stability across histories to allow for an explanation in which there are cause-effect regularities; if the underlying causal structure is too diverse, look for causal units that stably replicate across patients, explain them, and then put these well-understood parts together to explain individuals. Another advantage of this approach is that it promises to treat patients as individuals rather than clumping together into kinds that they only approximately fit.

This zoom-in approach is exemplified by Bentall (2005) and Spaulding et al (2003). They both look for the smallest unit of reliable explanation or manipulation. Spaulding et al. argue from a clinical perspective for seeing patients as collections of “problems”, with the problems in each case drawn from a repertoire of clinically significant phenomena that different patients exemplify in different ways. Bentall argues that diagnoses like depression or schizophrenia have proved helpless in the face of all the variation that patients exhibit. He sees individual cases as mosaics made up of recurring symptoms like hallucinations, which can be separated out and studied in isolation. According to this view, there is no such thing as schizophrenia. That's not because there is nothing wrong with schizophrenics, but because schizophrenia is not a natural kind. The natural kinds of psychiatry are specific pathologies that schizophrenics display in shifting conjunctions. These are distinct psychotic phenomena that should be approached separately and treated as distinct symptoms or complaints.

Bentall thinks “we should abandon psychiatric diagnoses altogether and instead try to understand and explain the actual experiences and behaviours of psychotic people” (2005, 141). But in fact he tries to relocate diagnosis at a more reliable level. For example, Bentall (ch. 15) objects to the old model of inferring thought disorder on the basis of disordered speech. He sees disordered speech as a failure of communication, which is especially likely when subjects are emotionally aroused. In Bentall's tentative model, initial deficits in working memory caused by emotional arousal interact with other deficits in semantic memory, theory of mind and introspective monitoring. The result is a failure to

communicate and a lack of self-awareness of one's failure to communicate (which distinguishes psychotic patients from normal subjects in the grip of powerful emotions who are struggling to get their ideas across). This is a stable phenomenon, in the sense that we can give the same causal story in all cases of thought disorder, thus giving us a robust account that transfers across patients. A general theory of schizophrenia would have too many qualifications and varieties to transfer in this way.

Zooming-in faces two problems. First, where do we stop when looking for smaller units? There is always the chance that some finer causal discrimination will uncover an even more stable structure.

A proponent of zooming-in can always argue that the idea, as in any science, is to develop the descriptive apparatus empirically in a way that ultimately fits one's theory of the domain. Bentall and Spaulding et al. are betting that any clustering of problems or symptoms will not line up neatly with the DSM categories, and this is probably a good bet for any approach. Bentall's chosen approach stresses the cognitive science of psychotic phenomena, whereas Spaulding et al. tend to look for correlations between clinically significant problems and a wider variety of underlying phenomena; for Spaulding et al, the relevant sciences for identifying problems span many levels of analysis from the intracellular to the socio-cultural. Their examples include low levels of cortisol and learning deficits (127–130).

Nonetheless, there will always be differences across patients in the phenomena that we ultimately zoom in to. This is likely to be especially problematic when these independently characterised lower-level phenomena start to combine in actual patients. Zooming in should not, therefore, be thought as an alternative to idealization, since any search for commonalities across patients will involve some degree of idealization. The rival approach is another form of idealization, but zooms out to think in terms of idealized patients, rather than idealized problems.

Zooming out begins by accepting that mental illness almost always depends on complex interactions between diverse lower-level phenomena and tries to deal with this variety by treating disease networks as idealizations that are more or less similar to actual histories. It is a recurring theme in the history of psychiatry that the textbook expression of a disorder can be realized in individuals in different ways. Charcot (1887–88) distinguished *archetypes*, which are ideal types of disorder, from *formes frustes*, which are the imperfect forms the ideal type manifests in individual subjects. Birnbaum's (1923) distinction between the *pathogenic* features of a psychosis and its *pathoplastic* features is similar. The former define its essential structure, and the latter express the personal circumstances of the patient.

Murphy (2006) argues that the variety in mental illness requires us to explain psychiatric phenomena not by looking for stable regularities but by constructing exemplars. Murphy sees the exemplar as an imaginary patient who has the ideal textbook form—something like Thagard's network, and the narrative it provides—of a disorder, and only that disorder, although the textbook needs to be thought of as a statement of the final theory, and not any

current work in psychiatry. As with zooming-in, the approach assumes that commonalities across patients will be revealed by inquiry. The idea is to model all the causes that contribute to a natural history so as to explore how they work together in different contexts to produce the various outcomes of interest, and then understand patients as idiosyncratic instances of the exemplar. When we explain a disease, we construct an exemplar and model it. But the causal structure that explains the exemplar resembles real world patients to varying degrees. So when we talk about individuals we explain their symptoms by identifying processes in the patient and showing how they resemble some part of the model. Godfrey-Smith (2006), following Giere, suggests that there may not be a helpful general treatment of these resemblance relations. They are not formal relations but the sort of comparisons between imaginary and real states of affairs that we all perform effortlessly. If there is a general theory, then, it is likely to be found not by philosophers but by cognitive scientists studying analogical thought (e.g. Hummel and Holyoak 2005). And because it is instrumental concerns that drive our search for explanations in psychiatry, as in medicine more generally, the clinically significant relations between disease, model and patient are likely to be highly context-specific; they will be determined in part by whether they offer opportunities for successful therapeutic interventions, which depend not just on how the world is arranged, but also on what our resources and opportunities are.

The exemplar lets us identify *robust processes* (Sterelny 2003, 131–2, 207–8) that are repeatable or systematic in various ways, rather than the actual processes that occur as a disorder unfolds in one person. But we do not stop there: the ultimate goal is causal understanding of a disease. We build a model to serve this end. It aims to represent the pathogenic process that accounts for the observed phenomena in the exemplar. To explain an actual history in a patient is to show how the processes unfolding in the patient resemble those that are assumed to occur in the exemplar. Exemplars provide an idealized form of the disorder that aims to identify the factors that remain constant despite all the individual variation. Not every patient instantiates every feature of an exemplar, and so not every part of a model will apply to a given patient. Once we understand the resemblance relations that exist between parts of the model and the exemplar, we can try to manipulate the model so as to change or forestall selected outcomes in the real world.

In commenting on this approach, Mitchell (2009) points out that her own work (2003) contains an alternative approach to model-based explanation in complex sciences. Her “integrative pluralism” aims to isolate individual causes and model them individually, seeing how each makes a causal contribution on its own. Theorists then put together a collection of models of individual phenomena and try to integrate them by applying multiple models as seems necessary to explain a particular case. Mitchell's approach resembles the zoom-in in its search for decomposing, but it tries to isolate causes that can recombine—such as genes or interpersonal difficulties—rather than searching for explanations of particular clinical phenomena like thought disorder (see also Tabery 2009). This is similar to the way Spaulding et al. seek the development and integration of causal models in the service of developing a comprehensive clinical understanding of people and their problems.

Mitchell (2009, 131) suggests that we might need both her models and Murphy's as circumstances dictate. And there might be room for zoom-in models as well. There seems no reason why psychiatric research could not aim to be pluralist about explanation and combine elements drawn from all these explanatory styles as we learn more about what works when we try to figure out mental illness.

Ghaemi (2003, ch 12) offers a different defense of zooming-out. He argues that current DSM diagnoses function as ideal types in Weber's sense. One way to understand Weber's idea is just as a forerunner of modelling, in which essential variables are isolated and inessential ones put aside. Ghaemi, though, locates it in a tradition of hermeneutic understanding that is most closely associated in psychiatry with Karl Jaspers (1997). This approach looks for meaningful psychological connections between phenomena and is contrasted with causal, scientific explanation. In Ghaemi's view the DSM categories are designed to foster understanding of this type by directing clinical attention to aspects of the patient's life that are relevant to the disorder at hand. He sees this as an application of the methods of the humanities, rather than those of the sciences.

One argument that might be made against zooming-out is that it assumes that mental disorders are categorical, rather than dimensional. The neo-Kraepelinian approach is often faulted for assuming that disorder is an all or nothing affair, whereas it might mark an extreme point on a dimension that is present to a greater or lesser degree throughout the population (Kendell 1975, McHugh and Slavney 1998, Poland et. al 1994, Widiger and Sankis 2000). So schizophrenia might be seen either as condition that is present or not—just as one either is of voting age or not. But we might also see it as like having high blood pressure—marking a point on one or more dimensions where normal psychological traits become aggravated enough to constitute a population of clinical concern. Some psychiatrists have supported one approach or the other. Others, like Ghaemi (203), have argued that some conditions, like the major psychoses and mood disorders, are probably categorical, whereas other conditions, especially the personality disorders, are dimensional.

This is clearly an empirical issue in the end, so if the zooming-out approach is committed to a categorical approach it is a cause for concern. However, a zooming out approach could treat its models as end points on a spectrum that usefully capture differences between populations (Ghaemi, 199); this is compatible with looking for causal stories that explain the model. Furthermore, in so far as the zooming-in approach relies on idealizations too, it faces the same problem. Bentall's approach, for instance, faces the issue of what degree of language difficulties counts as pathological, and Spaulding et al. need to find a way of distinguishing a clinical problem from a phenomenon that is normal but unwelcome.

3.5 Cross-Cultural Issues

Successive editions of the DSM have been published in North America, and questions have always arisen about the applicability of its categories outside the West. DSM-5 follows its predecessor in coping with cross-cultural variety in two ways. First, some conditions, such

as depression and anxiety, are treated as universal – found in every culture, despite local differences in character. We may term this the unificationist project. The DSM-IV-TR averred that major DSM categories occur all around the world but with symptoms and courses that are influenced by local cultural factors. Second, DSM-IV-TR employed the concept of culture-bound syndromes, which were defined as “recurrent, locality-specific patterns of aberrant behavior and troubling experience that may or may not be linked to a particular DSM-IV diagnostic category” (p. 898). Culture-bound syndromes occur in a limited number of specific societies or cultural areas. They are local, often traditional, diagnostic categories that make sense of patterns of thought and behaviour that are treated as deviant or troubling in that context.

In theory, culture-bound syndromes are those folk illnesses in which alterations of behaviour and experience figure prominently, relative to local norms. Standard DSM diagnoses are not thought of in this way, even if they are culturally limited. Multiple Personality Disorder/ Dissociative Identity Disorder, for example, was widely diagnosed in the USA in the 1980s, but seldom found elsewhere. That did not lead psychiatrists to treat it as a culture-bound syndrome, even though it seems to fit the definition. The assumption in mainstream psychiatry is that Western conditions are not culture-bound; they represent abnormalities in a universal human endowment. This is the case even for conditions (such as some kinds of body dysmorphia) for which prevalence data only exists in a few western nations.

DSM-5 has moved away from concept of a culture-bound syndrome and adopted that of “cultural concepts of distress” (p.758). These are ways in which “cultural groups experience, understand and communicate suffering, behavioural problems or troublesome thoughts and emotions”. DSM-5 distinguishes three main types of cultural concepts: syndromes, idioms (ways of expressing distress), and local explanations. The last is particularly interesting: a standard picture of mental illness in both non-Western and premodern societies represents it as explained by supernatural forces like ghost possession. But Edgerton (1966) reported that the East African tribes he studied were divided over the explanation of psychosis. Some tribes did understand it in terms of ghost possession, but others thought of it as having an organic cause (“brain worms”). DSM-5 makes room for such findings as part of the integration of local explanations into a theory of mental illness.

DSM-5 supplements this conceptual reform with a thorough overhaul of the “Outline for Cultural Formulation”. This is an interview outline introduced in DSM-IV-TR, in part in response to pleas from cross-cultural psychiatrists dissatisfied with the old ways (e.g. Canino and Alegria 2008). It reflects an unprecedented investment in applying anthropology to psychiatry by the DSM-5 Cultural Issues Subgroup, who ran field trials of the new Cultural Formulation Interview (CFI) in 6 sites around the world (Aggarwal 2013). The CFI is designed to assess both the cultural or ethnic groups that the subject belongs to and the ways in which those groups understand the problem and affect the subject’s experience of it. It comprises a series of questions designed to gather information from the point of view of the patient and his or her cultural peers, including specific questions that tell the patient not to repeat what the doctor has told them but to express their problems as

they see them and describe which aspects of their identity are most important to them. It could be used with any subject anywhere, and the manual acknowledges that all mental illness, including the DSM diagnoses, is shaped by the surrounding culture. On the other hand, the CFI and its associated conceptual material are poorly integrated into the DSM-5 – many diagnoses pay no attention to cultural formulation at all – and the culture-bound syndromes do in fact survive. They have been renamed “Cultural Concepts of Distress” but the appendix listing them survives, so one may wonder how thorough the overhaul has been.

The other approach to cultural variation, as mentioned, is to treat different diagnoses as manifestations of the same underlying DSM condition. The issue is how different a diagnosis needs to be before it stops being a form of a DSM condition and becomes a culture-bound syndrome. It has been suggested that anxiety disorders, for example, should have their “worry domains” expanded because people from other cultures may worry about things that Euro-Americans don’t fret over (Lewis-Fernandez et al 2010). The assumption here is that anxiety disorders are common enough in all human cultures to merit treating as a universal, but that this can be missed if clinicians do not appreciate the different triggers, reactions and conceptions that distinguish the anxiety disorders in different populations. We need to recognise the greater variety of symptoms among sufferers from anxiety who live in Western cultures but don’t show the typical symptoms of a Westerner. Lewis-Fernandez and his team (2010) argue that a majority of Latino sufferers from *ataque de nervios* would meet panic disorder criteria, for example, if the DSM amended the requirement that a panic attack lasts only a few minutes. *Ataque de nervios* was listed as a culture-bound syndrome in DSM-IV-TR. Its symptoms commonly include a mix of classic panic attack symptoms like palpitations and sensations of bodily heat and loss of control. But it may also involve uncontrollable shouting and threats of violence, together with other symptoms less often found in panic attacks, and it may last much longer. Lewis-Fernandez is interested in the idea that *ataque de nervios* can be assimilated to existing anxiety diagnoses provided we take the variation in symptomatology more seriously.

This unificationist approach may be used to argue that the diagnostic entity itself is the same in all cases – some shared underlying process that adapts or responds to different cultural settings to produce different behaviors. Some scholars argue that the underlying abnormality, even if it exists, is beside the point. Our response to the condition should be determined by its expression, even if there is an underlying psychological kind. Kleinman argued that depression in the West and in East Asia are different enough in their expression and course that they should be seen as different conditions (1987, 450): “Depression experienced entirely as low back pain and depression experienced entirely as guilt-ridden existential despair are such substantially different forms of illness behavior with different symptoms, patterns of help-seeking, course and treatment responses that though the disease in each instance may be the same, the illness rather than the disease is the determinant factor.”

Horwitz and Wakefield (2007, p. 199) opt for universality in response: “we agree with Kleinman’s distinction between disease as a universal underlying dysfunction and illness

as the culturally shaped expression of a given dysfunction ... [but] if there are indeed underlying common dysfunctions, then treatment presumably depends in large part on the science of identifying and intervening in such dysfunctions irrespective of their cultural presentation.”

The unificationist project is an empirical bet that is based on the assumption that human psychology is enough alike to treat at least some culturally distinctive forms of mental disorder as expressions of a shared human nature that is differently shaped by local culture. The appeal to culture bound syndromes is also a perfectly good bet, reflecting the idea that some forms of behaviour are entirely local and not assimilable to a universal model. Both unificationism and recognition of culture-bound syndromes could be pursued at the same time. Given the boring but probably correct observation that people around the world are both alike and different, that may be the obvious course. However, pursuit of these strategies should be freed from the apparent assumption in DSM-5 that western minds are the blueprint for the world. The western form of a condition may be unusual: the more we look, the more it seems as though the cognitive aspects of depression – the complex psychology of self-blame and immiserating introspection – may be distinctively Western. There may be conditions that DSM treats as universal but which in fact are culture-bound syndromes that only exist in NATO countries.

3.6 Clinical Application

Psychiatric models explain idealization of some sort. The point of it all, though, comes when that understanding lets us help patients. Guze (1992) calls this the move from the biological perspective to the clinical one. In making the move we shift from a general description of a disease process to a specific description of the biology of an individual. The clinician takes the imaginary ideal patient and specifies more real-world detail, so as to produce a description of a smaller set of patients, perhaps even a single case. The scientific project idealises, whereas the clinical one uses the resources of the science to help individuals. In displaying causal relations that give rise to features of the model we present opportunities for therapeutic action, whether we have zoomed in or out. The model defines a set of relations that differ from those present in a real patient along various dimensions. One is precision: the causal relations the model represents may need to be made more precise when we look at real patients. The degree to which a symptom is present, for example, might need to be specified precisely in a clinical setting, whereas in the exemplar the symptom can be defined as inhabiting some range of values, any one of which might apply in nature. Or we can supply a determinate story that embraces the details of the beliefs and other intentional states of the patient, instead of just citing the fact that a particular information-processing pathway is implicated in patients of that type.

4. Philosophical Psychopathology

4.1 Varieties of Philosophical Psychopathology

The work surveyed so far in this entry mostly counts as philosophy of science, where psychiatry is the science in question, and the philosophy is concerned with topics familiar to philosophers of science, such as explanation. Another tradition in philosophy of psychiatry belongs to what Graham and Stephens (1994, 6) call applied philosophy of mind, which is one of the approaches they distinguish in their rough taxonomy of approaches to philosophical psychopathology. The remaining category of work Graham and Stephens recognises looks at the ethical and experiential aspects of mental disorder. This section surveys some recent work in applied philosophy of mind. It does not aim to be comprehensive, since this area has probably seen the most intense flowering of recent philosophical work. We will look at some representative work to get a sense of the varieties of topics and projects under philosophical discussion.

Philosophers writing in this vein (exemplified by Graham, 2010) look to incorporate empirical results into philosophical reflection on topics like personal identity, rationality or voluntary action. They may also seek distinctively philosophical explanations of psychopathologies, which try to understand them in terms borrowed from the philosophy of mind. Whereas Murphy (2006), for example, sees the scope of the mental as defined by the sciences of the mind, Graham thinks of the mental as possessing the twin hallmarks of intentionality and consciousness. Disturbances of intentionality and consciousness are what count as mental illnesses, therefore, and philosophical theories that treat of intentional or conscious states can help us to understand mental disorders better.

This philosophical project involves, but is not limited to, conceptual clarification. Philosophers have made useful contributions by interrogating the details of psychiatric concepts. This way of doing philosophical psychopathology is exemplified by Part One of Potter (2009). She looks at key concepts in the current understanding of Borderline Personality Disorder, which is characterized by feelings of emptiness or loss of self, impulsive or self-destructive behaviour, bursts of anger and unstable relationships. A staple of the literature on the condition is the idea that borderline patients are manipulative. Potter (ch.6) usefully asks what this means. A therapist may often feel outmanoeuvred by a patient who has managed to bring about an outcome that the therapist did not prefer—for instance, a patient who still wishes to be admitted to hospital despite her therapist's judgement to the contrary may deliberately injure herself to force the hospital to admit her. In this case, the patient has certainly taken extreme measures to get her way, but it is not clear that it counts as manipulation. The patient has defied medical opinion, but she has not changed the therapist's mind or been deceptive—we might prefer to think of manipulation as deliberately changing someone else's mind through charm, trickery or some other psychological process. That kind of ability to get your own way is often ascribed to psychopaths, but in her work Potter makes us wonder if the same idea really applies to borderline patients. Perhaps they should just be seen as desperate people who are trying to cope with their distress, or socially inept (115–16) rather than agents setting out to bend

others to their will. And even if they are trying to get their own way, well, why shouldn't they? The concern Potter raises is that the concept of manipulateness is being used by clinicians to just mean something like 'behaviour that makes my job harder'.

4.2 Thought Insertion

The program set out by Graham and similar theorists has a different agenda that goes beyond conceptual clarification. It involves both using concepts from psychiatry to shed light on philosophical problems and using existing philosophical positions to explain, or at least make clearer, some of the issues that daunt psychiatrists. Take, for example, delusions of thought insertion. Subjects with this condition say that they have thoughts inside their minds that have been put their by someone else or are actually being thought by someone else, as in the following famous case reported by Mellor (1970): "I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his ... He treats my mind like a screen and flashes thoughts onto it like you flash a picture." (Eamonn Andrews was a British TV personality.)

It is very hard to imagine what thought insertion could be like: how can I be introspectively aware of something as an episode in my mind, yet at the same time experience it as someone else's thought? Graham and Stephens (2000) suggest that thought insertion should be seen as alienated self-consciousness. They draw on Frankfurt's (1988, 59) idea that only some of things I do actually express my agency. Some bodily movements can be movements of my limbs without counting as actions of mine—perhaps because someone else is controlling my movements. The same can be true of mental events. If a tune runs through my mind unbidden, it is an episode in my mental history. But it does not count as my mental activity in the same way as the mental arithmetic I do to balance my checkbook counts as mine. The latter, but not the former, comprises intentional thought that expresses my agency. Graham and Stephens (2000, 152) call these properties of thought the sense of subjectivity and the sense of agency. They propose that the two properties can come apart, so that in thought disorder the sense of subjectivity occurs without the sense of agency. On this picture, attributing the thought to someone else makes sense, because it is the kind of mental activity that must be produced by something, carrying as it does the hallmarks of agency (just not *my* agency).

Such a theory does not explain where the abnormal experiences come from. Rather, it tries to use philosophical resources to make the experience somewhat less unfathomable, and perhaps clear some ground for an explanation by pointing the way towards the sort of mental processes that would need to be involved. A different approach to philosophical psychopathology looks to the science to illuminate a philosophical debate.

4.3 Psychopaths and Moral Psychology

The debate over psychopathy seems to be a case where empirical findings have a clear philosophical contribution to make. Findings about psychopathy have been appealed to in two debates. First, they are involved in the debate over whether moral judgments are intrinsically motivating, or whether it is possible to know what morality requires yet remain unmoved. Jesse Prinz (2007, 42) has called psychopathy a “test case”, which helps us to discriminate between the claim that moral judgments are intrinsically motivating and the contrary claim that you could form a moral judgment without being motivated to act on it. A second, albeit related issue in moral psychology is the extent to which moral judgments are the product of reason or of affective sentiments.

Hume (1751, section IX) imagined the “sensible knave”, who, “in particular incidents, may think, that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That *honesty is the best policy*, may be a good general rule; but is liable to many exceptions: And he, it may, perhaps, be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions.” Hume's sensible knave knows what virtue requires but does not care, because he lacks the sentiment of benevolence that normally counteracts self interest. The sensible knave cannot be argued into morality but not because of any intellectual deficiency; he just does not feel the pull of morality at all. Shaun Nichols (2004) has argued that psychopaths are actual cases of sensible knavery, since they seem to fit Hume's criteria.

According to Nichols (2004), psychopaths are sensible knaves—amoral but fully rational agents. They lack prosocial emotions (especially empathy), and this lack leads them to disregard the claims of morality because they do not respond normally to the sufferings of others. Nichols argues that because the moral flaws of psychopaths are traceable to their emotional deficits, rather than any shortcomings of reason, psychopaths refute what he calls “empirical rationalism”. Empirical rationalism is the idea that ethics is based on, or is just a province of, practical reason. Our moral judgments are explicable in terms of the same rational standards that apply to ordinary thought and action. Simply put, being moral is just indistinguishable from being rational. If psychopaths refute this view, thinks Nichols, they count as evidence in favour of a broadly Humean approach to morality

Nichols distinguishes empirical rationalism from moral rationalism. Empirical rationalism claims that it is an empirical fact that moral judgment in humans is a kind of rational judgment; i.e. our moral judgments derive from our rational faculties or capacities. Conceptual Rationalism states that it is a conceptual truth that people who make moral judgments are motivated by them. That is, if you understand what morality requires you will want to do it (because it is the rational thing to do). Nichols concentrates on the empirical case—the extent to which moral psychology is a form of practical reason.

Heidi Maibom (2005) has challenged Nichols' treatment on several grounds. Most directly, Maibom points to literature showing that psychopaths do in fact suffer from rational impairments. Psychopaths seem to have trouble learning from experience, lack a grasp of

means-end reasoning, and lack an ability to plan for the longer term or formulate and pursue longer-term goals. Maibom argues that because of these deficits psychopaths fail to meet the rational constraints on practical reason, such as coherence and consistency. Maibom also argues that to support a Humean position, it is not enough to tie these blemishes in the reasoning of psychopaths to ultimate abnormalities in their emotional lives. For a Humean, she argues, emotional deficits should have a direct effect on morality via an absence of empathy, not an indirect one via reasoning failures caused by a general lack of emotions.

It does seem to be correct that psychopaths suffer from various rational deficits. But nobody is perfectly rational, and the case has to be made that psychopaths are different enough from normal subjects. For Maibom to make her case, the rational shortcomings of psychopaths need to make them a different class of subject, since Nichols' argument depends on their being as rational as the average subject. We remain unsure about the nature of the rational deficits in psychopathy, but also about the significance of the various models of psychopathic irrationality. Perhaps it is wrong to think of practical reason as a unitary phenomenon or faculty. Psychopaths seem to be perfectly normal when it comes to theory of mind and general intelligence, so maybe their reasoning deficits are circumscribed in ways that leave them rationally in the same class as normal subjects when it comes to the reasoning systems that matter for morality. It may be in the end that psychopathy involves a combination of emotional and rational shortcomings that reflect a failure of normal neuropsychological development. If that were the case, then they would be less suitable to adjudicate between Humean approaches and Kantian approaches that see morality in terms of practical reason, rather than emotional responses.

Controversy also exists over the extent to which psychopaths really understand moral concepts, which they would have to do in order to count as amoralists, rather than as people who do not understand morality. Although psychopaths can employ moral language and use emotion terms, some theorists (e.g. Hare 1993, ch 3) have argued that they seem incapable of attaching the same sense as the rest of us to the moral concepts they use. This may suggest that they do not really grasp moral concepts. However, it may still be that their failure to grasp moral concepts is due to their emotional stuntedness rather than some deficit in reasoning (Prinz 2007, 46–47).

4.4 Multiple Personality

Philosophers have also sought scientific support for their theories by appealing to Multiple Personality Disorder, or Dissociative Identity Disorder as it was called in DSM-IV. This has been a very controversial diagnosis. Some philosophers have thought that it shows something very interesting about the development of the self in normal subjects. Other theorist have been more sceptical, both about the philosophical claims and about the very reality of the condition itself.

The basic picture of MPD/DID is that a sufferer has several distinct personality states, known as “alters”. At various times, one of the alters is in charge of the subject's behaviour.

Each alter has a peculiar, enduring pattern of perceiving, thinking about and relating to the environment and self, as well as different patterns of speech and physical comportment. The “alters” emerge spontaneously and involuntarily to control behaviour, and function more or less independently of each other. Not all alters are known to the patient. This results in significant amnesia (not ordinary forgetting) on the patient's part for those periods of time when an alter was in control.

Scepticism about MPD in the clinical community was fostered by the circumstances under which it became prominent in the 1970s and 1980s. The existence of dissociation is not controversial. DSM-IV (477) defines it as “disruption in the usually integrated functions of consciousness, memory, identity or perception of the environment. The disturbance may be sudden or gradual, transient or chronic”. Dissociation does not have to be pathological: it can occur in normal life when, for example, one becomes absorbed in a task to the exclusion of other aspects of the environment. And the idea of an alter personality assuming control of behaviour emerged in the nineteenth century. But the sceptics point out that fewer than a dozen cases had been reported before the early 1970s, when a sudden epidemic flowered up in North America, featuring subjects who had several, sometimes dozens, of alters. Proponents of the diagnosis pointed to greater awareness of child abuse as the key factor that made widespread diagnosis possible, since it took great efforts to overcome public reticence on that key causal factor.

By the mid-nineties, the epidemic was waning, as more and more psychiatrists began to conclude that the symptoms of MPD/DID were being produced in the clinic by hypnosis and suggestion, in concert with portrayals of the condition in the media. Questions were also raised in retrospect about the validity of the diagnosis in some of the key cases that inspired the great take-off in the diagnosis (McNally, forthcoming ch 5). But there is no doubting the reality of the symptoms that are displayed by people who receive the diagnosis. A tendency has been to ask whether multiple personality is a real disorder or the product of social circumstances, a culturally permissible way to express distress? We should reject the presupposition that there is an important contrast between being a real disorder and being a product of social circumstances. The fact that a certain suite of mental symptoms appear together only in specific historical or geographical contexts does not imply that it is not real (See Hacking 1998 on transient mental disorders). Philosophical discussion of dissociation is not imperilled if the sceptics about MPD carry the day. However, philosophical theories which are tied strongly to the more controversial claims about the nature and etiology of MPD do look less plausible.

Dennett and Humphreys draw on MPD/DID to support a philosophical proposal about the nature of the self. They do seem to tie their interpretation of MPD/DID to the specifics of the causal story, since they treat the condition as evidence for a particular theory about the development of the self in normal cases. Different alters, as they put it, express exaggerated moods but retain memories of their periods of control that host personalities lack. These memories become aggregated into a distinct self (or self-representation). The causal story they favour (49–51) is borrowed from the MPD community, and asserts that children endure abuse by dissociating and leaving an alter behind to deal with the abuse. Then the

alter splits into pieces that express different emotional relations to what is going on (including conflicting feelings towards the parent who is also the abuser).

Are these alters different selves? Dennett and Humphreys distinguish two theories of the self.

1. The “Proper Self”: a ghostly supervisor who lives inside the head. The seat of thought and consciousness.
2. The “Fictive Self”: a “center of narrative gravity” of a biography, but with no real causal power. This is a representation of what you are like that does not correspond to anything inside the head.

Dennett and Humphreys dislike the “proper Self” idea because they think that it is untrue to the way the human mind is organized. They argue that the mind is a collection of unintelligent subcomponents, with no overall executive control center. Like a termite colony, a human being is made up of a set of independent systems: different functional units responsible for (e.g.) vision, language processing, theory of mind, working memory, emotional appraisal, spatial cognition and so on. All these simple systems give rise to intelligent behaviour. But unlike a termite colony, a human being has a figurehead: each of us “first creates—unconsciously—one or more ideal fictive selves and then elects the best supported of these into office as her head of mind” (1998, 41)

On this view, normal development of the self is a matter of finding the representation that makes the most sense. The representation that makes the most sense does not have to be accurate. It reflects a desire for coherence (and a role in social relations) rather than accuracy. In MPD, Dennett and Humphreys argue that this process fragments into lots of competing systems, since there are too many inconsistent pictures of the “real you”: so the alters persist as different representations that make partial sense of different aspects of the self. The normal integration of different representations into a coherent one is prevented by the strains of development.

Owen Flanagan (1994) has developed a version of this view that is slightly different. He argues that the self is a narrative: an unfolding rationale for what in fact happens to you. Normally we assume that there is one self per person, and that the self permits continuity across change as we develop. But in the normal case, Flanagan thinks, the self is multiplex—it permits some differentiation across situations. The notion of a complete, well-integrated personality is a normative concept—a way we think people ought to be—rather than anything that is usual in psychology. It is a goal we work towards that can be achieved in different ways.

These views treat MPD as evidence for a particular theory of how the self comes into being via developmental work at integrating different aspects of the person. Ian Hacking (1995a, chs 16–18), on the other hand, is sceptical of any claims that MPD/DID is evidence for positions in the philosophy of mind. It may illustrate positions—if you believe Dennett and Humphreys' theory of the self then you can see how their view makes sense of the phenomenon of MPD/DID. But the philosophical theory has to be confirmed elsewhere.

This is partly because the understanding of the syndrome the theory rests on is so recent: Hacking argues that people “have been diagnosed with double consciousness or multiple personality for two centuries `now. But they began to talk the way they do now—using the symptom language described by Dennett—only very recently. Today they all say such things, or at least suspect that they ought to say them. That is how they learn to describe themselves in therapy” (226). Dissociation may be a genuine psychological condition, but the details of dissociative identity disturbances are too contentious to bear much philosophical weight. Hacking doubts that there is any stable condition to have knowledge about, because the very act of being classified alters human behaviour via what he calls the “looping effect” (1995b, 2007. For discussion, see Cooper 2004, Murphy 2001, 2006 (ch 7), Tsou 2007). The looping effect comes about when individuals change their behaviour in response to classification. Such a feedback loop could, in principle, keep behaviour stable, but Hacking is interested in those cases in which the behaviour is rendered unstable. He has also (1998) examined the case of transient mental illnesses, which appear to flare up and then go out of existence.

4.5 Delusions

Of all psychiatric concepts, it is probably delusions that have received the most extensive recent treatment from philosophers. Here the issues straddle philosophy of mind and philosophy of science. Some theorists have raised conceptual difficulties for accounts of delusions, others have sought to frame scientific hypotheses that can explain delusions in terms of information-processing deficits or other subpersonal problems. And some philosophers have looked for a distinctively philosophical explanation of delusions, which draws on concepts previously formulated in the philosophy of mind.

Like many other psychological concepts, the concept of delusion has a fairly explicit and self-conscious scientific use and a variety of commonsense uses. There is considerable overlap between some psychiatric uses and some casual employments of “delusion”. But there are clearly everyday occasions when “delusion” merely refers to a belief that seems obviously false or unwarranted to the speaker. I might call you delusional when you announce that you expect to buy a four bedroom house in your neighbourhood for what you can afford to pay, without meaning that your reasoning is symptomatic of a psychotic illness; it is just wishful thinking. But delusions in a technical sense are manifestations of psychosis. DSM-IV-TR (819) defines a delusion as ‘A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary’. Furthermore, the delusion must be inexplicable in cultural terms, so that religious beliefs are not counted as delusional, no matter how bizarre. The fact that the patient belongs to a culture or community in which the relevant belief is treated as a religious claim means that it does not count as a delusion. This proviso has seemed odd to many people. It may just testify to the disinclination of psychiatrists to take sides in religious disputes, but perhaps the best way to view it is as an acknowledgment of the

importance of testimony or social context in the normal acquisition of our beliefs. We expect people to acquire religious, or otherwise culturally specific, beliefs that might strike outsiders to the culture as odd. But the weirdness of the belief does not impugn the normality of the process of culturally specific belief transmission, so the products of such cultural transmission should not be treated as pathological, no matter how strange we may think they are.

Other aspects of the definition of delusion have also come under attack. Gipps and Fulford 2004 ably summarize some standard objections. These include the idea that a delusion has to be false. Jaspers (1997, 106), for instance, argues that someone who is pathologically jealous and suffers from delusional beliefs about their partner's infidelity (“the Othello Syndrome”) might in fact be correct. The partner could actually be unfaithful. In this case the truthfulness of the belief is beside the point, since we can recognise pathological jealousy without regard to truth. Ryan (2009, 24) notes that someone with Ekbom's syndrome (the belief that one is infested with bugs) might coincidentally suffer scabies. Fulford (1989, 204–5) also mentions a patient whose only symptom of psychopathology was the delusion that he was mentally ill—if he really was mentally ill, then that couldn't have been a delusion. On the other hand, as Graham and Stephens (2007, 194) note, many theorists have doubted that delusions are really beliefs at all. Delusions often lack properties that beliefs have been thought to possess. In particular, delusions often seem poorly integrated with the rest of the subject's beliefs, and they can lack the affect associated with firmly-held beliefs. These objections are rebutted by Bortolotti (2009). The boundaries between delusion and other states, like self-deception, have also been discussed (Bayne and Fernandez 2009. For general philosophical accounts of delusion see Bortolotti 2009, Radden 2010.)

Analyses of delusion can look like inversions of the attempts made by analytic epistemologists to define knowledge. We may agree that a delusion is a false belief, just as knowledge is true belief, but, as with knowledge, we do not rest there. Knowledge is true belief plus something that ties the true belief to the world in the right way. So too, we try to find that extra property of the false belief that converts it from a false belief into a delusion. This conceptual program is seen as the prelude to the development of empirical theories of delusion, in which philosophers work alongside cognitive scientists (Davies and Coltheart 2000).

These theories have tended to assume that when ‘delusion’ is used in the technical sense it denotes a psychological natural kind (Samuels 2009) whose basic psychological structure can be worked out via attention to paradigmatic cases. Call this *the explanatory project*. We can distinguish two versions; one involves explaining delusions as caused by failures of normal relations among components of our cognitive architecture, or at least those parts of it that have to do with the fixation of belief. Another way of proceeding is more philosophical. It is less bothered with causal theories of cognitive architecture and more concerned to find resources in the philosophy of mind that characterise delusions in an illuminating way.

Neuropsychological theories of the origin of monothematic delusions (Those characterised by one dominant delusional belief about one specific matter) have come to fall into two camps: one factor and two factor theories. According to one familiar version of the two-factor theory (Davies and Coltheart 2000), an unusual experience is the first factor in the aetiology of a delusion, but there must also be a second factor. The first factor can be seen as an impairment that effects the reception of information about the world. The delusion belief explains that input—your spouse seems odd, so you conclude that he has been replaced by a robot. That explains the oddness, but in order for you to form that belief there has to be a second impairment, in some system that forms or evaluates beliefs. This will typically be some neurocognitive deficit that interacts with the experience. The impairment in your reasoning system prevents you from rejecting the explanatory belief, despite its lack of warrant. The disagreement with one factor theories (Davies et al. 2005) turns on whether the first deficit suffices to cause the delusion.

The more purely philosophical approach to explaining delusions concentrates on the role the delusion plays in the subject's mental life, as understood in terms of some theories in philosophy of mind. Arguably, these attempts at philosophical understanding of delusions originate with Jaspers' (1997) concept of a *primary* delusion, which is a subjectively meaningful transformation in the patient's experience of the world. Recent work in this tradition includes the *delusional stance* idea of Graham and Stephens (2007). In a similar vein to their explanation of thought disorder in terms of identifying with one's own thoughts or being alienated from them, they suggest (194) that delusions are constituted by a higher-order stance towards lower-order intentional states. This stance involves an identification with the lower order states, which become seen as an expression of one's own agency or mental nature. As in Jaspers' treatment, it is the role of the delusion in the subjective mental life of the deluded person that is critical.

These two approaches represent two ways of doing philosophical psychopathology, as a foundational contribution to a scientific inquiry or as a piece of philosophy of mind. They are not incompatible; the philosophical projects may help us to better describe the phenomena psychiatrists care about and make the explanatory project easier. On the other hand a two-factor explanation, or some other neuropsychological account, does not have to falsify a philosophical description of delusions. It may be that philosophical accounts make unwarranted empirical assumptions, but they may simply illuminate the condition as it is experienced by those who suffer from it, or guide us towards a philosophical account of the mind that is informed by real cases rather than armchair ones. The last decade has seen a great increase in work on the epistemology and philosophical significance of psychiatry, which is becoming a recognisable philosophical field. The full implications of this development, both intellectually and professionally, are still a long way off.

Bibliography

- Aggarwal N.K., 2013 “Cultural Psychiatry, Medical Anthropology, and the DSM-5 Field Trials,” *Medical Anthropology*, 32: 393–398
- American Psychiatric Association, 2000. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, textual revision, Washington, D.C.: American Psychiatric Association; referenced above as **DSM-IV-TR**.
- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition, Arlington, VA: American Psychiatric Association; referenced above as **DSM-5**.
- Andreasen, N.C., 1997. “Linking Mind and Brain in the Study of Mental Illnesses: A Project for a Scientific Psychopathology,” *Science*, 275: 1586–93.
- —, 2001. *Brave New Brain*, New York: Oxford University Press.
- Ankeny, R.A., 2000. “Fashioning Descriptive Models in Biology: Of Worms and Wiring Diagrams,” *Philosophy of Science*, 67 (Supplement): S260-S272.
- —, 2002. “Reduction Reconceptualized: Cystic Fibrosis As A Paradigm Case For Molecular Medicine,” in L.S. Parker and R.A. Ankeny (eds.), *Mutating Concepts, Evolving Disciplines: Genetics, Medicine and Society*, Dordrecht: Kluwer: 127–141.
- Bayne T. & J. Fernandez (eds.), 2009. *Delusion and Self-Deception*, New York: Psychology Press.
- Beebe, H. & N. Sabbarton-Leary, 2010. “Are Psychiatric Kinds ‘Real’,” *European Journal of Analytic Philosophy*, 6: 11–27.
- Bentall, R.P., 2003. *Madness Explained*, London: Penguin.
- Birnbaum, K., 1923. “The Making of a Psychosis,” Tr. H. Marshall, in S. R. Hirsch & M. Shepherd (eds.), *Themes and Variations in European Psychiatry*, Bristol: John Wright. 1974: 197–238.
- Black, K., 2005. “Psychiatry and the Medical Model,” in E. Rubin & C. Zorumski (eds.), *Adult Psychiatry*, 2nd edition, Malden, MA: Blackwell, 3–15.
- Bolton, D. & Hill, J., 2005. *Mind, Meaning and Mental Disorder: The Nature of Causal Explanation in Psychiatry*, 2nd ed., New York: Oxford University Press.
- Bortolotti, L., 2009 *Delusions and Other Irrational Beliefs*, New York: Oxford University Press.
- Broome, M. & Bortolotti, L. (eds.), 2009. *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, Oxford: Oxford University Press.
- Bynum, W.F., 1994. *Science and the Practice of Medicine in the Nineteenth Century*, Cambridge: Cambridge University Press.
- Campbell, J., 2008. “Causation in Psychiatry”, in Kenneth Kendler and Josef Parnas (eds.), *Philosophical Issues in Psychiatry*, Baltimore: Johns Hopkins University Press, 196–216.

- Canino, G. & Alegria, M., 2008. "Psychiatric diagnosis: Is it Universal or Relative to Culture?" *Journal of Child Psychology and Psychiatry*, 49: 237–50.
- Carter, K.C., 2003 *The Rise of Causal Concepts of Disease*, Aldershot: Ashgate.
- Charcot, J-M., 1987. *Charcot, the Clinician: The Tuesday Lessons*, C.G. Goetz (trans.), Philadelphia: Lippincott, Williams & Wilkins.
- Cooper, R., 2004. "Why Hacking is wrong about human kinds," *British Journal for the Philosophy of Science*, 55: 73–85.
- —, 2007. *Psychiatry and Philosophy of Science*, Stocksfield: Acumen Publishing.
- Craver, C. F. and Bechtel, W., 2007. "Top-down causation without top-down causes," *Biology and Philosophy*, 22: 547–563. [[Reprint available online](#)].
- Davies, M & Coltheart, M., 2000. "Pathologies of Belief," *Mind and Language*, 15: 1–46.
- Davies, M., Davies A., & Coltheart M., 2005. "Anosognosia and the Two-Factor Theory of Delusions," *Mind and Language*, 20: 241–57.
- Dennett, D. & Humphreys, N., 1998. "Speaking for Ourselves," in D. Dennett, *Brainchildren*, Cambridge, MA: MIT Press, 31–57.
- Edgerton, R. B., 1966. "Conceptions of Psychosis in Four East African Societies," *American Anthropologist*, 68: 408-425
- Flanagan, O., 1994. "Multiple Identity, Character Transformation and Self-Reclamation," in G. Graham & L. Stephens (eds.), *Philosophical Psychopathology*, Cambridge, MA: MIT Press, 135–162.
- Frankfurt, H., 1988. *The Importance of What We Care About*, Cambridge: Cambridge University Press.
- Fulford, K.W.M., 1989. *Moral Theory and Medical Practice*, Cambridge: Cambridge University Press.
- Fulford, K.W.M, Thornton, T. & Graham, G., 2006. *Oxford Textbook of Philosophy of Psychiatry*, New York: Oxford University Press.
- Ghaemi, S. N., 2003. *The Concepts of Psychiatry*, Baltimore: Johns Hopkins University Press.
- Giere, R., 1988. *Explaining Science: A Cognitive Approach*, Chicago: University of Chicago Press.
- Gipps, R. and Fulford, K. W. M., 2004. "Understanding the Clinical Concept of Delusion: From an Estranged to an Engaged Epistemology," *International Review of Psychiatry*, 16: 225–235.
- Glymour, C., 1992. "Freud's Androids," in J. Neu (ed.), *The Cambridge Companion to Freud*, Cambridge, Cambridge University Press, 44–85.
- Godfrey-Smith, P., 2006. "The strategy of model-based science," *Biology and Philosophy*, 21: 725–740.

- Gold, J. & Gold, I., 2014. *Suspicious Minds: How Culture Shapes Madness*, New York: Free Press.
- Graham, G., 2010. *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness*, London: Routledge.
- Graham, G. & Stephens, L. (eds.), 1994. *Philosophical Psychopathology*, Cambridge, MA: MIT Press.
- —, 2000. *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*, Cambridge, MA: MIT Press.
- —, 2007. “The Delusional Stance”. In Chung, M., Fulford, K. and Graham G. (eds.) *Reconceiving Schizophrenia*, New York: Oxford University Press, 193–215.
- Guze, S. B., 1992. *Why Psychiatry Is a Branch of Medicine*, New York: Oxford University Press.
- Hacking, I., 1995a. *Rewriting the Soul*, Princeton: Princeton University Press.
- —, 1995b. “The looping effects of human kinds,” in D. Sperber, D. Premack & A.J. Premack (eds.), *Causal cognition: A Multidisciplinary Debate*, Oxford: Clarendon Press: 351–394.
- —, 1998. *Mad Travelers: Reflections on the Reality of Transient Mental Illnesses*, Charlottesville, VA: University Press of Virginia.
- —, 2007. “Kinds of people: Moving targets,” *Proceedings of the British Academy*, 151: 285–318.
- Hare, R. D., 1993. *Without Conscience*, New York: Guilford press.
- Heinrichs, R.W. 2001. *In Search Of Madness; Schizophrenia and Neuroscience*, New York: Oxford University Press.
- Hempel, C., 1965. “Fundamentals of Taxonomy,” in C. Hempel, *Aspects of Scientific Explanation*, New York: The Free Press, 137–154.
- Horwitz, A. & Wakefield, J., 2007. *The Loss of Sadness*, New York: Oxford University Press.
- Hummel, J. E., & Holyoak, K. J., 2005. “Relational Reasoning in a Neurally-plausible Cognitive Architecture: An Overview of the LISA Project,” *Current Directions in Cognitive Science*, 14: 153–157. [[Reprint available online](#)].
- Hume, D., 1751. An Enquiry Concerning The Principles of Morals. In *Enquiries concerning Human Understanding and concerning the Principles of Morals*, edited by L. A. Selby-Bigge, 3rd edition revised by P. H. Nidditch, Oxford: Clarendon Press, 1975.
- Insel, T. R., Cuthbert, B., Garvey, M., Heinsen, R., Kozalk, M., Pine, D. S., Quinn, K., Sanislow, C., and Wang, P., 2010. “Towards a New Classification Framework for Research on Mental Disorders,” *American Journal of Psychiatry*, 167: 748–751.
- Jaspers, K., 1997 *General Psychopathology*, 7th edition, J. Hoenig & M. W. Hamilton (trans.), Baltimore: Johns Hopkins University Press.

- Kandel, E., 2005. *Psychiatry, Psychoanalysis and the New Biology of Mind*, Arlington, VA: American Psychiatric Publishing.
- Kendell, R. E., 1975. *The Role of Diagnosis in Psychiatry*, Oxford: Blackwell Scientific Publications.
- Kendler, K. S. & Campbell, J., 2009. "Interventionist Causal Models in Psychiatry: Repositioning the Mind-Body Problem" *Psychological Medicine*, 39: 881–887.
- Kendler, K. S. & Prescott, C. A., 2006. *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders*, New York: The Guilford Press.
- Kendler, K.S., Zachar, P.&Craver, C., 2011. "What Kinds of Things Are psychiatric Disorders?" *Psychological Medicine*, 41: 1143-1150.
- Kincaid, H., 2008. "Do We Need Theory to Study Disease? Lessons from Cancer Research and Their Implications for Mental Illness," *Perspectives in Biology and Medicine*, 51: 367–378.
- Kleinman, A., 1987. "Anthropology and psychiatry. The role of culture in cross-cultural research on illness," *British Journal of Psychiatry*, 151: 447–454.
- Kraepelin, E., 1899. *Psychiatry: A Textbook for Students and Physicians*, 6th ed. Repr. Canton, MA; Science History Publications 1990.
- Lamb, K., Pies, R., & Zisook, S., 2010. "The Bereavement Exclusion for the Diagnosis of Major Depression: To Be, or Not to Be," *Psychiatry*, 7: 19–25.
- Lewis, D.A. & Levitt, P., 2002. "Schizophrenia as a Disorder of Neurodevelopment," *Annual Review of Neuroscience*, 25: 409–432.
- Lewis-Fernández R, Hinton D. E., Laria A. J., Patterson E. H., Hofmann S. G., Craske M.G., Stein D.J, Asnaani A.& Liao B., 2010. "Culture and the Anxiety Disorders: Recommendations for DSM-5," *Depression and Anxiety*, 27: 212–229.
- Maibom, H., 2005. "Moral Unreason: The Case of Psychopathy". *Mind & Language*, 20: 237–57. [[Reprint available online](#)].
- Marr, D., 1982. *Vision*, San Francisco: W.H.Freeman.
- McHugh, P. & T. Slavney, 1998. *The Perspectives of Psychiatry*, Baltimore: Johns Hopkins University Press.
- McNally, R.J., forthcoming. *The Meaning of Madness*, Cambridge, MA: Harvard University Press.
- Meehl, P.E., 1977. "Specific Etiology and Other Forms of Strong Influence: Some Quantitative Meanings," *Journal of Medicine and Philosophy*, 2: 33–53.
- Mellor, C.S., 1970. "First Rank Symptoms of Schizophrenia," *The British Journal of Psychiatry*, 117(536): 15–23.
- Murphy, D., 2001. "Hacking's Reconciliation: Putting the Biological and Sociological Together in the Explanation of Mental Illness," *Philosophy of the Social Sciences*, 31: 139–162.

- —, 2006. *Psychiatry in the Scientific Image*, Cambridge, MA: MIT Press.
- —, 2008. “Levels of Explanation in Psychiatry,” in K. Kendler & J. Parnas (eds.), *Philosophical Issues in Psychiatry*, Johns Hopkins University Press, 99–125.
- —, 2009. “Psychiatry and the Concept of Disease as Pathology,” in M. Broome & L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, Oxford University Press: 103–117.
- Nichols, S., 2002. “How Psychopaths Threaten Moral Rationalism, or is it Irrational to Be Amoral?” *The Monist*, 85: 285–304.
- Parnas, J. Nordgaard, J. & Varga, S., 2010. “The Concept of Psychosis: A Clinical and Theoretical Analysis,” *Clinical Neuropsychiatry*, 7: 32–37.
- Pearl, J., 2009. *Causality*, 2nd ed. Cambridge: Cambridge University Press.
- Poland, J., Von Eckardt, B. & Spaulding, W., 1994. “Problems with the DSM Approach to Classifying Psychopathology,” in G. Graham & L. Stephens (eds.), *Philosophical Psychopathology*, Cambridge, MA: MIT Press.
- Potter, N., 2009. *Mapping the Edges and the In-between: A Critical Analysis of Borderline Personality Disorder*, New York, Oxford University Press.
- Prinz, J., 2007. *The Emotional Construction of Morals*, Oxford: Oxford University Press.
- Radden, J. (ed.), 2000. *The Nature of Melancholy from Aristotle to Kristeva*, New York: Oxford University Press.
- —, 2010. *On Delusion*, London: Routledge.
- Ross, D., C. Sharp, R. E. Vuchinich & D. Spurrett, 2008. *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling*, Cambridge, MA: MIT Press.
- Ryan, C., 2009. “Out on a Limb: The Ethical Management of Body Integrity Identity Disorder,” *Neuroethics*, 2: 21–33.
- Sadler, J.Z., 2004. *Values in Psychiatric Diagnosis*, New York: Oxford University Press.
- Samuels, R., 2009. “Delusion as a Natural Kind,” in M. Broome & L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: philosophical perspectives*, Oxford: Oxford University Press, 49–79.
- Sanislow, C. A, Pine, D. S., Quinn, K. et al., 2010. “Developing Constructs for Psychopathology Research: Research Domain Criteria,” *Journal of Abnormal Psychology*, 119: 631–639.
- Schaffner, K.F., 1993. *Discovery and Explanation in Biology and Medicine*, Chicago: University of Chicago Press.
- —, 1994. “Reductionistic Approaches to Schizophrenia,” in J. Sadler, O. Wiggins & M. Schwartz (eds.), *Philosophical Perspectives on Psychiatric Diagnostic Classification*, Baltimore: Johns Hopkins University Press, 279–294.
- —, 1998. “Genes, Behaviour, and Developmental Emergentism: One Process, Indivisible?” *Philosophy of Science*, 65: 209–252.

- —, 2012. “A Philosophical Overview of the Problems of Validity for Psychiatric Disorders,” in K. Kendler and J. Parnas (eds.), *Philosophical Issues in Psychiatry II: Nosology*, Oxford: Oxford University Press, 169–189.
- Spaulding, W., Sullivan, M. & Poland, J., 2003. *Treatment and Rehabilitation of Severe Mental Illness*, New York: Guilford Press.
- Spirtes P., Glymour C. & Scheines, R., 2001, *Causation, Prediction and Search*, 2nd edition, Cambridge, MA: MIT Press.
- Steiger, H., Gauvin, L., Israel, M., Koerner, N., Ng Ying Kin, N.M.K., Paris, J. & Young, S.N., 2001. “Association of Serotonin and Cortisol Indices with Childhood Abuse in Bulimia Nervosa,” *Archives of General Psychiatry*, 58: 837–843.
- Sterelny, K., 2003. *The Evolution of Agency and Other Essays*, Cambridge: Cambridge University Press.
- Tabery, J., 2009. “Difference Mechanisms: Explaining Variation with Mechanisms,” *Biology and Philosophy*, 24: 645–664.
- Thagard, P., 1999. *How Scientists Explain Disease*, Princeton: Princeton University Press.
- Thornton, T., 2007. *Essential Philosophy of Psychiatry*, New York: Oxford University Press.
- Tsou, J. Y., 2007. “Hacking on the Looping Effects of Psychiatric Classifications: What Is an Interactive and Indifferent Kind?” *International Studies in the Philosophy of Science*, 21(3): 329–344.
- Wachbroit, R., 1994. “Normality As A Biological Concept,” *Philosophy of Science*, 61: 579–591.
- Wakefield, J., 1992. “Disorder As Harmful Dysfunction: A Conceptual Critique of DSM-III-R's Definition of Mental Disorder,” *Psychological Review*, 99: 232–47.
- Weisberg, M., 2007. “Who is A Modeler?” *British Journal for the Philosophy of Science*, 58: 207–233.
- Whitbeck, C., 1977. “Causation in Medicine: The Disease Entity Model,” *Philosophy of Science*, 44: 619–637.
- Widiger, T.A. & Sankis, L.M., 2000. “Adult Psychopathology: Issues and Controversies,” *Annual Review of Psychology*, 51: 377–404.
- Woodward, J.F., 2003. *Making Things Happen*, Oxford: Oxford University Press.
- Zachar, P., 2014. *A Metaphysics of Psychopathology*, Cambridge, MA: MIT Press.
- Zisook, S., Shear, K & Kendler, K.S., 2007. “Validity of the bereavement exclusion criterion for the diagnosis of major depressive episode,” *World Psychiatry*, 6: 102–107.