EuJAP | VOL. 6 | No. 1 | 2010
ORIGINAL SCIENTIFIC PAPER
UDK 159.972
616.89:1

COMPLEX MENTAL DISORDERS: REPRESENTATION, STABILITY AND EXPLANATION

DOMINIC MURPHY
University of Sydney

ABSTRACT

This paper discusses the representation and explanation of relationships between phenomena that are important in psychiatric contexts. After a general discussion of complexity in the philosophy of science, I distinguish zooming-out approaches from zooming-in approaches. Zooming-out has to do with seeing complex mental illnesses as abstract models for the purposes of both explanation and reduction. Zooming-in involves breaking complex mental illnesses into simple components and trying to explain those components independently in terms of specific causes. Connections between existing practice and zooming-out are drawn, and zooming-in is criticised.

Keywords: explanation, mental illness, scientific representation, depression, reduction

1. Introduction

Theories of psychopathology have always struggled with the diversity of symptoms, course and underlying physical manifestations in many mental disorders. Most mental disorders are, in this sense, complex, the paradigmatic examples being the major psychoses and affective disorders. They have numerous diverse symptoms, which vary across patients and change over time, and lack any straightforward physical basis that remains constant across cases.

In this paper, I will discuss some recent approaches that aim to make this problem tractable – what are our best bets for dealing with the complexity of mental disorders? I will discuss two ways in which the complexity of mental illness makes a difference to our thinking about psychiatry. On the one hand, it raises questions of how we represent, or conceive of, mental illnesses – that is, what sorts of thing are they? A simple way of thinking about a physical illness is that it is a specific pathophysiology; a destructive process realized in body tissue with a distinctive associated set of signs and symptoms. Although some psychiatrists think of mental illness this way (Murphy 2009) we are a long way

from being able to represent mental illnesses in this fashion, since they vary greatly in their manifestations and we are mostly ignorant of their causes. Maybe mental illnesses aren't natural kinds at all, but something else. In discussing how we should represent mental illnesses, though, I hope to put worries about the underlying metaphysics aside. Whatever their real nature(s), we need a way to represent mental disorders as explananda. So the first question I will ask is, what do we explain when we explain a mental illness? I will treat this as a question about representation. We uncover lots of statistical relations among phenomena of clinical interest, and we need some kind of pattern or structure that makes these relations into objects of empirical inquiry. The metaphysics of mental illness is a separate question.

The second question I'll wonder about is: how do we explain mental illness? A traditional scientific response to complexity is idealization. Once we have an idealised representation of the phenomena we can go on to explain that. The typical explanatory strategy in the life sciences is to take a large scale phenomenon apart – to decompose it in to the bits that come together to produce it. To do this, we need to represent the phenomenon of interest in a way that lets us take it apart, and then show how the nature of the component parts, and the relationships between them, bring about the phenomenon we started with. In recent years philosophers have stressed the way in which this approach amounts to a search for mechanisms (Bechtel and Richardson 1993; Craver 2007; Schaffner 1993; Tabery 2009). Rather than seeing explanation as a search for laws, we look at various levels of explanation for the parts whose structure and activities explain the phenomena we started with. Philosophers disagree over exactly how to characterize mechanisms, but it is agreed that mechanisms comprise (i) component parts that (ii) do things. They dispute turns on whether the activities of the parts should be seen as constituents of a mechanism or just activities of the constituents (for a review, see Tabery 2004).

A mechanistic explanation shows how these parts and activities interact to give rise to the phenomenon we want to explain. Craver's (2007, 4-6) example is the mechanism by which neurotransmitters are released. This involves finding answers to questions such as why does depolarization of an axon terminal lead to neurotransmitter release, and why are neurotransmitters released in quanta? The answers involve pointing out various entities, such as specific types of calcium and various intracellular molecules, and showing their properties allow them to engage in patterns of activities. They interact with each other in a way that allows us to answer the questions and understand how interacting entities jointly give rise to the phenomenon that we want to explain. I will assume that this picture applies to mental disorders just as it does to the activities of the normal mind. I will not take a stand on what the relevant mechanisms are, or on how we should draw their boundaries, except to say that I assume that the mechanisms which explain mental illness will be those which explain normal psychology, albeit behaving in abnormal ways, and mediated specific pathogenic environmental and genetic causes.

Representation and explanation are linked. To explain a phenomenon we first have to see how to represent it. In this paper I will argue for a way of representing mental

disorders and a way of explaining them. The representational strategy involves seeing mental disorders as idealizations derived from statistically related phenomena. The explanatory strategy involves explaining the relations between components of the idealization and then showing how those relations are similar to relations in actual patients. I will call this the zooming-out approach, because it aims at an abstract representation of a diagnosis that prescinds from details of individual patients. By way of contrast, I will discuss Bentall's (2003) approach to these issues, which is, at first look, a competing approach. I call it 'zooming in'. Bentall is looking for phenomena that have enough stability across case histories to permit explanations that avoid variation altogether.

I will try to bring out the philosophical assumptions underlying zooming-in and criticize them. I will argue that zooming-in fails to confront the problem of the interrelations among the small scale kinds that it looks for. They are not stable enough to do the job Bentall wants. My second criticism of Bentall is that the connections he draws between representation and explanation are too tight. Because he represents mental illnesses as essentially psychological phenomena, he assumes that the explanations they receive must cite psychological mechanisms. I regard this as a constraint on causal explanation – essentially, that it must be intra-level, that we do not need to accept.

2. Complexity

Two patients may be diagnosed with the same condition yet share few symptoms. Even if they do share symptoms when diagnosed, the way they manifest the conditions may nonetheless diverge over time. DSM-IV-TR (American Psychiatric Association 2000, 356), for example, lists nine depressive symptoms, none of which is necessary for the diagnosis, and also notes that the condition can vary in course, with most cases going into remission after about four months, although in a small minority of cases the condition lasts for years (2000, 354). Distinguishing depression gets no easier in anatomical or physiological terms. Major depressive episodes are associated with several kinds of pathophysiology, but none of these are present in all depressive subjects, and no one of them is specific to depression (2000, 353). One leading contender for the physiological basis of major depression has been the monoamine depletion hypothesis. It holds that the underlying pathophysiology in depression is a shortage in the central nervous system of the neurotransmitters serotonin, norepinephrine or dopamine. But this hypothesis - which has led to the mass-marketing of SSRI drugs for the correction of serotonin imbalances - remains unconfirmed despite decades of effort (Licinio and Wong 2005, 78). The more distal causes of depression are also diverse. Twin studies, and observation of afflicted families, do show that there are genetic risk factors for depression. Optimistic judgements that the gene for depression has been found, however, have always been premature. The widely touted 5-HTT gene, for instance, is involved in building the receptors that help to control the neurotransmitters picked out by the monoamine hypothesis. A form of the 5-HTT gene is involved in emotional regulation and response

to threat (Hariri and Holmes 2006), but the gene is not, as far as we can tell to date, the gene for depression - although it probably makes a difference to some people in some contexts, so promoting their chances of depression.

The situation we face is not one in which there is a gene for depression that affects different people differently in different contexts. There are lots of genes and lots of other causes that interact, so that the causal complexity of depression is just as daunting as its diversity of symptoms. As Kendler et al. showed on the basis of extensive twin studies (2006, 115) major depression is a classic "multifactorial disorder". A range of factors affect your chances of contracting major depression. Genes certainly make a difference, but so do things like the extent of the child abuse you suffered, the state of your marriage and your history of substance abuse, as well as stressful environmental events, such as losing your job. The association between these stressful life events and major depression is, say Kendler and Prescott (2006, 281), at least partly causal, and the question of causes leads us into the question of explanation.

Complexity poses a problem for reductive explanations. The strategy we have developed for explaining complex systems is what Bechtel and Richardson (1993) termed decomposition and localization. We explain the behaviour of the components and aggregate those explanations to explain the whole. This simple reductionist approach runs into trouble when the details of the explanation depend on the way in which the parts are put together. In what Bechtel and Richardson call a component system, it is enough to know what the parts do and how they are put together; to explain the whole system we need to know the details of the organisation, but the parts themselves are unaffected by the way they are put together. In an integrative system, the behaviour of the parts themselves depends on the organization of the system, so that explanation needs to be top-down as well as bottom-up. The basic explanatory strategy remains that of taking the system apart and putting it back to together to explain its behaviour. So it is in that sense reductive. But it is not fully reductive, since the overall behaviour of the system is not just explicable in terms of its components. The arrangement of the components and the relations between them, especially feedback, are just as important as the properties of the individual components themselves.

It is very likely that the same will apply to understanding complex mental illnesses. The brain seems to be an integrative system, and certainly simple reductive attempts to explain mental disorders in terms of (e.g.) genetics have been unsuccessful. Kendler and Prescott (2006, 333-8) present an etiological model of major depression in men based on a large-scale twin study. Their data suggest that there are three main pathways to major depression. They depend on the interaction of genes with psychological characteristics like neuroticism and low self-esteem, as well as other mental illnesses. The model also incorporates accidents of biography, such as the early loss of a parent, sexual abuse in childhood, divorce, and insufficient social support. They also report (2006, 159) that episodes of "humiliation in a public setting" are among the most powerful predictors of major depression. They assumed (2006, 336) that the relations between variables in their model were additive and linear, but they acknowledge that this is known to be false. There is no straightforward pathway from gene to depression, but a complicated system

of causal relations running back and forth between a host of variables mediating between genes and phenotype. A simple reductive approach is unlikely to work, so what should we do?

3. Representation

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. Discovering statistical relations among phenomena is a large part of what science does; on this approach to representing mental disorder, we take the statistical relations that are uncovered and use them as a representation of what needs explaining.

This fits the picture that Kendler and Prescott present for depression quite neatly; we have a number of phenomena that cluster together in major depression, and we uncover the statistical relations among them. What are these relations, though? Kendler and Prescott's path models are designed to incorporate both causal and non-causal statistical relations between variables. Thagard says that his causal-statistical networks exhibit "a kind of narrative explanation of why a person becomes sick" (115). They incorporate information about the typical ways a disease 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.

In neither case, though, do the networks specify how, or whether, the relations among phenomena produces the effects. The way in which humiliation acts on a vulnerable system to produce depression is not mentioned in the model. The precise biochemical, information-processing or physiological mechanisms that explain the outcome are filled in as we interpret the relations between parts of the model. At the outset, then, no particular assumption about the nature of these causal connections needs to be made.

A disease network is really a descriptive model of a disease – a representation 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 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. Rachel Cooper (2007) argues that to the extent such narratives work, it is because they instantiate what she calls "natural history explanations". 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 say that Laura hears voices because she has schizophrenia. Our confidence that she is schizophrenic warrants pessimistic predictions about her future that depend on our being able to place her in a certain kind.

Cooper suggests (2007, 174, n.2) 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 when the details of the underlying causal structure evade us, because the story we tell in each case will be different. Sydenham used the logic of natural histories in this way to argue that smallpox and cowpox are not the same condition in the seventeenth century, and Kraepelin applied it to psychiatry as the basis for differential diagnosis, for example between dementia praecox (schizophrenia) and other forms of insanity (1899, 173-5). It is a familiar idea that DSM-IV'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-Krapelinian 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.

We assume that the disease narrative reflects that underlying causal structure. In psychiatry, sceptics wonder if the degree of variation makes the predictions too unreliable for this approach to really work.

4. Explanation. 1: Zooming-out

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 of a disorder, and only that disorder. A more precise way to think of the exemplar is as one of Thagard's network, or a specific instantiation of a model in Kendler and Prescott's sense. The network doesn't just give us a qualitative understanding, it specifies relations among phenomena in the exemplar quite precisely, and therefore directs our attention to the key features we need to explain. This will be more important depending on how stable the relations in the exemplar tend to be. Stability in this sense (see Woodward forthcoming for a more technical treatment) is a kind of counterfactual dependence – a relationship is stable in so far as it would continue to hold even if background factors were different. Kendler (2005, 1248), discussing the idea of "genes for" disorders, calls this a noncontingent association, by which he means "that the relationship between gene X and disorder Y is not dependent on other factors, particularly exposure to a specific environment or on the presence of other genes." A classic Mendelian disorder would be an example of a noncontingent association between a gene and a disorder. On the other hand, to take an example Kendler uses to make a slightly different point (2005, 1249-50), suppose we have a putative gene for liking Mozart, but the causal pathway runs from

allele to Mozartophilia via perfect pitch. Then we might want to say that what we really have is a gene for perfect pitch. If circumstances had been different the subject would still have perfect pitch, but would perhaps have been introduced to different music early on, and grown up loving the Jesus and Mary Chain instead.

If the exemplar incorporates stable relationships then, in explaining them, we can hope to point to *robust processes* (Sterelny 2003, 131-2, 207-8). Robust processes are repeatable or systematic in various ways, whereas the actual processes that occur as a disorder unfolds in one person might be idiosyncratic or unstable. If circumstances vary just a little, unstable processes will vary too, and we will therefore not be explaining robust and stable features of a mental illness, but unstable ones. These unstable features might be culturally specific or even peculiar to a particular individual one.

I assume that the ultimate goal is causal understanding of a disorder, and I will have a bit more to say about this below. We build a model to serve this end. It aims to represent the pathogenic process that accounts for the observed phenomena in the exemplar. Then we show how those relations, in their turn, resemble the ones that exist in the actual condition as realized in particular patients.

Our knowledge of the pathophysiology is typically scantier in psychiatry than in general medicine, in which we have very often developed our models to such an extent that we can think in terms of just a (perhaps partly) completed model that shows how the symptoms depend on unobserved processes But logically there are (and historically there have been) at least these steps: first, the study of patients; second, the construction of an exemplar by isolating those features which the patients share; third, the explanation of why the exemplar takes the form it does; fourth, relating the exemplar to its realization in individuals.

To do this, we represent the pathogenic process that accounts for the observed phenomena in the exemplar, namely the stable relations among variables in the model. To explain an actual history in a patient is to show how the processes unfolding in the patient resemble those that 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. Like Thagard's network, an exemplar is what needs explaining. But I do not presume that the relata in the exemplar are causally rather than probabilistically related: exemplars display relations among phenomena but they are not, by themselves, explanations. The underlying causal relations do the explaining.

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 qualitative forerunner of modelling, in which essential factors are isolated and inessential ones put aside (Engerman 2000).

Ghaemi, though, locates it in a tradition of hermeneutic understanding that is most closely associated in psychiatry with Karl Jaspers. The hermeneutic approach looks for psychologically meaningful 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.

Ghaemi's point is a sensible one in the context of clinical application – we do need to understand the lived experience of people with mental disorders, and of course any given patient will appear to us as a specific individual in specific circumstances. We do not see types, but individuals. However, to understand the individual, we must relate it to a type. However, it is not necessary to think of this as tied to the phenomenological or hermeneutic traditions. The basic idea is simply idealization, which is the standard scientific response to complexity. Wachbroit (1994, 588) argues persuasively that when medicine or physiology says that an organ is 'normal', the relevant conception of normality "is similar to the role pure states or ideal entities play in physical theories." Such an idealization represents actual organs or systems in unperturbed states (cf. Ereshefsky, 2009). To understand a real case we add information to develop a model that resembles actual hearts (Wachbroit 1994b, 589). For instance, Gross (1921) was able to establish post mortem that anastomotic communication between main arteries increases over a typical lifespan, thereby establishing that we need to model younger and older hearts differently. The point of such idealisations is not to represent the statistically average heart, but to describe hearts in a way that allows departures from the ideal to be recognised and to serve as template from which more realistic models can be built.

In commenting on Murphy's 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 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. In Mitchell's picture, we do not start with an ideal representation of the whole system, but a set of partial representations that we then put together. However, there is no reason why zooming-out cannot incorporate Mitchell's basic idea; even if the representational strategy of an exemplar-based approach is different, it is entirely consistent with Mitchell's explanatory procedure. If a mental disorder is complex, as it is pretty much bound to be, we might need both her models and Murphy's as circumstances dictate (Mitchell 2009, 131). Psychiatric research should aim to be pluralist about explanation and combine elements drawn from different explanatory styles as we learn more about what works when we try to figure out mental illness. Complexity is likely to require both top-down and bottom-up approaches, as we saw above when discussing Bechtel

and Richardson. The components in a psychiatric system cannot be understood in isolation but will typically take different values depending on their interactions with other components. So we need models that both work on the component ingredients of the whole system and take into account their relations with each other and with the system as a whole. It is the fact that these explanations need to be both bottom-up and top-down that is the main drawback, I think, to Bentall's zooming-in strategy, to which I now turn. I have two main objections to Bentall; he mixes up representation and explanation, and his strategy is too bottom-up to work.

5. Explanation. 2: Zooming-in

I have discussed idealisations as a response to complexity. Another way to make psychiatric variation empirically tractable is to look for some stable phenomena that do not vary. One such approach looks for the smallest units of psychiatric interest that repeat reliably across patients. These are not likely to be found at the level of diagnoses, so we must search for smaller units of explanation. This approach aims at finding natural kinds in psychiatry, but is sceptical of many current diagnoses. If we are to carve nature at its joints, we must descend to a lower level than that of current diagnoses. This zoom-in approach is exemplified by Bentall (2003). He argues that diagnoses like depression or schizophrenia have proved useless for research in the face of all the variation that patients exhibit. Instead of thinking in terms of diagnostic categories, Bentall sees cases as mosaics of symptoms like hallucinations, which recombine in idiosyncratic ways across patients, and 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 occur in shifting conjunctions. These are distinct psychotic phenomena that should be approached separately and treated as distinct symptoms or complaints.

Bentall thinks that "we should abandon psychiatric diagnoses altogether and instead try to understand and explain the actual experiences and behaviours of psychotic people" (2003, 141). But in fact what is he is trying to do is relocate diagnosis at more reliable level. For example, he (ch. 15) objects to inferring thought disorder from disordered speech. Disordered speech he thinks of as a failure of communication, especially likely in emotionally aroused subjects. 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. The zooming-in approach assumes that people who are diagnosed as schizophrenics are indeed mentally ill, but not in the same way; there is no shared diagnosis here. Rather, there are problems

that cluster together in some unpredictable ways, so that one patient may suffer from A,B,C & D, while another has A,B & E, a third C,E & F, still another D, F & G, and so on. Where a current diagnosis might treat all members of this class as sharing a mental illness of diverse manifestation, the zooming-in approach says that the real picture is just what I just described; a collection of people with a partly overlapping pool of symptoms but no one diagnosis in common. Indeed, Bentall denies that there is a useful distinction between schizophrenia and manic depression, on the grounds that there are too many overlaps in symptoms, outcomes genetics and drug responses among schizophrenics and bipolar patients. The patients have psychosis in common – but that is not a helpful category when it comes to generalizing across them. (In order to have a helpful shorthand description I will refer to the diverse problems that the zooming-in approach looks for as "component psychoses".)

A general theory of schizophrenia would have too many qualifications and varieties to transfer in this way: it will not generalize across all patients. Where the zooming-out approach looks for an ideal model that more or less resembles subjects with the diagnosis, Bentall's zooming-in approach seeks projectible kinds. If the manifestations of mental illness look too diverse, says the proponent of zooming-in, then we can look instead for component psychoses that stably replicate across patients. But 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 is 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. Nonetheless, there will always be differences across patients in the way in the manifestations of 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 if idealized patients, rather than idealized problems. As far as the representation of a disorder goes, then, Bentall needs idealisation too.

When we turn to explanation, Bentall's picture seems to privilege psychology, and he seems to think that a picture of explanation falls out of his picture of representation. But in fact, the view of explanation he is committed to only follows if a number of assumptions about explanation are made, and those assumptions are not only not defended, they are unwarranted. I will now try to spell out why I say this, but I will also argue that the zooming-in strategy, considered as a representational device, ends up turning into an exemplar approach.

So how does the representation work? The representation of a mental illness, on

Bentall's account, is a network of psychological problems such as paranoia, theory of mind deficits, feelings of hopelessness, hallucinations and incoherent speech. Bentall represents these cognitive and affective ailments in the familiar "boxological" style as boxes connected by arrows showing the relations between psychotic phenomena. There are other boxes, too; these are representations of other processes that are not themselves observable psychological complaints but are part of the sequence of processes out of which those complaints emerge. So, for example, 'stored knowledge about the self' feeds into "paranoid beliefs" via 'current beliefs about the self' and 'external personal attribution' (and there is a feedback loop between those systems) (2003, 410).

This is all very similar to Thagard's picture, except that the phenomena are exclusively psychological. Bentall denounces genetic approaches as too reductive, but it is the genetics rather than the reductionism he objects too. He opposes what he calls the "Kraepelinian paradigm" in which a genetic etiology is assumed to lead to a pathological anatomy which gives rise to a set of symptoms. His rival approach presumes that mental disorders are psychological phenomena. That is fair enough, in so far as most symptoms of mental illness are either descriptively psychological or can, probably, be traced to underlying subpersonal information-processors. Bentall's picture of psychology is scientific, drawn from cognitive psychology. It is not folk psychological. But he seems to assume that psychological ailments so conceived must have psychological explanations via decomposition into exclusively psychological components. Once the complaints have been explained in terms of psychological processes, there is nothing left that also requires an illness, because the complaints "are all there is" (2003, 405).

But even if the complaints are all there is, why suppose that all there is to explaining them is psychology? Bentall notes the connectedness between psychological phenomena (2003, 414) and argues that functional relationships between psychological systems and symptoms give is a much better way to understanding the clustering of psychiatric complaints than the Kraepelinian paradigm can provide. But why suppose they are in competition. The Kraepelinian paradigm, according to Bentall, hypothesizes a path from gene to pathophysiology and on to symptom. It seems that this approach should be consistent with Bentall's view that the symptoms, psychologically described, cluster together in connected ways. Indeed, that picture emerges from Kendler et al.'s work and was assumed by the idea of an exemplar, in which psychological, genetic, environmental and other factors may all be incorporated.

Yet Bentall treats psychological decomposition as a competitor to the multi-level causal explanations that the exemplar approach relies on. This may be because of an equivocation in Bentall's account between representation and description.

Bentall represents mental illnesses as collections of psychological phenomena connected by what he calls "functional relationships'. Bentall borrows the concept of a functional relationship from algebra (2003, 408): if there is a functional relationship between two variables we can graph the relationship between them. So, as x changes, values of y vary in step, as described by the equation expressing the relationship.

However, Bentall also assumes that these functional relationships are also causal ones, or at least that the functional relationships he picks are causal - and that just does not seem to follow. Essentially, Bentall assumes that once we have a representation, we can read off the causal relationships from the statistics, and the causal relationships will all be psychological.

What we have, as on Thagard's account, is a set of statistical relations between phenomena, and those relations need explaining. Although there is a causal explanation of why we see a given pattern of correlations, that causal explanation is, of course, not necessarily expressed by the correlations themselves. This is easiest to grasp in the case of common causes; it may be that there is a functional relationship between being a registered Republican and opposing gay marriage, but that doesn't mean that joining the Republican party makes you into someone who dislikes the idea of married gay people. It could be that some prior trait (readers can insert one here that expresses their idea of Republicans) leads to both membership in the Republican party and hostility to gay marriage.

The worry that some people are bound to have, as Bentall sees (2003, 405), is that there might be an additional etiological story to tell, about "how the complaints came into being in the first place". Bentall's answer is that his approach can incorporate a variety of developmental, genetic and environmental influences (2003, chs. 16-18). But this is agreed on by many other theorists. As we noted above, there is little reason to expect a simple genetic etiology for any major mental illness. Bentall argues that on his account it is fruitless to ask for the ultimate cause of one or more psychotic complaints. This too is a commitment of Kendler and Prescott's story, in which complex diseases have many causes. There is absolutely nothing in Bentall's model to rule out the idea that genetic factors produce a constitutional biological vulnerability or abnormality that interacts with environmental factors to produce just that set of functional relationships among psychological processes that he depicts. So why is the stress on functional relationships as opposed to other factors?

Bentall's picture is relentlessly psychological, although perhaps the logic of his picture is consistent with a more reductive account. On the other hand, his vision of inquiry appears to leave out developmental information: it takes a snapshot of psychological relations among cognitive systems in a mature brain and explains psychopathology in terms of relations among those systems or failures within them. The worry is that this leaves no room for other sorts of causal explanations, such as developmental ones. It may be that part of the explanation for why the mature brain has the features it does lies in developmental processes that Bentall's picture misses.

Bentall argues (2003, 405-6) that mental disorders can be completely explained by citing psychological processes, so it follows that other factors, such as genes or environmental influences or brain development, are in some way unexplanatory. I think this is because he is looking for clean causal stories; he complains in several places that the effects of genes and other factors are simply too non-specific to be useful. The idea seems to be that because the relations between symptoms and genes is typically nonspecific, since

there is no one-to-one relation between genes and symptoms. However, Bentall does seem to think that there is a specific relation between symptoms and a particular network of psychological systems, so it is that relationship, between symptoms and psychology, which is explanatory. In that sense, Bentall's picture is a simple reductive one, in which all symptoms need to be explained via a set of specific proximal relationships, thus allowing more distal factors to be downgraded.

This involves a big philosophical assumption – essentially, the commitment that mental illnesses are decomposable systems, despite their complexity. The way Bentall deals with complexity is to argue that there are no complex psychoses, but collections of non complex symptoms. A different possibility would be that the component psychoses are not really independent of each other. Instead, they represent different ways in which a causal pathway can unfold. Imagine that the developing organism is mostly buffeted against change, in the sense that a human being can end up with a healthy brain from many different starting points – but some starting points move the system outside the space within which development is buffered. Within that wider space, initial differences that are close together in the wider space in development can nonetheless take the system to final states that are far apart. We could use this insight to preserve the idea that 'schizophrenia' names a developmental problem, and even track down the problem, but having that explanation of what is going on would still leave us without very detailed inductive knowledge of the phenomena. At that point we might seek to use Bentall's methods to describe the specific psychological phenomena caused by the developmental process that leads to schizophrenia. But it makes a difference if some component psychoses are a result of that processes and others are not – they come about because of stroke, say. The same psychological story could apply to a psychological system – say, a language system underlying a pathology of speech or comprehension - even though the damage has been caused in different ways. And those distal causal differences might be important.

Let me take stock – I am arguing that Bentall's approach, despite being located at the psychological level, shares the reductionist logic of the gene-centred approaches he criticises. It seeks to explain complex mental phenomena by arguing that they are really simple; the way to understand them is to decompose them into units under the control of specific causes. It's just that those specific causes will be psychological rather than genetic.

Can this approach work? It could be, after all, that decomposition and localization won't work for genes because the pathways from genes to symptoms are too baroque, but that it might work for psychology, because the pathways there are short and specific. In fact, though, Bentall's own account shows us why that won't do. He notes (2003, 412-3) that there are often connections between different psychotic symptoms and other underlying psychological phenomena. He notes the mutual feedback relations between delusions and hallucinations and argues for similar connections between paranoia and incoherent speech, mediated by working memory problems. The connections between component psychoses are very numerous and complex. There are feedback relations between them which means that when the component psychoses are linked together, as they typically are in the patients Bentall worries about, the particular explanations of symptoms that

zooming-in has given to us will need to be altered in the light of relations to other phenomena. That is just to say that the system is not fully decomposable at all, since the relationships between components, and between them and the overall system, make a difference to the behaviour of the system. If mental illnesses are complex in this way, as they seem to be, then zooming-in can't work, because it amounts to a bet that at some level, simple reductionism will work.

Bentall's treatment has many virtues that I have overlooked so far; the stress on cognitive psychology is an important corrective to many accounts that leave out psychological mechanism when trying to explain the abnormal mind. And his discussion of functional relations among psychological phenomena is useful. But as a representational strategy, there seems to be no reason to limit the exemplar to purely psychological variables, since so many other kinds are important. And as an explanatory strategy, zooming-in looks to have made the wrong bets.

REFERENCES

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

Bechtel W. & Richardson R. C. 1993. *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton University Press, Princeton.

Bentall, R. 2003. *Madness Explained*. London: Penguin.

Cooper, R. 2007. *Psychiatry and Philosophy of Science*. Stocksfield: Acumen.

Craver C. F. 2007. *Explaining the Brain*. Oxford University Press, New York.

Engerman, S. L. 2000. Max Weber as Economist and Economic historian. In S. Turner (ed) *The Cambridge Companion to Weber*. Cambridge: Cambridge University press; 256-71.

Ereshefsky M. 2009. Defining 'Health' and 'Disease'. *Studies in the History and Philosophy of Biology and Biomedical Sciences* 40, 221-7.

Ghaemi, S. N. 2003. *The Concepts of Psychiatry*. Baltimore: Johns Hopkins University Press.

Hariri A. R. & Holmes A. 2006. Genetics of emotional regulation: the role of the serotonin transporter in neural function. *Trends Cogn Sci* 10: 182-91.

Kendler K. S. 2005. A gene for…: the nature of gene action in psychiatric disorders. *American Journal of Psychiatry* 162: 1243-52.

Kendler, K. S., Gardner, C. O. & Prescott, C. A. 2006. Towards A Comprehensive Developmental Model for Depression in Men. *American Journal of Psychiatry* 163: 115-24.

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.

Kraepelin, E. 1899. *Psychiatry: A Textbook for Students and Physicians*, 6th ed. Repr. Canton, MA; Science History Publications 1990.

Licionio, J & Wong, M-L. 2005. *The Biology of Depression*. Hoboken: Wiley VCH.

Mitchell, S. 2003. *Biological Complexity and Integrative Pluralism*, Cambridge: Cambridge University Press.

Mitchell, S. 2009. Taming Causal Complexity. In *Philosophical Issues in Psychiatry*, ed. Kenneth Kendler and Josef Parnas. Johns Hopkins University Press. 125-31.

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

Schaffner K. F. 1993. *Discovery and explanation in biology and medicine*. Chicago:

The University of Chicago Press.

Sterelny, K. 2003. *The Evolution of Agency and Other Essays*. Cambridge: Cambridge University Press.

Tabery, J. 2004. Synthesizing Activities and Interactions in the Concept of a Mechanism", *Philosophy of Science* 71: 1-15.

Tabery, J. 2009. *Difference mechanisms: explaining variation with mechanisms*. *Biology and Philosophy* 24: 645-64.

Thagard, P. 1999. *How Scientists Explain Disease*. Princeton: Princeton University Press.

Received: April 9, 2010
Accepted: April 15, 2010

Unit for History and Philosophy of Science
University of Sydney
Carslaw F07
Camperdown NSW 2006 Australia
dominic.murphy@sydney.edu.au