Mental Illness from the Perspective of Theoretical Neuroscience

Paul Thagard

ABSTRACT Theoretical neuroscience, which characterizes neural mechanisms using mathematical and computational models, is highly relevant to central problems in the philosophy of psychiatry. These models can help to solve the explanation problem of causally connecting neural processes with the behaviors and experiences found in mental illnesses. Such explanations will also be useful for generating better classifications and treatments of psychiatric disorders. The result should help to eliminate concerns that mental illnesses such as depression and schizophrenia are not objectively real. A philosophical approach to mental illness based on neuroscience need not neglect the inherently social and historical nature of mental phenomena.

THIS ARTICLE ADDRESSES four major problems in the philosophy of psychiatry: objectivity, classification, treatment, and explanation. The objectivity problem revolves around whether mental illnesses such as schizophrenia and depression are real physical disorders or merely social constructions. The classification problem concerns the validity of taxonomies of mental illnesses, such as those provided by the widely used *Diagnostic and Statistical Manual of Mental Disorders* (DSM). The treatment problem centers on whether it is possible to improve upon current practices that are often haphazard in prescribing drugs and

Department of Philosophy, University of Waterloo, Waterloo, ON N2L 3G1, Canada.
E-mail: pthagard@uwaterloo.ca.

The author wishes to thank Brandon Aubie, Abninder Litt, and Fred Tauber for comments on an earlier draft. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Perspectives in Biology and Medicine, volume 51, number 3 (summer 2008):335–52

other remedies for mental illnesses. Finally, the explanation problem concerns the gap between neuroscientific accounts of how the brain works and the mental experiences that are part of psychiatric illnesses.

I argue that a solution to the explanation problem has the potential to solve the other three problems as well. Improved theories about the neural mechanisms that produce mental experience should help to establish the biological objectivity of psychiatry, provide a more reliable basis for classifying psychiatric disorders, and generate more reliable treatment regimens. As the most likely source for the needed theories, I look to the emerging field of theoretical neuroscience, which uses mathematical analysis and computational modeling to characterize neural mechanisms that can explain a broad range of mental phenomena.

Because this essay was prepared for the conference "Future Horizons for the Philosophy of Medicine?" I begin with a brief discussion of the nature of philosophy and its application to medicine. I argue against conceptions of the nature of philosophy that take its intellectual role to the pursuit of necessary truths or conceptual clarifications. These conceptions see philosophy as standing above or below the sciences, whereas I see it as standing side by side, hand in hand, aiming to develop general theories about the nature of reality, knowledge, and right and wrong. This article illustrates this naturalistic approach to the philosophy of medicine, using theories from cognitive science and neuroscience to address philosophical issues about the nature of medical knowledge of mental illness.

WHAT IS PHILOSOPHY OF MEDICINE?

What is philosophy, and how is it related to science? To answer these questions, I want to reject two highly influential approaches to philosophy, the Platonic and the Wittgensteinian. According to Plato, the kind of empirical knowledge gained by science is a mere shadow of fundamental knowledge that can be gained by reflection on the Forms, which are abstract ideas that capture the essence of things. From this perspective, philosophy should aim for truths that are a priori (gained prior to any sense experience) and necessary (true in all possible worlds, not just ours). Many philosophers have sought such necessary truths, from Plato to Aquinas to Leibniz to Kant to Kripke. The problem is that no one seems to have found any: the best candidate for universal adoption that I know of is Hilary Putnam's (1983) rather trivial "truth" that not every statement is both true and false. In particular, if anyone in the philosophy of medicine has arrived through a priori reflection at some necessary truths, the news does not seem to have hit the journals.

Wittgenstein propounded a view of philosophy much less ambitious than the Platonic aim to find necessary truths. In the *Tractatus Logico-Philosophicus* (1971), he stated:

> Philosophy is not one of the natural sciences.

> (The word "philosophy" must mean something which stands above or below, but not beside the natural sciences.)
>
> The object of philosophy is the logical clarification of thoughts.
>
> Philosophy is not a body of doctrine but an activity.
>
> A philosophical work consists essentially of elucidations.
>
> Philosophy does not result in "philosophical propositions," but rather in the clarification of propositions. (p. 49, propositions 4.111–4.112)

On this view, the role of philosophical activity is the clarification of concepts using logic, or, in the later Wittgenstein, ordinary language. Whereas the Platonic view of philosophy places it above the sciences, establishing more solid kinds of knowledge, the Wittgensteinian view places philosophy below the sciences, at most clarifying and elucidating its results. Philosophy "leaves everything as it is" (Wittgenstein 1968, p. 49e, section 124).

I have many objections to this view of philosophy, but will here mention only one. The view that philosophy is concerned with conceptual clarification rests on a false view of concepts that supposes that they have meaning independent of the theories in which they are embedded. There are both philosophical and psychological arguments that the meaning of concepts is utterly bound up with theories that contain them (Medin 1989; Quine 1963). It is impossible to undertake conceptual clarification without critically examining the truth of the scientific and philosophical theories that employ the concepts. To take an example from the philosophy of medicine, it is pointless to clarify concepts such as "disease" without scrutinizing explanatory theories of what causes diseases. It turns out for theoretical reasons that the idea of a conceptual confusion is itself a conceptual confusion. Hence, the meager ambitions of the Wittgensteinian approach to philosophy are unachievable.

I prefer a naturalistic approach to philosophy, which in the past century is most associated with the philosopher W.V. O. Quine, although he has had many illustrious predecessors. Thinkers whose philosophical investigations have proceeded in intimate connection with scientific research include Aristotle, Locke, J. S. Mill, Peirce, Dewey, and Bertrand Russell in his later work. The three main branches of philosophy are epistemology (the theory of knowledge), metaphysics (the theory of reality), and ethics (the theory of right and wrong). From the naturalistic perspective, epistemology is tightly connected with psychology and the other cognitive sciences, and metaphysics draws on all the sciences, from research in physics relevant to the nature of space and time to research in psychology concerning the relation of mind and body. Ethics is not a stand-alone subject that searches for a priori moral truths or elucidation of ethical concepts, but rather collaborates with relevant sciences such as psychology, economics, biology, and anthropology to develop theories of right and wrong.

The main differences between philosophy and the sciences are that philoso-

phy is more general and more normative. Philosophy is more general in that it deals with knowledge and reality in all domains, whereas the sciences try to develop knowledge of reality in particular spheres such as terrestrial life, which is the concern of biology. Philosophy is more normative in that its investigations in epistemology and ethics are unavoidably concerned not just with how things are, but also with how things ought to be. Philosophy addresses not only how we think and act, but also how we can think and act *better*. Some sciences also have a substantial normative component, for example educational psychology, which tries to find better ways of learning and teaching. Medicine also has a large normative aspect, looking for ways to improve health as well as explain disease. Even physics, when it is applied in engineering, can be concerned with how to do things better. But in all of these sciences most research is directed less toward normative issues than toward descriptive ones, such as how the mind works, how diseases arise, and how physical forces operate. In philosophy, however, normative issues are unavoidably central in ethical and epistemological investigations.

I will now try to show how naturalism can contribute to the philosophy of medicine. After reviewing what theoretical neuroscience is increasingly contributing to the understanding of mind, I will project what it can potentially offer for central problems in the philosophy of psychiatry. Finally, I defend a neuroscientific, naturalistic approach to mental illness from the charge that it ignores the inherently social and historical nature of mental phenomena.

THEORETICAL NEUROSCIENCE

Just as medical specialties such as cardiology and dermatology depend on biological knowledge of organs like the heart and skin, so psychiatry depends on neuroscience, which investigates the brain and nervous system. Neuroscience uses experimental techniques such as single-cell recording and brain scanning to make generalizations about the operations of neurons and brain regions. Bringing order to the masses of generalizations compiled by experimental neuroscience requires theoretical investigations of the mechanisms by which brains accomplish such cognitive functions as perception, memory, and problem solving.

In contemporary cognitive science, there are two main theoretical approaches (for reviews, see Thagard 2005, n.d.-a). The first, symbolic approach grew up in the 1950s and 1960s at the intersection of psychology and artificial intelligence. Researchers such as Herbert Simon, Allan Newell, and John R. Anderson have developed computational models of how people can use rules to represent the world and solve complex problems. The second, connectionist approach arose in the 1980s, with arguments that various kinds of thinking are better explained in terms of sub-symbolic representations consisting of multiple, interacting, neuron-like units that compute in parallel, roughly the way that that the brain does. However, connectionist models are only crudely brain-like, neglecting many features such as modular organization and the use of very large numbers of neurons

and connections. Theoretical neuroscience differs from connectionism in attending much more closely to the structures and processes that operate in the brain (Dayan and Abbott 2001; Eliasmith and Anderson 2003; O'Reilly and Munakata 2000). The study of rule-based symbolic systems is also moving closer to neuroscience, as investigators attempt to tie such systems' functioning with neural operations (Anderson et al. 2004).

Theoretical neuroscience uses mathematical and computational methods to characterize the operations of nervous systems at many levels, from individual neurons to interacting brain areas. Dayan and Abbott (2001) describe how theoretical neuroscience uses both descriptive models, which summarize experimental data, and mechanistic models, which use known anatomy and physiology to explain how nervous systems operate. Philosophers of science are increasingly recognizing the importance of mechanistic explanations, in which an event or regularity is explained by showing how it is produced by a mechanism, or a system of components whose properties, relations, and interactions produce regular changes (Bechtel and Abrahamsen 2005; Bechtel and Richardson 1993; Craver 2007; Machamer, Darden, and Craver 2000; Salmon 1984). Mechanistic explanations are ubiquitous in medicine, where a disorder is explained by describing how multiple causes interact with biological systems to produce the symptoms and development of a disease (Thagard 1999, 2006).

Theoretical neuroscience uses mathematical techniques drawn from linear algebra, calculus, and statistics to characterize neural mechanisms. The primary components in these mechanisms are neurons, cells whose most important property is their ability to generate electrical signals in response to chemical inputs from other cells. When a neuron's electrical potential reaches a threshold, it fires (spikes) by sending chemical inputs to other neurons via synaptic connections. Neurons encode sensory stimuli by becoming tuned to fire when particular stimuli are presented. Differential equations elegantly capture the process by which a neuron accumulates electrical signals from other neurons and passes them on to others, but the behavior of networks of neurons quickly becomes too complex to be handled by mathematical analysis alone: computer models are required to predict and explain the behavior of many interacting neurons. Neural networks are able to adapt to new stimuli by learning, which consists of adjusting the strengths of the synaptic connections between neurons in response to environmental conditions. In sum, a neural mechanism is a network of neurons whose synaptic connections and learning capabilities enable the network to encode stimuli and adapt to changes in the environment.

Theoretical neuroscience is a young field that began to flourish only in the 1990s. Nevertheless, impressive progress has been made in explaining many psychological phenomena concerned with perception, learning, memory, and decision making. I will now consider the potential of theoretical neuroscience for providing mechanistic explanations of mental illness.

EXPLAINING MENTAL ILLNESS

A biological mechanism explains the normal functioning of an organism by showing how a system of interacting components enables the organism to operate in its environment. A disease is a breakdown in normal functioning that impedes the organism's performance. Breakdowns arise because internal or external factors affect the properties and interactions of the components of a system in such a way that they no longer produce the regular changes that the organism needs to function well in its environment. For example, the circulatory system consists of a set of components—the heart, veins, arteries, and blood—that interact to provide nutrients to the rest of the body. This mechanism is susceptible to many kinds of breakdown, such as defects in the heart valves, blockage in the arteries due to plaque and blood clots, and abnormal growth of blood cells. These breakdowns can arise because of many kinds of interacting causal factors, from internal ones such as defective genes to external ones such as infectious agents.

Similarly, the explanation of mental diseases requires specification of the normal functioning of the brain and other relevant organs, along with precise description of the different kinds of breakdown that can impede mental functioning. The causes of many such diseases are known approximately; for example, epileptic seizures are the result of disturbances of the normal electrical functions of the brain produced by many factors such as brain tumors. But much remains unknown about how particular kinds of brain malfunction produce particular kinds of mental symptoms, such as the hallucinations and delusions that afflict patients with schizophrenia.

Theoretical neuroscience should help to fill these explanatory gaps by developing computational models of normal and diseased brain operation. Normal operation is characterized by a network of neurons whose interactions carry out standard functions such as perception, inference, and memory. Diseased operation is characterized by alterations of the standard network that produce limitations in these functions. The crudest kind of alteration would be to lesion the simulated network by deleting a subset of artificial neurons, corresponding to real lesions produced in human brains by strokes or assaults. Most kinds of brain disease are much more subtle and require electrochemical interactions among large numbers of neurons. There is more to disease than abnormality, which can include relatively benign conditions such as left-handedness and synesthesia. Mental illnesses involve malfunctions in brain mechanisms that are harmful.

An impressive recent illustration of how computational modeling can contribute to the understanding of mental disorders is the account of attention deficit/hyperactivity disorder (ADHD) by Frank et al. (2007). Diagnostic criteria for ADHD include the inability to attend to detailed tasks such as schoolwork, excess physical activity, and impulsivity. To explain such symptoms, Frank and his colleagues employed a computational model of dopamine function and dys-

function that was previously applied to learning, decision making, and Parkinson's disease (Frank 2005; Frank and Claus 2006). Dopamine is also a key factor in neurocomputational models of schizophrenia (Cohen, Braver, and Brown 2002; Grossberg 2000a, 2000b), and theoretical neuroscience is just beginning to tackle the task of developing detailed models of the most dramatic symptoms of schizophrenia: hallucinations and delusions (Behrendt and Young 2004; Hoffman and McGlashan 2001; Kapur 2003; Smith et al. 2007).

Theoretical neuroscience is also beginning to shed light on the most common mental illness, depression. Neural mechanisms relevant to depressed mood seem to include the activity of serotonin and other neurotransmitters, and also the development of new neurons, or neurogenesis (Kramer 2005). Evidence is mounting that a major neural cause of depression is stress-induced decreases in neurogenesis in the dentate gyrus, part of the hippocampus (Jacobs, van Praag, and Gage 2000). Becker and Wojtowicz (2007) have proposed a mechanism by which altered neurogenesis affects mood state through contextual-memory formation and the generation of appropriately contextualized responses to emotional stimuli.

But what about the major symptom of depression, feeling sad? Graham and Stevens (2007) argue that what makes an illness *mental* is its effect on conscious representational experience, which is important for both emotional reactions such as sadness and perceptions such as hallucinations. Philosophers who defend a dualism of mind and body have claimed that there is an unfillable explanatory gap between neuroscience and conscious experience, which must be understood phenomenologically rather than scientifically. To fill this gap, theoretical neuroscience has to draw connections, both between neural mechanisms and behavior and between mechanisms and conscious experience. One effort in this direction is the model of emotional consciousness proposed by Thagard and Aubie (n.d.), which uses interactions among many brain areas—including the prefrontal cortex, thalamus, amygdala, and dopamine reward systems—to explain diverse aspects of emotional experience, including differentiation, integration, intensity, valence, and change. These brain areas integrate perceptions of bodily states of an organism with cognitive appraisals of its current situation. Emotions are neural processes that represent the overall cognitive and somatic state of the organism, and conscious experience arises when neural representations achieve high activation as part of working memory. So far, the model has only been applied to normal mental functioning, but it should be possible to consider how breakdowns in its operation can lead to affective disorders such as depression.

We can begin to sketch the mechanisms that might connect neurogenesis with the feelings associated with depression. When stress or other causes leads to insufficient formation of new neurons in the hippocampus, people are unable to encode experiences flexibly and generate a broad range of explanations for them. An ambiguous stimulus, such as a person being a bit unfriendly, will there-

fore be interpreted in a manner consistent with other negative experiences, such as other people being unfriendly. People will then find all these experiences emotionally coherent with such negative beliefs as that they are inherently unlikeable, which will promote further negative interpretations of new experiences, increasing sadness.

Theoretical neuroscience is only beginning to solve the problem of explaining mental illness by showing how neural processes can produce abnormal behavior and aberrant conscious experience, and the research program I envisage will require decades or centuries. But this section has attempted to show how, in principle, the explanatory gap between neuroscience and psychopathology can be filled by theoretical investigations that employ computational models. Hence there is a possible route to solving the explanation problem in the philosophy of psychiatry, with correlative solutions to the classification, treatment, and objectivity problems. However, I must emphasize that explanation of mental disorders should not be restricted to neural mechanisms but needs also to attend to psychological and social ones.

CLASSIFYING MENTAL DISORDERS

The standard way of classifying and diagnosing mental disorders today is the DSM, produced by the American Psychiatric Society (APA 2000). A similar document is Chapter 5 of the *International Classification of Diseases*, produced by the World Health Organization (1992). The DSM lists typical symptoms for hundreds of problems such as different kinds of depression and schizophrenia. These classifications have been useful for clinical practice and research but have been criticized for being insufficiently informed by a biomedical understanding of mental illness.

Murphy (2006) makes three trenchant criticisms of the latest version of the DSM, DSM-IV-TR. First, it is incoherent, in that it officially regards causal information and theoretically informed observation as impermissible but nevertheless relies on discriminating between disorders causally and describing their symptoms theoretically. Second, it is heterogeneous, in that it lumps together dissimilar conditions and separates similar ones. Third, it is provincial in restricting attention to conditions that have historically been the province of psychiatry, while neglecting relevant research in areas such as cognitive psychology and neuroscience. Murphy argues that to overcome this problem psychiatry needs a taxonomy based on causal discrimination rather than clinical phenomenology. Additional criticisms of the DSM approach are made by Bentall (2004) and Galatzer-Levy and Galatzer-Levy (2007). The APA is planning to produce DSM-V, which is projected to place less emphasis on symptoms and much more on neural pathology (Kupfer, First, and Reiger 2002).

A good example of the value of a causal theory for taxonomy is the recent controversy over whether Pluto is a planet (Sofer 2007). Pre-Copernican astron-

omy included as planets Mercury, Venus, Mars, Jupiter, Saturn, and the sun and moon, but the acceptance of the heliocentric theory led to the deletion of the sun and moon as planets, and the addition of Earth. Pluto was added in 1930 but deleted by the International Astronomical Union in 2006. The reclassification was based on a theoretically motivated redefinition of a planet as a body massive enough to dominate its orbital zone by flinging smaller bodies away, sweeping them up in direct collisions, or holding them in stable orbits. This definition is not based on the simple properties of a body, but on causal theories of the dynamics of gravitation.

The history of infectious diseases provides a medical example. Many infections produce similar symptoms such as fever, and before the development of specific germ theories there was no good way of distinguishing different febrile diseases. However, once specific bacterial, viral, and fungal agents of infection were identified, it became possible to distinguish diseases such as malaria and tuberculosis based on their causes and not just on their symptoms (Thagard 1999).

Similarly, it would be desirable if neuroscience could provide the basis for a causal taxonomy of mental illnesses. Emil Kraepelin is viewed in the history of psychiatry as having made a major advance in distinguishing between dementia praecox (now called schizophrenia) and manic depression (now called bipolar disorder). But schizophrenia often involves symptoms such as affective flattening that can be part of depression, and the manic episodes in bipolar disorder can have some delusional aspects such as inflated self-esteem or grandiosity. Naturally, DSM contains an intermediate condition, schizoaffective disorder, which combines criteria for schizophrenia with criteria for mania and/or depression. How might theoretical neuroscience help to sort out this conceptual mess?

Suppose the explanatory gap between neuroscience and mental symptoms is progressively filled by developments in theoretical neuroscience that describe the molecular and neural mechanisms that affect behavior and conscious experience. For example, it might turn out that a genetically acquired combination of genes for dopamine receptors leads to excessive pruning of synaptic connections in the prefrontal cortex, which leads to insufficient cortical constraints on perception, thereby producing symptoms such as hallucinations. Suppose further that symptoms of major depression such as suicidal thoughts turn out to be the result of stress-induced depletion of hippocampal neurons and connections. If these causal explanations are reliable, and if techniques are developed to measure non-invasively the degrees of cortical connectivity and hippocampal depletion, then medicine would have a strong basis for distinguishing the perceptual symptoms of schizophrenia from the affective symptoms of depression. The measurement problem is not insurmountable, thanks to the availability of techniques such as functional magnetic resonance imaging for identifying levels of brain activity in specific regions, and positron emission tomography for monitoring levels of brain chemicals such as dopamine. Like the catch-all category of "fever" in Hippocratic humor-based medicine, the concepts of schizophrenia and depression

would be superseded by more precise concepts tied to causal explanations. Then psychiatry would undergo something like the substantial conceptual change experienced by medicine in the wake of the germ theory of disease. The concept of psychosis would be differentiated much more finely than is now possible with rough symptom-based categories like schizophrenia, bipolar disorder, and Alzheimer's disease.

Reclassification of mental illnesses in accord with theoretical neuroscience would help enormously to solve the problems that Murphy identified with the DSM approach. Not only would classification based on biological theory enable principled ways of establishing similarities and differences between disorders, but the provinciality of psychiatry would be overcome by relating it to neuroscience and cognitive psychology. The result would be a systematic, theory-based conceptual change, analogous to the dramatic changes in the classification of animals brought about by Darwin's theory of evolution (Thagard n.d.-b). And such progress on the explanation problem should advance treatment of mental illness.

TREATING MENTAL ILLNESSES

Although treatment for maladies such as schizophrenia and depression is sometimes highly effective, psychiatrists and other physicians are unable to predict what drugs will be most useful for particular patients. Prescription of medications is often haphazard, with only rough heuristics or trial and error dictating drug regimens for a given patient. For example, a patient with depression may be serially prescribed drugs such as Prozac, Zoloft, and other antidepressants until one works, and there is no guarantee that any drug will be effective. Similarly, psychiatrists often try a variety of antipsychotics, such as chlorpromazine, haloperidol, and others, in a sometimes unsuccessful attempt to find one that resolves a patient's symptoms without generating intolerable side effects. Moreover, medications for treating depression, schizophrenia, and other mental illnesses often have unpleasant side effects that make patients reluctant to take them, and only trial and error enables physicians to identify medications that minimize side effects for particular sufferers. In sum, psychiatry suffers from the treatment problem that therapy for mental illnesses is often hit or miss.

To be fair, it must be acknowledged that the same problem afflicts many other areas of medicine. For example, a patient with hypertension can be treated with a variety of drugs, such as calcium channel blockers, ACE inhibitors, diuretics, and beta blockers, and the physician rarely knows in advance which of these, or which combination, will be most effective in reducing blood pressure without serious side effects. Similarly, cancer treatments employing different kinds of chemotherapy and radiation have varying effectiveness for different patients and for different stages of illness in a particular patient, and they often have serious side effects. However, there is hope that deeper understanding of the molecular

basis of cancer can lead to much more targeted therapies that have a greater chance for effectiveness and avoidance of side effects. Examples of targeted cancer drugs based on molecular mechanisms include Gleevec for one kind of leukemia and Herceptin for one kind of breast cancer.

Analogously, deeper understanding of the biochemical and neural mechanisms responsible for normal mental functioning, along with identification of the kinds of breakdowns that produce dysfunctions, should enable treatment of mental illness to become less haphazard. Ideally, it should become possible to diagnose a mental patient on the basis of more than behavioral symptoms. Analysis of genetics, biochemical balance, and brain area activity should not only enable a more precise diagnosis of the causal origins of a disease, but also provide the basis for identifying treatments that are likely to be most effective while minimizing side effects.

Theoretical neuroscience is only beginning to develop the kind of mechanistic understanding that will help to make such improvements in treatment possible, but there are promising developments on a number of fronts. As the molecular mechanisms of dopamine functioning become better understood, it should become possible to develop more focused treatments for diseases that involve excess dopamine activity (schizophrenia, ADHD) or insufficient acitivity (Parkinson's disease). For example, Frank's (2005) neurocomputational explanation of how dopaminergic medication produces both improvements and side effects may inspire pharmacological innovations that increase the former and decrease the latter. More imminently, investigations along the lines of Frank et al. (2007) on ADHD and Smith et al. (2007) on schizophrenia should have implications for improving treatments of these diseases.

The account of depression I gave above, linking neurogenesis and emotional consciousness, is highly rudimentary. But it has the potential to explain why the most effective treatments for depression are often a combination of antidepressants, which increase neurogenesis, and cognitive-behavioral therapy, which enables people to replace irrational beliefs about themselves with more coherent ones that promote positive emotions. Hence understanding and improving treatments for depression will require attention to molecular, neural, and psychological mechanisms.

Although it is hard to identify pharmaceutical or psychotherapeutic discoveries generated by theoretical neuroscience, advances have already been made in explaining why existing treatments for depression, ADHD, Parkinson's, and schizophrenia are sometimes successful. Further investigations should make possible more substantial advances in the difficult problem of how to treat mental illnesses. As I discuss in the conclusion, treatments of mental illness based on neural mechanisms can naturally be combined with psychological and social treatments.

THE OBJECTIVITY PROBLEM

Finally, we come to the objectivity problem, the most fundamental one in the philosophy of psychiatry. Writers such as Foucault (1965), Laing (1967), and Szasz (1961) have challenged the legitimacy of the whole idea of mental illness. Psychiatry assumes that mental disorders are as biologically objective as diseases like infections and cancers. In contrast, Galatzer-Levy and Galatzer-Levy (2007) report how the anti-psychiatry movement contended that "psychiatry acts primarily as an instrument of social control. In this view, psychiatric diagnoses are applied to socially undesirable behaviors in order to rationalize oppression of offensive groups, including sexual and religious minorities, political dissenters, and those who cause social 'scandals'" (p. 168). Support for the anti-psychiatry view comes from such embarrassing historical episodes as the diagnosis of "drapetomania" for the desire of slaves to flee their masters, the inclusion of homosexuality in DSM-II and its agitated removal from later versions, and the use of psychiatric prisons for political repression in the Soviet Union.

There are many embarrassments in the history of medicine, from the millennia-long use of bloodletting as a treatment to the now discredited theory that peptic ulcers are primarily caused by stress. However, modern medicine also claims unchallengeable successes, such as the microbial theory of infectious diseases and their treatment by antibiotics. The crucial question is whether contemporary psychiatry can achieve a scientific, medical basis that would nullify the claim that it is merely a tool for social control.

By helping to solve the explanation, classification, and treatment problems, theoretical neuroscience can go a long way to solving the objectivity problem. Epidemiological evidence of mental illnesses across cultures suggests that diseases such as schizophrenia and depression are not merely social constructs. Neuroscience is beginning to provide the biological basis for the view that mental illnesses are just as objectively real as infections, cancers, and other diseases. There is ample reason to reject the extreme social constructivist claim that *all* diseases, like all scientific constructs, lack objectivity (Thagard 1999). Given the evident suffering of those afflicted and the modest success of biologically based treatments, the claim that mental illnesses are merely constructed instruments of social control is already implausible. Attacks on the objectivity of mental illness will become even more ridiculous if mental illnesses can be effectively classified on the basis of the causal mechanisms that produce them, if these causal mechanisms provide detailed pathways from genetic and environmental conditions to mental symptoms, and if biological understanding paves the way for improved therapies. I expect that advances in theoretical and experimental neuroscience will satisfy all these requirements.

Philosophical Naturalism and Neuroscience

My claims about the philosophical relevance of theoretical neuroscience have been largely speculative, dependent on extrapolations about how progress in understanding the biological underpinnings of mental illness may continue. Many philosophers and even some scientists will naturally remain skeptical about how scientific developments could possibly have such large relevance to philosophical issues. Science, after all, cannot provide the a priori, necessary truths that many have thought to be required for philosophical understanding. How can empirical and theoretical methods in neuroscience and other fields possibly shed light on philosophical problems? I will now try to sketch generally how science is directly relevant to such problems.

Here are what seem to me to be the four most important philosophical questions in metaphysics, epistemology, and ethics, the central areas of philosophy: what is reality, how do we know reality, what is the difference between right and wrong, and why is life worth living? The greatest philosophers, such as Plato, Aristotle, Hume, and Kant, have offered explicit or implicit answers to some or all of these questions. For Plato, Kant, and many other thinkers, the answers need to be sought in supernatural directions involving Platonic Forms, God, or truths that can be established for all possible worlds, not just the one that science studies.

In contrast, philosophical naturalism attempts to answer these questions based on scientific investigations of the world that people actually inhabit. The first question, about the nature of reality, cannot be answered a priori but depends on the results of empirical/theoretical investigations in all the sciences, from physics, chemistry, and biology to psychology and even sociology. These sciences are each restricted to particular aspects of reality—for example, physics to laws of motion that apply to various kinds of objects, or sociology to principles that describe the operation of people in groups and organizations. Philosophical investigations in metaphysics are much more general than research in any particular science, looking at questions about the overall nature of what exists, including the nature of space and time.

Why should we prefer a naturalistic approach to reality to supernatural ones? There are two main reasons: the failure of supernaturalism to establish any kind of consensus about the nature of reality, and the success of science in producing reliable interactions with reality. Supernatural metaphysics based on a priori reasoning or religious faith has failed to provide any justification for preferring one answer about what exists to competing answers. Is there one god or many? Are the gods benevolent or malevolent? Was the world created recently or eons ago? Do souls precede conception, as Hindus and ancient Greeks believed, or are they created at conception, as many Christians believe? As disagreements among theologians and a prioristic philosophers show, no one has come up with a way to determine which of alternative supernaturalistic theories gives better answers to questions about the nature of reality.

Science also has its disagreements but has a common set of methods for dealing with them: comparative evaluation of competing theories based on which ones best explain the results of observation and experiment. These methods have ancient precedents, but their systematic application only began in the 16th and 17th centuries. The main reason we have for thinking that scientific methods have often provided a grip on reality is the many technological applications of experimentally established scientific theories. To take just two examples, physical theories about atoms and electrons have enabled technologies such as computers, and biological theories about diseases and germs have produced medical advances such as antibiotics. (For a more general argument about science and truth, see Thagard 2007.) Science has been much more successful in grasping reality than supernaturalism, so philosophy should ally itself with science in its metaphysical investigations. In particular, if philosophy of psychiatry wants to understand the nature of minds and the diseases that afflict them, it stands to learn more from scientific investigations in psychology and neuroscience than from inquiries that are supernatural or purely conceptual. Cognitive neuroscience has already made substantial progress in understanding such psychological phenomena as perception, attention, memory, problem solving, and language (Smith and Kosslyn 2007).

Because science is the best way to purse the nature of reality, it is also highly relevant to answering the central epistemological question of how we know reality. Lacking supernatural souls with a special access to a priori truths, humans must rely on the minds we have, which fortunately can be investigated by the cognitive sciences such as psychology and neuroscience. Much has been learned about the workings of human minds, ranging from basic perceptual processes shared with other minds to high-level scientific reasoning. (For an accessible introduction to cognitive science, see Thagard 2005.) Psychology and neuroscience are learning more and more about how minds gain knowledge about the world, so it is pointless for epistemology to proceed in the a priori and introspective modes that have been preferred by many philosophers. This is not to say that epistemology can be replaced by psychology and neuroscience, whose investigations have tended to be much more narrowly focused on particular perceptual and cognitive processes. Epistemology remains important for raising general questions about the relation between mental states and the world, as well as for normative questions about how minds can work better to learn more about the world. But these general and normative questions are dependent on a basic understanding of how minds work, which requires psychology and neuroscience. Hence epistemology, like metaphysics, should be naturalistic. Even more controversially, a case can be made that ethics can best be approached naturalistically, addressing questions about the nature of right and wrong and the meaning of life.

Descombes (2001) offers a critique of cognitivism, which he takes to be the view that mental phenomena such as intentions are purely internal to a person

and so can be understood as states of the brain. In contrast, he draws on the later Wittgenstein to argue that intentional states should be understood anthropologically in terms of a person's social history and education. It might seem that if Descombes is right then my attempts to use neuroscience to explain mental illness are seriously defective: *mental* problems cannot just be brain problems, because mental states such as those found in people suffering from schizophrenia and depression are inherently social. Mental illnesses afflict persons, not brains.

Descombes's arguments might be effective against narrow versions of the cognitive approach such as that advocated by his major target, Jerry Fodor (1987). But current research in cognitive neuroscience, including the computational modeling that occurs in theoretical neuroscience, is thoroughly compatible with the view that mental states need to be understood partly in terms of their relation to the external world, including social history. First, brain representations are understood as patterns of neural firing that result from interactions that tune neurons to collectively encode real-world physical magnitudes (Eliasmith and Anderson 2003). Such interactions depend on the nature of our perceptual systems, so that brain states are a function of the structure of our bodies, as well as of the world and previous brain states. Saying that mental states are brain states does not make them purely internal as Descombes charges, because brain states are formed in part through a history of interactions with the world. Second, neuroscience is increasingly aware that many of these interactions are social, involving communication with other people. There is a rapidly growing field, social cognitive neuroscience, that investigates the neural mechanisms of social cognition and social interaction in humans and animals (Easton and Emery 2005). Such investigations imply that brain states depend in part on an organism's history of social interactions. Hence there is no basis to the claim that neuroscience cannot explain mental illness because its account of mental states neglects their external, social, and historical nature. Of course, neuroscience does neglect what supernaturalists take to be the spiritual nature of mental states, but such neglect is justified by lack of evidence that minds are souls.

CONCLUSION

Proponents of a more humanistic approach to psychiatry will be inclined to reject my enthusiasm for theoretical neuroscience as a manifestation of a misguided scientism and reductionism. But mine is not an unrestricted scientism, as I readily acknowledge that there are normative issues about right and wrong that science by itself does not resolve. I have been concerned in this article with the epistemology of mental illness, but a complete philosophical discussion would also require a discussion of psychiatry from the perspective of an ethics of care (Tauber 2005).

And I am certainly not defending a kind of ruthless reductionism, in which the only legitimate scientific explanations are those that occur at the most basic

physical or biological level. Rather, I have stressed elsewhere that understanding emotional thinking requires attention to interconnected mechanisms at four different levels, the social, cognitive, neural, and molecular (Thagard 2006). Similarly, although this article has focused on neural and biochemical explanations of mental illness, it is compatible with the view that psychiatry should also attend to psychological and social causes (Bentall 2004). It follows immediately that treatment of mental illness should employ psychological and social interventions such as psychotherapy and stress reduction, in addition to pharmaceutical ones inspired by advances in neurobiology.

I have argued that attention to current and potential advances in the burgeoning field of theoretical neuroscience can help to solve four of the major problems in the philosophy of psychiatry: explanation, classification, treatment, and objectivity. In accord with the approach defended at the beginning of this essay, my argument has been naturalistic, drawing philosophical conclusions from connections with the best scientific work in many fields, including neuroscience, medicine, psychology, and molecular biology. I have proposed no conclusions as a priori truths, and sought conceptual clarification, not by the abstract analysis of the use of words, but by exploring the implications of scientific theories well supported by experimental evidence. In contrast to the quietist Wittgensteinian view of philosophy, I have aggressively proposed how psychiatry ought to progress through close association with theoretical neuroscience and allied fields such as biochemistry. Like science, philosophy should aim not just to analyze concepts, but to change them.

REFERENCES

American Psychiatric Association (APA). 2000. *Diagnostic and statistical manual of mental disorders*, 4th textual revision ed. (DSM-IV-TR). Washington, DC: APA.

Anderson, J. R., et al. 2004. An integrated theory of the mind. *Psychol Rev* 111:1030–60.

Bechtel, W., and A. A. Abrahamsen 2005. Explanation: A mechanistic alternative. *Stud Hist Philos Biol Biomed Sci* 36:421–41.

Bechtel, W., and R. C. Richardson. 1993. *Discovering complexity*. Princeton: Princeton Univ. Press.

Becker, S., and J. M. Wojtowicz. 2007. A model of hippocampal neurogenesis in memory and mood disorders. *Trends Cogn Sci* 11(2):70–76.

Behrendt, R., and C. Young. 2005. Hallucinations in schizophrenia, sensory impairment, and brain disease: A unifying model. *Behav Brain Sci* 27:771–830.

Bentall, R. 2004. *Madness explained: Psychosis and human nature*. London: Penguin.

Cohen, J. D., T. S. Braver, and J. W. Brown. 2002. Computational perspectives on dopamine function in prefrontal cortex. *Curr Op Neurobiol* 12(2):223–29.

Craver, C. F. 2007. *Explaining the brain*. Oxford: Oxford Univ. Press.

Dayan, P., and L. F. Abbott. 2001. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge: MIT Press.

Descombes, V. 2001. *The mind's provisions: A critique of cognitivism*. Princeton: Princeton Univ. Press.

Easton, A., and N. Emery, eds. 2005. *The cognitive neuroscience of social behaviour*. London: Psychology Press.

Eliasmith, C., and C. H. Anderson. 2003. *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge: MIT Press.

Fodor, J. 1987. *Psychosemantics*. Cambridge: MIT Press.

Foucault, M. 1965. *Madness and civilization: A history of insanity in the age of reason*. New York: Pantheon.

Frank, M. J. 2005. Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci* 17(1):51–72.

Frank, M. J., and E. D. Claus. 2006. Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113(2): 300–326.

Frank, M. J., et al. 2007. Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology* 32(7): 1583–99.

Galatzer-Levy, I. R., and R. M. Galatzer-Levy. 2007. The revolution in psychiatric diagnosis: Problems at the foundations. *Perspect Biol Med* 50(2):161–80.

Graham, G., and G. L. Stephens. 2007. Psychopathology: Minding mental illness. In *Philosophy of psychology and cognitive science*, ed. P. Thagard, 339–67. Amsterdam: Elsevier.

Grossberg, S. 2000a. How hallucinations may arise from brain mechanisms of learning, attention, and volition. *J Int Neuropsychol Soc* 6(5):583–92.

Grossberg, S. 2000b. The imbalanced brain: From normal behavior to schizophrenia. *Biol Psychiatry* 48(2):81–98.

Hoffman, R. E., and T. H. McGlashan. 2001. Neural network models of schizophrenia. *Neuroscientist* 7(5):441–54.

Jacobs, B. L., H., Praag, and F. H. Gage. 2000. Adult brain neurogenesis and psychiatry: A novel theory of depression. *Mol Psychiatry* 5(3):262–69.

Kapur, S. 2003. Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry* 160(1):13–23.

Kramer, P. D. 2005. *Against depression*. New York: Penguin.

Kupfer, D. J., M. B. First, and D. A. Regier, eds. 2002. *A research agenda for DSM-V*. Washington, DC: APA.

Laing, R. D. 1967. *The politics of experience*. New York: Pantheon.

Machamer, P., L. Darden, and C. F. Craver, 2000. Thinking about mechanisms. *Philos Sci* 67:1–25.

Medin, D. L. 1989. Concepts and conceptual structure. *Am Psychologist* 44:1469–81.

Murphy, D. 2006. *Psychiatry in the scientific image*. Cambridge: MIT Press.

O'Reilly, R. C., and Y. Munakata. 2000. *Computational explorations in cognitive neuroscience*. Cambridge: MIT Press.

Putnam, H. 1983. There is a at least one *a priori* truth. *Realism and reason: Philosophical papers, volume 3*, ed. H. Putnam, 98–114. Cambridge: Cambridge Univ. Press.

Quine, W. V. O. 1963. *From a logical point of view,* 2nd ed. New York: Harper Torchbooks.

Salmon, W. 1984. *Scientific explanation and the causal structure of the world*. Princeton: Princeton Univ. Press.

Smith, A. J., et al. 2007. Linking animal models of psychosis to computational models of dopamine function. *Neuropsychopharmacology* 32(1):54–66.

Smith, E. E., and S. M. Kosslyn. 2007. *Cognitive psychology: Mind and brain*. Upper Saddle River, NJ: Pearson Prentice Hall.

Sofer, S. 2007. What is a planet? *Sci Am* 296(Jan.):34–41.

Szasz, T. S. 1961. *The myth of mental illness: Foundations of a theory of personal conduct*. New York: Harper and Row.

Tauber, A. I. 2005. *Patient autonomy and the ethics of responsibility*. Cambridge: MIT Press.

Thagard, P. 1999. *How scientists explain disease*. Princeton: Princeton Univ. Press.

Thagard, P. 2005. *Mind: Introduction to cognitive science*, 2nd ed. Cambridge: MIT Press.

Thagard, P. 2006. What is a medical theory? In *Multidisciplinary approaches to theory in medicine*, ed. R. Paton and L. A. McNamara, 47–62. Amsterdam: Elsevier.

Thagard, P. 2007. Coherence, truth, and the development of scientific knowledge. *Philos Sci* 74:28–47.

Thagard, P. n.d.-a. Cognitive architectures. In *The Cambridge handbook of cognitive science*, ed. K. Frankish and W. Ramsay. Cambridge: Cambridge Univ. Press, forthcoming.

Thagard, P. n.d.-b. Conceptual change in the history of science: Life, mind, and disease. In *Handbook of conceptual change*, ed. G. Hatano and S. Vosniadou. Mahwah, NJ: Erlbaum, forthcoming.

Thagard, P., and Aubie, B. n.d. Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Conscious Cogn*, forthcoming.

Wittgenstein, L. 1968. *Philosophical investigations*, 2nd ed., trans. G. E. M. Anscombe. Oxford: Blackwell.

Wittgenstein, L. 1971. *Tractatus logico-philosophicus*, 2nd ed., trans. D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul.

World Health Organization (WHO). 1992. *ICD-10: The international statistical classification of diseases and related health problems*, 10th ed. Geneva: World Health Association.