

# Mental Disorder, Illness and Biological Disfunction

DAVID PAPINEAU

## 1 Introduction

This paper will be about the relationship between mental disorder and physical disorder. I shall also be concerned with the connection between these notions and the notion of 'illness'.

I shall begin with the 'anti-psychiatry' view that the lack of a physical basis excludes many familiar mental disorders from the category of 'illness'. My response to this argument will be that anti-psychiatrists are probably right to hold that most mental disorders do not involve any physical disorder, but that they are wrong to conclude from this that these mental disorders are not illnesses.

My arguments in this paper will be philosophical rather than empirical. My conclusions will be informed by recent philosophical analyses of notions like 'mental state' and 'biological purpose'. How they might apply to particular mental disorders, such as depression, say, or schizophrenia, is not for me to say. This will depend on the detailed empirical nature of such disorders, and this is something about which I am relatively ignorant, and about which even the experts disagree. My aim here is to clear away some of the more abstract issues, so that we will know better what to say if we ever are in full command of the empirical facts.

## 2 The 'anti-psychiatry' argument

A simple version of the anti-psychiatry argument might go as follows. 'What reason do we have for supposing that mental disorders are generally based on physical disorders? It is true that a few mental abnormalities are uncontroversially symptoms of underlying physical disorders: incapacities due to brain lesions, for instance, or dementia in tertiary syphilis. But there is no obvious evidence that other mental disorders—neuroses, affective disorders, schizophrenia—have a similar physical basis. However, if these latter disorders are specific to the *mental* level, and not due to any underlying physical injury or ailment, then surely they should not count as *illness*. They are social abnormalities, not med-

## David Papineau

ical ones.' (For a less simple version of this argument, and related theses, see Szasz, 1961.)

Note that this argument involves two steps. (1) Many familiar mental disorders are independent of any physical disorders. (2) They are therefore not illnesses. I am happy to concede (1), but will deny that (2) follows.

### 3 The autonomy of mental disorders

At first sight it might seem that by conceding (1), I am conceding the anti-psychiatrists too much. None of us, I take it, wants to be a Cartesian dualist any more, with a separate metaphysics of mental and physical substances. But if we deny Cartesian dualism, then doesn't it follow that all mental disorders must in some sense be physical disorders too? Surely the idea of someone with a healthy brain, yet a disordered mind, only makes sense from a Cartesian point of view.

But I think this is the wrong response to the anti-psychiatry argument. Recent work in the philosophy of mind shows that, even for non-Cartesians, there is a perfectly good sense in which a disorder can be purely mental and not physical at all. What is more, it is arguable that many familiar mental disorders fall into this category—though exactly which these are will of course depend on empirical facts.

This is why I am happy to concede to the anti-psychiatrists the possibility of a purely mental, entirely non-physical disorder. However, the way the modern philosophy of mind makes room for this possibility will lend no support to the anti-psychiatrists' further idea that such purely mental disorders are not illnesses. This is because illness is centrally a matter of *biological disfunction*; and, as we shall see, there is no reason why a purely *mental* disorder should not also be a *biological* disfunction. So my eventual view will be that many mental disorders are purely mental and not physical, but that nevertheless they are illnesses in a full biological sense. (Let me say at this point that much of this article, and in particular the idea that mental disorders are illnesses because they involve biological disfunction, follows the argument of Karen Neander's unpublished PhD thesis, *Abnormal Psychobiology*, 1984.)

### 4 Physicalism without reductionism

Let me now sketch the way that I, and most other contemporary English-speaking philosophers, think about the mind-brain rela-

## Mental Disorder, Illness and Biological Disfunction

tion. We are *physicalists*, but we are not *reductionists*. We are physicalists because we deny any substantial realm of being separate from the physical realm. Yet we are not reductionists because we hold that mental categories in general fail to correspond to physical categories. Mental categorisation is concerned with a level of structure which abstracts from the physical details of how the structure is realised, and indeed is compatible with the structure being realised differently in different cases. (See, for example, Fodor, 1974.)

The software–hardware distinction from computer science provides a helpful analogy. (Minds are not digital computers; but on this specific point the analogy is useful.) To say that a computer is running Microsoft Word is not to say anything about the internal physics of the computer, as is shown most forcefully by the fact that computers with different internal physics—such as PCs and Macintoshes—can both run this same program. They are both running this program because they are *structurally* similar even though physically different: they are both disposed to respond to given inputs with a sequence of states which bear certain causal relations to each other; this sequence of states will be physically different, given that the machines have different hardware, but they will be structurally similar, in that they embody the same system of causal relations.

Similarly, says contemporary philosophical orthodoxy, with mind and brain: to be in pain, say, does not require any specific internal physics or physiology, but only that you instantiate a certain architecture of causal relations. And this is shown most clearly by the fact that when a human and an octopus are both in pain, there is nothing physically in common between them. The kinds of molecules and cells involved in each case are quite different. What the human and the octopus have in common is rather the same causal structure, in that they each have states (physically different in the two cases) which are typically caused by bodily damage, and typically give rise to avoidance behaviour.

### 5 What about species-specific reductions?

These last remarks will give some indication of the direction I am heading in. If mental categorisation in general abstracts from physical embodiment, then perhaps mental disorders in particular float free of physical disorders. But before pursuing this idea in more detail, I would first like to consider a general objection to my overall line of attack.

## David Papineau

This objection queries whether analogies with computers, or examples involving octopuses, are really relevant to the issue at hand. After all, we are concerned with *human* psychology, not octopus or computer psychology. And the points made in the last section might still seem to leave room for a reduction of mind to brain *within* the human species (and within other species). Even if pain involves different physiological mechanisms in humans and octopuses, it does not follow, and it seems very unlikely to be true, that it involves different physiological mechanisms in different human beings. But this now suggests that, among humans at least, there will be a close match between mental and physical categories after all, and in particular between mental disorders and physical disorders.

### 6 Design and variable realisation

Despite this objection, I think that there is no close match between mental and physical categories, even within the human species. True, the pain mechanism probably has the same physical basis in all human beings. But I think that pain is an exception in this respect, and that most other mental kinds, including kinds of mental disorder, fail to match up with physical classifications in the same way.

To see why this should be so, it is helpful to ask a more general question. Why do we often find the same structure being realised in different materials? This question is not asked as often as it should be, by philosophers working in this area, but it is a good question, for there is surely something to explain about the existence of physically different systems which nevertheless link up physical inputs (bodily damage, for example) with physical outputs (avoidance behaviour) in the same way. In general, when we find regular physical patterns in nature, we expect there to be some uniform explanation, in terms of some common physical mechanism responsible for the behaviour. But that is just what current orthodoxy in the philosophy of mind denies. We observe the same input-output patterns in different species, but mediated by different physical mechanisms, which means these patterns lack any uniform physical explanation. So the puzzle is this: if the insides of octopuses and humans are so different, how comes both species move away from any sharp object that penetrates their skins?

I address this issue in some detail in chapter 2 of my recent book *Philosophical Naturalism* (1993). My response to the puzzle is that

## Mental Disorder, Illness and Biological Disfunction

biological organisms are in a sense *designed* systems—though of course their designer is blind natural selection, rather than any human or divine agent. Let me illustrate the case by considering the simpler case of objects which *are* designed by humans. Take thermostats, say. Thermostats are devices which switch off heating systems when the temperature gets too high, and turn them on again when it gets too low. Now, thermostats are prime examples of items which vary in their physical constitution: there are many different physical designs of thermostat, involving different materials arranged in different ways. So you can imagine somebody raising the puzzle: how come these physically quite different items all respond in the same way when the temperature gets too high or low? But in this case of course there is an obvious answer: they all do this, despite their different physical constitutions, because that is what they were *designed* to do. Their human designers choose their various physical constitutions precisely because those constitutions will yield the right responses to the right inputs. The designers do not mind about the internal make-up of the thermostats, as long as they link causes and effects in the right way.

A similar point applies to pain in octopuses and humans. The reason that there is the same link between inputs and outputs in both octopuses and humans, despite the fact that the physical make-up of the intervening mechanism is quite different, is that natural selection has designed the mechanism to create this link in each case. Natural selection chose these different mechanisms precisely because they both yield the right input-output links. Natural selection does not mind, so to speak, about the internal physics of the pain mechanisms, so long as they link causes and effects in the right way.

In support of the point made in this section, note that it implies that non-physical categories *will* reduce to physical ones in those areas of science where natural or other designers do not play a rôle. And this seems plausible. Even it is often difficult to carry out the reduction in practice, everybody accepts that in principle there must be uniform physical explanations for the general truths uncovered by chemists, geologists, meteorologists, and so on.

### 7 Learning as a natural designer

Let me now return to my main line of argument. I have just argued that the existence of some kind of designer can explain why given structures are often realised in different materials. The issue with which we left the main line of argument, however, was

whether we will find different physical realisations of mental structures *within* the category of humans. It might not be immediately clear how the point about designers bears on this specific issue. For, when natural selection is the designer, the normal result is still that the members within a species end up physically similar. Apart from a few special cases of polymorphism, natural selection tends to give all the members of a species the same genetic program, which produces the same proteins, and eventually the same adult physical characteristics. So, as with the pain example, we would expect to find the same physical basis for any given mental state in different people. The designs produced by natural selection might account for physically different mental processes across species, but they give us no reason to expect such physical differences within a given species.

The trouble with this line of thought is that it assumes that biological design stops with the intergenerational natural selection of genes. However, there is a quite analogous process which goes on during the course of individual development, and which does account for physically variable realisations of mental processes within a given species. This is *learning*. Learning is no less a designer than genetic natural selection, and because of this we should expect that learned mental states will vary in their physical realisation within a species, as much as innate ones vary in their physical realisation across species.

To take a simple example, consider the ability, which I take it we all share, to recognise the letter 'b' by sight. I take it that in each of us there is some neuronal mechanism which serves to produce signs of recognition (the sound *bee*, say) in response to the appropriate visual input. And I take it also that this neuronal mechanism was selected (reinforced, developed) *because* it produced that output given that input. But note that nothing in this idea requires that the *same* neuronal mechanism plays this role in different people. All that the learning mechanism cares about is we each end up with *some* mechanism which does the job, not that it be the same in each case. And there is ample evidence, both from neurophysiology and from computer simulations of learning by neural nets, to suggest that different arrangements of neurones will end up doing this job in different people. (It will depend on such chancy things as how your neurones were initially configured before learning started, and which other letters you might have learned before 'b'.)

This is of course an extremely simple example. But the point is general. Most mental states and processes are acquired in the course of individual learning, rather than fixed by species-wide

## Mental Disorder, Illness and Biological Disfunction

genes. So we can expect them to have different physically realisations in different individuals, and in particular in different humans.

### 8 Disorder at a non-physical level

I am now finally in a position to explain how a disorder can be purely mental, and not physical at all, even for those of us who reject Cartesian dualism. So far I have been talking, not specifically about disorders, but simply about mental and physical states or processes in general. But let me now focus on the notion of disorder. I take *disorder* to mean that something is not working as it is supposed to, not doing the job for which it was designed. The points made so far imply that this might be true of some mental faculty, not because anything is failing to do its job at the physical level, but solely because some structural task is not being carried out.

Let me return briefly to the computer analogy, to make the point graphic. Suppose that you and I are both using MS Word 5.0 as our word processing program, but that you are working on a PC while I am working on a Macintosh. Now suppose that there is some bug in the program. For example, suppose that whenever either of us tries to double-space a highlighted section, that section gets deleted. (This is just for the sake of the argument: I've never found this or any other bug in MS Word 5.0.) This obviously wouldn't show that there was anything *physically* wrong with our machines.

It would be silly to get the hardware engineer to come and solve the problem. The logic circuits are all working as they are supposed to. Rather, the fault lies entirely at the software level.

You might not be convinced by this example. Once someone is running the faulty version of MS Word 5.0, you might object, isn't there a sense in which the physical computer, made of circuits and switches, *isn't* doing what it is supposed to? After all, as currently programmed it is supposed to allow the user to process words without deleting them involuntarily—and the computer is certainly not doing this. So isn't there something physically wrong with the machine after all? Aren't its circuits and switches failing to do what they are supposed to?

But note that, while there is indeed a sense in which this computer is not working, the respect in which it is not working cannot be *physically* specified. There is no strictly physical description of what the computer is failing to do. For what it is failing to do is

## David Papineau

*avoid involuntary deletions of words*, and this is a structurally specified fault, not a physically specified one. (Note that *this* fault is *common* to both the PC and the Macintosh, even though their physical system of switches and circuits is different. This shows why it is a specifically structural rather than physical fault.)

### 9 Purely mental disorders

This illustrates my notion of a purely mental disorder: it is the failure to perform some function, where that function can only be specified in structural terms, and not physically. I think that this notion has direct application in psychiatry. For there might be humans whose brains are doing everything they are supposed to do at the physical level—they have all the right molecules, enough neurotransmitters, and so on—but are still failing to do something else they are supposed to do, in the sense that some aspect of their structural design has gone awry.

Now, as I indicated earlier, it is clearly an empirical matter, hinging on factual questions which are often controversial, which actual mental disorders fit this scheme. Even so, I would like to suggest that many familiar mental disorders may fit this model. Consider, by way of illustration, obsessional-compulsive behaviour. We can think of this as a disorder of our decision-making system. Normally this decision-making system weighs up different ends, considers different means, and chooses an appropriate course of action. But it may be that the procedures by which this is done lead in certain unusual cases to an exclusive concentration on certain concerns and associated behaviours, in a way that excludes all other projects, and causes distress to the unfortunate people captured by the compulsion. But there is no reason at all to think that this must be caused by a fault at the level of molecules and neurotransmitters. Perhaps there is just some bug in the program which governs our decision-making procedures, some kind of vicious loop which is triggered in certain special cases, and locks the unhappy subject into the treadmill of obsessional behaviour.

### 10 Mental disorders are still illnesses

This was the first point I wanted to make. Many mental disorders may well be purely mental, and involve no physical disorder at all. To this extent the anti-psychiatrists are right. But my second point, to which I now turn, is that these disorders may still be



## Mental Disorder, Illness and Biological Disfunction

medical *illnesses*. To the extent that they deny this, the anti-psychiatrists are wrong.

This second point is in fact a corollary of my argument so far. To show how it follows, I need to make two further assumptions explicit. Both involve the idea of *biological disfunction*.

- (i) Medical illnesses are a matter of biological disfunction—of things not doing what they are biologically supposed to do.
- (ii) Biological disfunction is nothing more than items not producing those effects they were designed to produce, that is, not producing the effects in virtue of which they were naturally selected.

There are perhaps some philosophers today who would disagree with these assumptions. But the majority would, I think, regard them as uncontentious. In any case, I shall take them as given here.

Given these assumptions, it follows immediately from my argument so far that even if a mental disorder is independent of any physical disorder, it is still uncontroversially an illness. For the possibility of mental-but-not-physical disorders arises, I have argued, because mental processes are *selected*, in the course of genetic evolution and individual learning, to play certain structural roles. The corollary is that when these processes fail to play these roles, that is, when they fail to produce those effects for which they were selected, then this is a matter of biological disfunction (cf. (ii) above); and so, given that biological disfunction implies illness (cf. (i) above), it follows that mental-but-not-physical disorders are *still* illnesses. (These remarks do cut one obvious corner. Even if biological disfunction is necessary for illness, it is clearly not sufficient. A disfunction only counts as an illness if it is also in some sense *incapacitating*. Exactly what kinds of incapacity qualify a disfunction as an illness is an interesting issue; but it is orthogonal to my concerns in this article.)

### 11 Do the mentally ill need doctors?

Those who argue that mental disorders are not illnesses do not wish only to make an intellectual point. Rather they are concerned with practical implications, and in particular with the question of whether mental patients should be dealt with by traditional doctors using standard pharmacological techniques. It may seem that on this latter point my argument supports the anti-psychiatry position, even if I disagree on the more specific issue of whether

## David Papineau

mental disorders are illnesses. For if mental disorders are often independent of physical disorders, as I have argued, then doesn't it follow that physical methods of treatment, like drug therapy, are inappropriate?

The question of what kinds of therapy are best for which ailments is a huge topic, which it would be absurd for me to try to cover here. But a couple of general points about the implications my argument will be in order.

Note first that if a disorder is purely mental, its causes may still be entirely physical, as opposed to mental or social. By way of illustration, consider how some dangerous drug might precipitate a mental illness. Suppose it raises the level of some neurophysiological activity, and that this in turn leads to some specifically mental development, which persists even after the drug has worn off. Think of yourself as *learning* something dysfunctional when you are drugged, and retaining the lesson afterwards. After the drug has gone, there need be nothing physically wrong with your brain, yet the purely mental disorder may persist.

Now consider the same process in reverse. Just as a purely mental disorder may be caused by a drug, so also may one be *cured* by a drug. It may be that the effect of a given drug on those with some mental disorder is first to produce some temporary physical effect, and thereby to produce some permanent mental effect. The drug could lead to the patients *unlearning* the mental disorder, in the sense that it produces a lasting alteration of the dysfunctional mental processes involved in their illness.

In pointing to this possibility, I do not want to suggest that orthodox pharmacological treatments are necessarily the best therapy for purely mental disorders. It is quite consistent with what I have just said that some mental disorders can only be remedied by more subtle therapies, or indeed only by changes in the personal or social circumstances which originally gave rise to the disorder. My point is a more limited one, namely that it would be a mistake to infer, just because a particular mental disorder is purely mental, that it cannot be cured by pharmacological means.

Copyright © The Royal Institute of Philosophy and the  
contributors 1994