ON MENTAL ILLNESS AND BROKEN BRAINS

This is the author's original version of the paper. The version of record is forthcoming in *Think – Philosophy for everyone*

Anneli Jefferson

Abstract: We often hear that certain mental disorders are disorders of the brain, but it is not clear what this claim amounts to. Does it mean that they are like classic brain diseases such as brain cancer? I argue that this is not the case for most mental disorders. Neither does the claim that all mental disorders are brain disorders follow from a materialist world-view. The only plausible way of understanding mental disorders as brain disorders is a fairly modest one, where we label brain differences we find in mental illness as pathological based on their link to mental dysfunction. How many mental disorders will turn out to be brain disorders on this understanding is an empirical question.

We often hear that mental illness, for example bipolar disorder (also known as manic-depressive disorder), stems from a chemical imbalance in the brain.[1] Mental illness is, we are told, a disease of the brain. There are many reasons for thinking that mental illness is brain disease: findings from psychiatry and neuroscience may show differences in the brains of people suffering from, e.g. bipolar disorder, compared to brains of people in the non-clinical population. We may also hope that the brain disorder hypothesis will provide an answer to the baffling aspects of mental illness: in schizophrenia, we may hear voices or think that someone has inserted thoughts into our heads. When suffering from depression, we may be convinced of our worthlessness, even though we are surrounded by love, care and even admiration. These conditions are puzzling because, while they are often similar to states and processes we find in healthy emotion and cognition, they are exaggerated to an extent that is

hard to comprehend. We have all experienced sadness and low mood, but this is normally responsive to changes in our environment; and we can snap out of it in a way that an individual cannot snap out of depression. In cases of mental illness, certain tendencies are amplified to an extent that can be hard to understand from the perspective of the healthy person. One candidate explanation for this is that we have a problem in the brain, some kind of physiological problem that leads to psychological dysfunction.

But what exactly does it mean to say that mental disorders are brain diseases? In this short paper, I will look at three ways of understanding this claim, each of which operates with a different understanding of the term 'brain disorder'. On the strongest reading, the claim that most, or even many, mental disorders are brain diseases is straightforwardly false. On the second, it is a very weak claim which defines any brain state underlying dysfunctional mental states as pathological based on a materialist world view. The resulting position is, however, scientifically useless. The third, moderate proposal lies between these two extremes, we can call it the Goldilocks position. On this proposal it's an empirical question which specific mental disorders are brain disorders, to be resolved by the sciences and further findings about the brain. But this third reading of the claim yields fruitful ways of understanding the mind-brain relationship and the role of brain differences in mental disorders. I will try to persuade you that we should opt for the Goldilocks position.

The strong reading

The strong reading says that mental disorders are brain disorders if they are like typical brain diseases such as brain-tumours, Parkinson, or Alzheimer's dementia. In these conditions, we can find a defect in the brain that causes the mental symptoms, and we could in principle

(though not always in practice) treat the problem by treating that defect. Some forms of psychological dysfunction have been successfully explained in this way in the past. For example, neurosyphilis, which can lead to psychotic symptoms such as hallucinations in the later stages, was found to be caused by a bacterial infection, treatable by penicillin.

The strong reading holds that in order for a mental health condition to be a brain disorder, the brain and only the brain needs to be the problem. This is why I call it the strong reading: it makes strong demands on what needs to be happening at the level of the brain for a condition to count as a brain disorder. We need to be able to find an underlying physical cause in the brain which can be specified as pathological just by looking at the brain, and which is in principle amenable to physical treatment. Given past successes in finding the physical basis of diseases like syphilis, some may hope that conditions like schizophrenia or depression will in future be successfully explained as a physical disease, too.

One attraction of this view is that it will lend weight to the idea that there is something objectively wrong with the person, or so people hope. A medical label, especially a medical label that certifies that we are suffering from a physical problem, validates our condition as a real disorder. The depressed person can point to their diagnosis to prove that they are not merely acting up, but that there genuinely is something wrong with them. If a condition is a brain disease, it is a medical condition rather than a problem of living.

Indeed, critical psychiatrist Thomas Szasz (1960) believed that the only way to substantiate the claim that mental disorders are genuine illnesses lay in showing that they were brain diseases. He argued that our understanding of illness is fundamentally somatic, i.e. based on the idea of illness of the body. From this, he concluded that mental disorders are only

correctly labelled as illnesses if there is indeed something independently wrong with the brain. Otherwise, they are more accurately understood as what he called ‘problems in living’. If Szasz is right, the very idea that mental illness is an *illness* depends on the idea that there is independent brain pathology causing mental distress.

Szasz himself drew a skeptical conclusion from his definition of brain disorder. He correctly noted that most mental disorders are not brain disorders in the strong sense he defined. Why is this? Let’s return to the past successes which have served as a model for the claim that mental disorders are brain disorders. In the case of neurosyphilis, we have a bacterial cause which leads to psychological symptoms. If we do not treat the condition, it will get worse and eventually lead to death. This is also the case for other paradigmatic brain diseases which are associated with mental symptoms, e.g. brain tumours, Alzheimer’s dementia. As a psychologist once said to me, ‘Of course mental illness is not brain disease, all these brain diseases are neurodegenerative conditions which get progressively worse.’

Mental disorders need not be like that: individuals can and do recover from conditions like depression, addiction or anorexia. Furthermore, the social, the psychological and the physical are intertwined in messy ways in mental health conditions. Depression can be caused by trauma, but it can also be caused by lack of light. Addiction is associated with changes of the brain, but treatment requires the sufferer to take responsibility for their using habits and requires the desire to change, even if that desire on its own is frequently insufficient and requires help from health professionals.

If we endorse Szasz' argument, we would have to conclude both that mental disorders are not brain disorders and that they are not genuine disorders in the first place. Let me restate the argument to clarify:

P1: Brain disorders are those conditions where there is a brain defect that can be specified independently of psychology.

P2: In most mental disorders, there is no independently specifiable brain defect.

C1 (from P1 and P2): Most mental disorders are not brain disorders.

P3: Only physical disorders are disorders, so mental disorders are only disorders if they are brain disorders.

Conclusion (from C1 and P3): Most mental disorders are not disorders.

Many philosophers and clinicians object to Szasz' third premise, that mental disorders have to be brain disorders in order to count as disorders at all. But the focus of this paper is the question of whether there are other ways of understanding the nature of brain disorders than the one proposed by Szasz. My target claim is P1, which proposes a specific understanding of what it takes to be brain disorder. Accepting this would be premature - there are other proposals on the table.

The weak reading

We have seen that on the strong reading just discussed, very few mental health conditions are suitable candidates for being brain diseases. However, some psychiatrists and philosophers argue that the path from mental disorders to brain disorders is a lot more direct: it simply follows from a scientific world-view that mental disorders are brain disorders. They start out with the plausible assumption that on a materialistic, scientific world-view, the mind is not

independent of the brain. Rather, the brain realizes mental states; whenever there are changes in the mind, something also has to be happening in the brain. The brain is the organ which does the thinking.

Given this assumption about the nature of the mind, they believe that illness of the mind is also illness of the brain. If there is something wrong in the mind, but the mind just is the brain, doesn't it follow that there must also be something wrong with the brain? Whatever is going on in the brain that is producing dysfunctional psychological processes must itself be dysfunctional, so the thinking goes. Led by this kind of reasoning, we even get psychiatrists asserting that to deny that mental disorders are brain disorders implies a form of dualism, positing minds and brains as separate entities and allowing spooky souls back into the picture.

The objection philosophers have raised against this way of thinking is that it neglects an important insight: One and the same psychological state can be realized differently in different organisms. This is called multiple realizability. Pain can have a different neurological physical basis in different kinds of animals, but still be pain. After all, it is caused by the same kinds of things (cuts, burns, etc) and leads to the same kind of behaviour (avoidance, utterances of distress etc). We may also have multiple realization within a species. One and the same psychological state can have different underlying physical states in different individuals (or indeed, in one individual over time).

But if my brain and its processes look very different from yours, and Jimmy's look still different, even though we are all suffering from depression, it seems that there is no one type of brain problem we all three have. Even if we admit that there cannot be any psychological

dysfunction which is not in some way realized in the brain, this does not imply that there is one type of brain process, a dysfunction that we can locate in different individuals. We may not find any *systematical* processes or patterns that are shared by the brains of different people suffering from depression, yet different from what we find in the brains of the non-clinical population. But this is what we would need to justify talk about a brain dysfunction, so the objection goes. Given recent findings in neuroscience, this is not an unreasonable objection. Finding systematic differences between the brains of people suffering from a mental disorder and the brains of people in the non-clinical population is a project that has been met with varying degrees of success in different mental health conditions.

Of course, the advocate of the weak view (we could also call it the 'brain-disorder-by-metaphysical-fiat' view) may now double down and say the following: No matter whether we find a systematic difference that is shared across people suffering from a disorder, it just follows that if what the brain is doing manifests itself in pathological psychological states, then it is itself doing something wrong. This is why I have called this position the weak view: other than being the brain of someone who suffers from mental illness, there are no criteria the brain has to meet to count as disordered.

We could respond to this in a theoretical vein or in a practical one: The theoretical response would be to say that there may be different standards for what counts as pathological at the level of the brain and for what counts as pathological at the level of the mind. It is theoretically possible to say that the brain is doing just what it should be doing, but things have gone wrong at the level of the mind. The hardware software distinction in computers is frequently employed to make this case – there may be nothing wrong at the level of the hardware, but we may still have a problem in the software. A PC and a Mac may both have perfectly functional hardware, but due to some programming error, they both have a word-processing bug.[2]

There is also a practical and, in my view, more pressing objection to the view that we should just label any brain processes underlying mental dysfunction brain dysfunctions: we do not get any descriptive or explanatory mileage out of saying that a mental disorder is a brain disorder if we just make this a definitional issue. On this account, the brain disorder label is just too cheap to come by. So it seems that if we take paradigmatic neurodegenerative brain diseases as a model, mental disorders are not brain disorders. If, on the other hand, we take the view that every mental disorder is automatically a brain disorder, we lose any substantial notion of what it takes to be a brain disorder. The brain processes and states of a person suffering from a mental health condition are just defined as pathological, without any guarantee that there will be anything substantial that brain research adds to describing and explaining the condition.

The Goldilocks reading

Is there still a fruitful notion of mental disorders as involving pathology of the brain to be had? The problem with the strong view was that it modelled the notion of brain pathology in mental disorders on diseases like neurosyphilis. This model does not fit most mental disorders. The weak model, on the other hand, did not even require any systematic brain differences to label the brain processes of those suffering from mental ill health as dysfunctional. But there is a model between these two extremes. When people like the former head of the National Institute for Mental Health in the United States claim that mental disorders are brain disorders, they are not claiming that they are conditions like neurosyphilis. They are also not saying that mental disorders are brain disorders by definition. Rather, they predict that we will, at least in some cases, find ways in which the brains of people suffering

from mental health problems systematically differ from those of healthy individuals, and these differences in brain function or structure can rightly be described as dysfunctional.

Two things are important to note here: First, this is still an empirical bet. The expectation is that we will find significant differences between the clinical and the non-clinical population which are useful in understanding and possibly treating a condition. But only time and further research will show whether we actually do find these systematic differences in, for example, depression. It is theoretically possible that we will not because there is too much variation in how brains realize psychological dysfunction.

The second thing to note is that the kind of approach I have been proposing will make what counts as brain dysfunction dependent on what we count as psychological dysfunction. Let me explain: based on differences in brain function in people suffering from addiction, some scientists claimed that addiction is a brain disease. The initial hope some clinicians and scientists had was that this would validate addiction as a 'proper' disease. In fact, this was supposed to be a major pay-off of the brain disorder view. Psychiatrists who pushed the view that addiction is a brain disease did so partly in order to destigmatize the condition. They hoped that showing that addiction is a brain disease would make society take it seriously as a disorder, rather than writing addicts off as weak-willed, selfish and undisciplined.

However, on the Goldilocks approach I am proposing, finding brain differences does not in itself decide the status of a condition as a genuine illness. If a difference in the brains of those suffering from addiction is only characterized as dysfunctional because of its link to psychological dysfunction, then it cannot be a way of showing that the person is really ill,

rather than just having a problem in living. Bear in mind that we find brain differences associated with non-pathological features of human psychology as well. For example, London taxi drivers exhibit brain differences, with areas involved in memory being larger than in the general population. This is a systematic difference, but not a pathological one – a better spatial memory is a good thing! So when we link brain dysfunction to psychological dysfunction in the way proposed, we do not get an independent way of validating conditions as genuine disorders. This does not mean that conditions like addiction are not a disorders, all it means is that whether they are will not hang on the fact that there are brain differences in addiction. We will still have to do the hard work of deciding whether we want to characterize a condition as an illness or a problem of living.

On the Goldilocks account, mental disorders are brain disorders if we can find dysfunction in the brain, even if that dysfunction is only understood as such because of its link to psychological dysfunction. It presents a more modest view of what might count as a brain disorder than the strong view advocated by Szasz. But it still allows for the notion of brain dysfunction to play a role in our understanding of a mental health condition. When we find out about differences in the way the brain functions in, for example, addiction, this gives us a fuller picture of the condition. It may also help us identify risk-factors, or avenues for treatment. These can, but need not be, through medication. As we have seen in the taxi-driver example, there are many ways of affecting changes in the brain.

Anneli Jefferson is Lecturer in Philosophy at Cardiff University

References

Jefferson, A., 'What does it take to be a brain disorder?', *Synthese* 197 (2020), 249-262

Papineau, D., 'Mental Disorder, Illness and Biological Disfunction', *Royal Institute of Philosophy Supplements 37* (1994), 73-82.

Szasz, T., 'The Myth of Mental Illness', *American Psychologist 15*, (1960), 113-118.

[1] This piece is loosely based on ideas from my recent paper 'What does it take to be a brain disorder?' (Jefferson 2020)

[2] For a very clear presentation of this argument as well as the multiple realizability one, please read David Papineau's (1994) paper on mental illness and biological dysfunction.