

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/126481/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Jefferson, Anneli ORCID: <https://orcid.org/0000-0002-1870-1361> 2020. What does it take to be a brain disorder? *Synthese* 197 , pp. 249-262. 10.1007/s11229-018-1784-x file

Publishers page: <https://doi.org/10.1007/s11229-018-1784-x>
<<https://doi.org/10.1007/s11229-018-1784-x>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



What does it take to be a brain disorder?

Anneli Jefferson¹ 

Received: 12 September 2017 / Accepted: 9 April 2018
© The Author(s) 2018

Abstract In this paper, I address the question whether mental disorders should be understood to be brain disorders and what conditions need to be met for a disorder to be rightly described as a brain disorder. I defend the view that mental disorders are autonomous and that a condition can be a mental disorder without at the same time being a brain disorder. I then show the consequences of this view. The most important of these is that brain differences underlying mental disorders derive their status as disordered from the fact that they realize mental dysfunction and are therefore non-autonomous or dependent on the level of the mental. I defend this view of brain disorders against the objection that only conditions whose pathological character can be identified independently of the mental level of description count as brain disorders. The understanding of brain disorders I propose requires a certain amount of conceptual revision and is at odds with approaches which take the notion of brain disorder to be fundamental or look to neuroscience to provide us with a purely physiological understanding of mental illness. It also entails a pluralistic understanding of psychiatric illness, according to which a condition can be both a mental disorder and a brain disorder.

Keywords Psychiatric illness · Brain disorder · Neuroscience · Hardware/software analogy

1 Introduction

In psychology and psychiatry as well as in philosophy, there has been strong resistance to the idea that all mental disorders are best understood as brain disorders (Graham

✉ Anneli Jefferson
a.jefferson@bham.ac.uk

¹ Department of Philosophy, University of Birmingham, ERI Building, Birmingham B15 2TT, UK

2013a; Miller 2010; Schramme 2013). Objections stem from practical worries about possible repercussions when it comes to treatment or the danger of stigmatisation, but there are also more theoretical objections. A number of authors have argued that mental disorders are autonomous from their physical realization (Boorse 1977; Graham 2013a; Papineau 1994; Schramme 2013). They argue that considerations regarding multiple realizability suggest we should be cautious about inferring that any brain state that realizes a mental dysfunction is *ipso facto* disordered and in assuming that we will in fact be able to find brain anomalies which underlie mental dysfunction for every kind of mental disorder. Drawing on this work, I argue that specific brain differences which are present in individuals suffering from mental disorder but not in those who do not, are a necessary condition for positing that mental disorders are also brain disorders. This means that those mental disorders for which we do not find reliable brain differences should not be labelled as brain disorders. I will then turn to the more controversial question of whether there are any valid reasons to deny that a mental disorder for which we have in fact found a brain difference that underlies the mental dysfunction should count as a brain disorder. This question has been insufficiently addressed in the literature, as many philosophers only use multiple realization arguments to show that we should not necessarily expect systematic differences at the level of the brain.

I will argue that where we find that there are anomalous brain processes or structures which are sufficient for dysfunctional mental processes, we should conclude that we are dealing with a brain disorder. I consider objections according to which inferring brain dysfunction from mental dysfunction is otiose and should therefore be avoided. I reject these arguments, but show that in some cases, my account leaves us with a derivative notion of brain disorders. The autonomy of mental disorders implies that for many psychiatric illnesses, the characterisation of brain dysfunction will be both epistemologically and metaphysically dependent on mental dysfunction. But this does not provide grounds for denying that mental disorders are also brain disorders unless one claims that ‘mental disorder’ and ‘brain disorder’ are mutually exclusive categories, and there are no good reasons to take this step. I will then outline the consequences of my more inclusive approach for the use and extensions of the terms ‘mental disorder’ and ‘brain disorder’.

In Sect. 2, I summarize the argument from multiple realizability and show why it implies that even if we accept physicalism, mental disorders need not be brain disorders. However, I show that brain differences still have an important role to play in validating psychiatric classification. In Sect. 3, I argue that once we have found anomalous brain processes and structures underlying specific mental dysfunction, we should conclude that we are dealing with brain disorders. In Sect. 4, I defend my account against a number of objections.

2 Multiple realizability and why some mental disorders may not be brain disorders

The position that mental disorders are *ipso facto* brain disorders is held by a number of psychiatrists, as in Thomas Insel’s programmatic statement that ‘mental disorders

are biological disorders involving brain circuits that implicate certain domains of cognition, emotion or behaviour' (Insel 2013). But the view can also be found in popular discourse on mental disorder. A post by the *Invisible Illness* Facebook group illustrates this:

We live in a world where if you break your arm, everyone runs over to sign your cast. But if you tell people you're depressed, everyone runs the other way... We are so, so, so accepting of any body part breaking down other than our brains. And that's ignorance. That's pure ignorance. And that ignorance has created a world that doesn't understand depression, that doesn't understand mental health. (Living with invisible Illness 2016)

A theoretical justification for this position takes the truth of physicalism to entail that any mental disorders are at the same time brain disorders. One might endorse an argument like the following: All mental processes are brain processes; X is a disordered mental process; therefore X is a disordered brain process.

But this claim rests on two assumptions: one is that we will be able to find reliable matches between brain anomaly and psychological dysfunction. The second is that dysfunction at the mental level carries over to dysfunction in the physical realizer. The computer analogy has frequently been invoked to cast doubt on both these claims. Looking at the distinction between hardware and software illustrates the point that just as software problems are distinct from hardware problems and the existence of one does not imply the existence of the other, mental problems can be distinct from physical problems (Boorse 1977; Miller 2010; Papineau 1994; Wakefield 2017).

The possibility that mental disorders may be realized in multiple ways in the brain provides a reason to think that attempts to show that mental disorders are brain disorders may remain unsuccessful even as our methods for investigating brain processes and structure improve. If one type of mental state or function can in principle be realized in different ways by the brain, regularities at the psychological level may not have corresponding regularities at the physiological level.

Thus, we cannot count on being able to find the same brain differences across individuals with one and the same psychological diagnosis. We may not even find any physiological abnormalities for the mental disorders we posit. Whether stable underlying physiological correlates can be found for all, or even most, mental disorders is an open empirical question. We will just have to see what the neurosciences find. In some cases, we have already found some interesting correlates between mental and brain processes, for example in the case of addiction (cf. Holton and Berridge 2017) or schizophrenia (Schaffner 2013), however, this does not mean that this will be the case across different psychiatric illnesses.

Finding brain differences that have explanatory and predictive power may not be possible, and if these cannot be found, this would preclude the claim that, for example, depression is a brain disorder. In order to usefully be able to talk about brain pathologies which underlie psychological dysfunction, we need the realizer to be stable across individuals. Otherwise there is nothing systematic we can say at the brain level. But we will still have regularities at the mental level which allow us to classify conditions as mental disorders. It follows that the hypothesis that any given mental disorder is a brain disorder is dependent on empirical findings and not something that can be

decided a priori. The purpose of the current account is not to decide the question which, if any, mental disorders are brain disorders, but to provide criteria for deciding whether a condition should count as both a mental and a brain disorder.

Of course, lack of a specific brain difference that underlies a given mental disorder does not mean that we cannot have a disjunct of anomalies in the brain which are clearly identifiable and can be subsumed under regularities at the mental level. However, the more disparate the brain differences underlying a specific psychological disorder or dysfunction are, the more brain disorders we'd have to posit. Furthermore, if there is no inter-individually stable brain difference, the level of the brain will not be useful for explanatory or treatment purposes. It follows that finding an inter-individually shared brain difference is not necessary for there to be a mental disorder, just as there are software bugs which are instantiated in different ways on different computers. Mental disorders do not rely on being brain disorders for their existence, and a defense of mental disorders as real entities need not be mediated by showing that they are brain disorders (cf. Graham 2014).

It is, however, worth pointing out that while finding brain anomalies that correspond to mental disorders is not essential to validate these mental disorders qua mental disorders, finding stable brain differences that underlie mental dysfunction does support the categorisation and explanation of the mental disorder in question. There is some controversy as to whether current diagnostic criteria for mental disorders cut psychological reality at its joints, i.e. whether current symptom based nosologies identify conditions which have the correct level of grain or whether diagnostic labels are too broad or too narrow.¹ So, while lack of fit between brain findings and findings at the mental level may result from the fact that there are many ways of realising a mental disorder in the brain, another possible cause is that there is a problem in our classification at the psychological level. The extent to which lack of underlying brain anomalies for specific disorders casts doubt on diagnostic categories will depend in part on how valid our diagnostic categories seem to be on the psychological level in the first place.

Whether we find reliable brain/mind associations more generally—not just in the case of disorder—will also be relevant to the question whether finding a brain signature for a specific disease is a realistic expectation. We know that due to brain plasticity, brain lesions can be compensated for to an astounding extent, so that noticeable brain damage is compatible with normal psychological function. Of, course, we cannot infer from the fact that a damaged brain can support a healthy psychology that a healthy brain can support a dysfunctional psychology. But if healthy psychological processes can be realized in such a variety of ways by the brain even in the face of structural damage, it is at least conceivable that there may be similar variation in the way the brain realizes unhealthy psychological processes. This allows for the possibility that it may

¹ For example, the RDoC project is trying to approach mental disorders in a different way, looking at circumscribed domains of functioning, such as cognitive systems, positive valence systems, negative valence systems etc. which are then divided into further subdomains, for example working memory or cognitive control and others in the case of cognitive systems (Cuthbert 2014) These comments are not intended as an endorsement of the RDOC system which has come under heavy fire for a number of reasons, one of which is that critics accuse it of disregarding the patient perspective (Parnas 2014). My aim is merely to show that existing problems in nosology may also lead to problems in finding systematic brain differences, because we may be looking in the wrong place.

be impossible to find systematic differences in the brain that correspond to a certain type of mental dysfunction. In principle, it is possible that mentally ill individuals have brains which do not exhibit specific differences in function or structure compared to those of individuals who are well, especially given the individual differences between brains that exist in the first place. While multiple realizability creates conceptual space for mental disorders or dysfunctions which do not have a stable counterpart in the brain, the extent to which psychological dysfunction is associated with systematic brain differences will only become clear with further research.

We can now see how, and to what extent, mental disorders are autonomous. While mental disorders are clearly realized by the brain, brain dysfunction is not necessary for mental dysfunction, the latter can exist without the former. What is more, the fact that they realize a mental disorder becomes the organizing principle for brain anomalies underlying mental disorders. As Schramme (2013) and others rightly point out, in the case of mental disorders, we decide what counts as dysfunctional by looking at the level of behaviour and mental states. On this model, where we can ascribe brain dysfunction because we find an underlying brain anomaly, this description of the brain as dysfunctional is inherited from the level of the mental. But should we really say that brain dysfunction can be inherited from mental dysfunction, or does brain dysfunction require its own, non-mental criteria?

3 Is a brain difference that realizes a mental dysfunction a brain dysfunction?

In the preceding section, I have shown that we cannot infer the claim that every mental disorder is *ipso facto* a brain disorder from physicalism. I now address the question whether we should ascribe brain dysfunction and disorder in those cases where we do find an underlying anomaly in brain structure and/or function. This question is of crucial importance, as I take the existence of brain dysfunction to be sufficient for the claim that there is a brain disorder. I defend this claim in Sect. 4. Here, I address the question of when we can posit dysfunction at the level of the brain. At first glance, this may seem like a non-issue. If there is an underlying brain anomaly, and it is sufficient for the occurrence of mental dysfunction, surely we are dealing with a brain disorder? Graham (2013a, b) has argued against this on the basis that this conclusion wrongly presupposes that there cannot be separate criteria for what counts as malfunction in the brain and what counts as a malfunction in the mind.

Once again, the computer analogy has been invoked to illustrate the point that just as software problems are distinct from hardware problems, mental problems can be distinct from physical problems (Boorse 1977; Graham 2013a; Papineau 1994; Wakefield 2017). As became apparent in Sect. 2, the computer analogy is helpful in bringing out the fact that the level of the physical need not inherit the ascriptions of dysfunction from the level of the mental just because the mental is realized by the physical. Graham argues that this leaves open the possibility of a healthy brain realising an unhealthy mental process. '[A] person may conceivably suffer from a mental disorder that is in the brain or is, in some sense, neurally neurobiologically/neurochemically/neurodynamically) based or inherent, but this does not

necessarily mean that the disorder is a brain deficit, illness, dysfunction or disease.” (Graham 2013b, p. 517)

Comparing human minds and brains to software and hardware is suggestive and highlights the question whether dysfunction can be inherited across levels of description. But it isn't helpful when it comes to deciding whether a specific brain difference underlying a psychological dysfunction should count as pathological because there is a fundamental disanalogy between brains and computers: We are the ones who invented computers, consequently, we have clear criteria for what counts as a hardware problem. There are specifications the hardware is supposed to meet, and if it does not, there is a hardware problem. We know what constitutes faulty wiring.

However, we do not have a specification for the brain, consequently, what constitutes a hardware problem is less clear cut when we look at the brain. It is worth stressing this point: We find ourselves in fundamentally different positions *vis a vis* computers and brains. We designed computers and so hardware malfunction can simply be read off from the way the hardware is designed and supposed to work. In contrast, we do not have a blue-print or a pre-existing specification for the brain. We are still finding out how the brain functions, and, correspondingly, we are still finding out what constitutes brain dysfunction.²

We do of course have some understanding of what constitutes a brain defect. Prototypical cases of brain defects are lesions, tumors or advanced neurodegenerative conditions, and they are specifiable without reference to the mental level. However, generally the characterization as defective results from a functional description which derives the label of defectiveness from the fact that the neurons or larger brain areas in question behave anomalously when compared to the normal function in comparison brains (Heinrichs 2014). If these differences in function have some kind of adverse effect, they can be described as pathological. Because we cannot just infer from brain difference to brain dysfunction, we need to show that in these cases the brain is not doing what it should be doing, that some normal function is disrupted and this has detrimental effects for the organism. In some cases, this will be done by linking brain anomaly or difference to psychological dysfunction (cf. Jefferson 2014).

Accordingly, some proposed types of brain dysfunction are derived from the psychological level, because they underlie disordered thought, volition etc. This can be illustrated by an example: the fact that the anomalies in the functioning of the amygdala found in psychopaths are labeled as dysfunctions results from the fact that there are mental and emotional impairments associated with the hypo-function we find in the amygdala (Blair 2008; Garson 2011).³ In other words, both epistemically, but also metaphysically, some types of brain dysfunctions derive their classification as dysfunctional from the psychological dysfunction they realize. But the fact that their

² Matters are further complicated by the fact that both the notion of biological function and the concept of dysfunction in medicine are highly contested (cf. Godfrey-Smith 1993; Nanay 2010; Neander 1991; Wakefield 1992). I will steer clear of these troubled waters in this paper.

³ For the purposes of the current discussion it isn't important what notion of dysfunction is embraced. We can work with a fairly minimal notion which characterises dysfunction as an anomaly where a trait or a system is not doing what it is supposed to do. Whether we flesh out what a system is supposed to do via a Cummins style functional analysis (Cummins 1975) of normally operating comparison traits or systems, a biostatistical account or an evolutionary account (Wakefield 1992) is irrelevant for current purposes.

dysfunctionality derives from the mental level of description does not make them ineligible to be brain dysfunctions, after all, the brain is the organ of the mind.

The central role the mental level plays for the description and individuation of certain types of brain dysfunction and the fact that brain dysfunction is inherited from the mental means that brain dysfunction is dependent on mental dysfunction. This in turn implies that the notion of brain dysfunction cannot be fundamental in psychiatric illness: we cannot move from an independent conception of brain dysfunction to the claim that there is a mental disorder. This also means that any hopes of making psychiatric illnesses more scientifically respectable by showing that they are also physiological dysfunctions is misplaced. While this is not a goal pursued by all proponents of biological psychiatry, it has been proposed that the only way for psychiatry to attain scientific respectability is by showing that mental disorders are physical disorders (Szasz 1960). If, as I have argued, disorders like, for example, depression are only brain disorders in as far as there are stable realizers of mental disorder to be found at the level of the brain, this thought needs to be given up.⁴ Finding underlying brain anomalies and dysfunctions is of course of some value to boost a diagnostic category's scientific respectability because finding differences in the brain supports the classification of mental dysfunction, as discussed in Sect. 2. It may also be helpful for prevention and treatment, but the normative criteria for dysfunctionality of mental disorders originate at the level of the mental (cf. Broome and Bortolotti 2009).

4 Clarification and response to objections

My claim that brain disorders can be essentially dependent on being the realizers of mental disorders requires further clarification as it yields a notion of brain disorder which is broader than our prototypical understanding of what brain disorder is. This also invites a number of objections. I will flesh out the notion of brain disorder that I propose by showing how my account deals with important objections. The criticisms I consider are all variations on one theme, which is that we already have a notion of brain disorder, and that mere presence of brain dysfunction in association with mental disorder will allow lots of conditions that do not fall under this concept to be labeled as brain disorders.

I readily concede that the account I propose has the potential to include conditions which do not fit our prototype of 'brain disorder', which is that of classic brain disorders such as Alzheimer's dementia or brain tumours. However, matching the prototype of the concept 'brain disorder' cannot be a necessary condition for being an instance of a brain disorder, because prototype theories of conceptual content do not provide necessary conditions that specify what falls under the extension of a concept. Rather, they give us typical instances (cf Margolis and Laurence 2014). For example, a penguin does not fit the prototype of bird (because it does not fly), but it is nevertheless a bird. So any criticism of the proposed account of brain disorders needs to argue for something

⁴ I take Hanna Pickard to be making a related point when she claims that while we may be able to find the physical basis of mental disorders have a scientific explanation for them, their status as mental disorders relies essentially on mental dysfunction, so this wouldn't vindicate them as the kind of physical disorders that Szasz takes to be real disorders (Pickard 2009).

stronger, namely that my proposal does not capture necessary conditions that need to be met in order for a disorder to qualify as a brain disorder. I now turn to the objections against my proposal.

4.1 The brain science objection

Arguably, part of our proto-typical understanding of brain disorders is that they are the disorders that brain science deals with. George Graham objects against brain disorders which essentially rely on mental dysfunction for their status as disorders on the basis that brain sciences don't deal with the mental, and that therefore a disorder that makes essential reference to the mental does not fit into the research framework of the brain sciences. Graham claims that de facto, the brain sciences don't deal with the mental and the corresponding conception of brain disorders is purely physiological. From this he concludes that essential reference to mental problems in the description of a brain condition as dysfunctional disqualifies said condition from being a brain disorder: "If a disorder's foundations and powers cannot be best described or sufficiently understood in brain science terms alone, if, that is, such terms are insufficient, and need to be complemented by, say, psychological terms then the condition should not be thought of as a brain disorder." (Graham 2013b, p. 520)

But why should the fact that some parts of brain science are entirely mechanistic and do not look at the mental imply that we cannot have a mentalized brain science? At best the argument rests on an unjustified bias for the status quo, at worst the description of the status quo itself is mistaken.

I would argue that we already do, to a large extent, have a mentalized brain science, as both cognitive neuroscience and biological psychiatry attest to. The claim that brain science stops being brain science if supplemented by psychology is needlessly focused on a subset of neuroscience and on molecular biology. It is similar to the claim that mental disorders cannot be brain disorders because classic instances of brain disorders are neurodegenerative disorders which get progressively worse. These kinds of arguments rely on some features of current practice or, in the latter case, on exemplary cases or prototypical instances of what we take to fall under the extension of the term 'brain disorder' and assume that the status quo determines the future extension of the term 'brain disorder'. But of course concepts are changeable and it is unsurprising that in fields which are still very much in the process of development, such as psychiatry and neuroscience, the extension of terms changes with discovery.

Brain disorders are those disorders which involve brain dysfunction and can be usefully examined, understood or treated with the help of the brain sciences. Given rapid changes in these sciences over the last decades, there will be a corresponding change in both the intension and the extension of the term 'brain disorder'.⁵ If the claim that brain dysfunction is derivable from mental dysfunction via the discovery of

⁵ Some may want to argue that we won't have a change of extension, rather, we will find out that further conditions fall under the extension of 'brain disorder'. I suspect that given the associated change in use of the term 'brain disorder' under the account I am proposing, a change of extension is more likely. In any case, we can agree that we are likely to end up labelling conditions as brain disorders which we don't currently take to be brain disorders.

underlying brain anomaly is true, then our understanding of brain science as the basic science under which psychiatric illness becomes subsumed will have to change.

4.2 The non-fundamentality of brain dysfunction objection

A related objection dismisses a category of brain disorders which relies on mental dysfunction with the following justification:

Given the purpose or intent behind the broken brain view, it is contrary to its scientific aspiration to propose that ‘brain disorder’ should be defined first in terms of the conceptually prior notion of mental disorder, and then and then only recommend that the notion of a mental disorder should be re-described as the criterion for a brain disorder, all the while continuing to use the criterion of a mental disorder as the criterion of a brain disorder. The whole effort of identifying a mental with a neural impairment would be viciously circular.” (Graham 2013a, p. 24)

One objection that can be extracted from this quote is that the notion of brain disorder is supposed to replace the notion of mental disorder and that on the characterization I have given, this is not possible. Graham’s criticism presupposes that the terms ‘mental disorder’ and ‘brain disorder’ should be understood as mutually exclusive or that the understanding of a disorder as a brain disorder is supposed to supersede the notion of that condition as a mental disorder. The school of thought which takes brain disorders to be more basic or fundamental can be found in popular thinking about psychiatric illness, as when people characterize depression as a chemical imbalance in the brain, but also in certain scientific projects.

While eliminativist and certain kinds of reductionist proposals do pursue the goal of showing that ‘mental disorders are really brain disorders’,⁶ it is clearly precluded by the relation between mental disorder and brain disorder that I have been proposing. In short, I agree with Graham that we cannot eliminate the notion of mental dysfunction if we make the individuation of brain disorders dependent on it. And, like Graham, I do not think replacing the notion of mental disorder with that of brain disorder is an appropriate goal: “The best picture of a mental illness is not likely to be found in a single, precise, biologically privileged ‘frame’ (viz. a biological marker). The best picture is more likely to be found in the overall manner of organizing the most useful perspectives about an illness that we have or otherwise achieve” (Graham and Flanagan 2013). Philosophers such as Dominic Murphy, who do favour thinking of mental disorders as brain disorders (Murphy 2005, 2013) also explicitly reject an understanding which takes the brain level as fundamental and instead champion a multi-level approach which combines different levels of descriptions and explanation. In sum, what this criticism shows is that a certain notion of brain disorder is not viable, but that is not the notion I am putting forward.

⁶ Murphy (2005) mentions Kandler as a proponent of this type of view.

4.3 The insubstantiality objection

There is, however, another worry in the offing here, which is that the notion of brain disorder that we are left with on my account is insufficiently substantial. This is a stronger objection than the one I discussed above in 3.1., which merely asserts that we presently use the term ‘brain disorder’ in such a way that it doesn’t fit the kind of derivative brain disorders that I propose. The more worrying objection maintains that the notion of brain dysfunction that remains is too unsubstantial to justify the label of ‘brain disorder’, because it does not provide a brain-specific notion of brain dysfunction.

One motivation for hanging on to a separate category of brain disorders proper may be that there are some disorders where we have a brain science specific criterion for what constitutes brain dysfunction, such as in neurodegenerative disorders. But even here, matters are slightly more complicated. Establishing when certain brain processes count as dysfunctional may well require that we look at consequences for the whole organism. For example, neurodegenerative diseases such as Alzheimer’s and Parkinson’s derive their status as pathological not just from what happens at the level of cells, but also from the effects on the memory and motor-system, and the fact that in the case of Alzheimer’s, the condition eventually leads to death. Effects on the body or on the mind play a central role in classifying even classic brain diseases as diseases. Alzheimer’s is both a mental disorder and a brain disorder, because it involves mental dysfunction and brain dysfunction. Similarly, autism appears to be a brain disorder to the best of our knowledge, but it is nevertheless a mental disorder, because it is associated with certain mental dysfunctions.⁷

In the case of Alzheimer’s dementia, the mental dysfunction is part of what makes that illness an illness, but nobody would claim that it is not a brain disorder because of that. One might object that the reason why Alzheimer’s counts as a brain disorder is that the changes in the brain will eventually lead to death. We therefore have a criterion for something being wrong in the brain *independently* of the mental effects of the changes in the brain.

I find the suggestion that fatality is a necessary condition for disease status dubious. If there was an Alzheimer’s*, which also had a clear brain signature that explained the memory loss without being fatal, would we want to deny that Alzheimer’s* is a brain disorder? This would require a further argument, and I can’t see what a good argument would look like. In both cases, a brain difference underlies a mental dysfunction.

4.4 The psychiatry-as-a-practical-discipline objection

In my arguments, I have very much focused on the question whether we should label certain brain processes as dysfunctional, assuming that if we can say that there is something wrong in the brain, we are justified in talking about a brain disorder. This assumes that what determines what kind of disorder we are dealing with is where

⁷ There is of course some controversy whether high functioning autism is a disorder at all, or merely a different way of being and thinking. My point is merely that in as far as it is a disorder, it is both a mental and a brain disorder.

we locate the *problem*. But, given that medicine is a practical discipline, shouldn't the question of whether a disorder is a brain disorder be answered by considering the question what discipline contributes most to the explanation and above all, the treatment of a disorder? Wakefield rejects this assumption out of hand, saying: "Critics of the disease model often incorrectly assume that the nature of treatment is tightly tied to the conception of the problem." (Wakefield 2017, p. 10) While I am sympathetic to this line of thought, I cannot adequately address the question whether psychiatric classification is best understood as targeting the location of a problem or should be understood more broadly and pragmatically to encompass the genesis of a condition and the way a condition is treated. What I can however show is that even if we do take treatment and causal explanations as central to the question whether a condition is a brain disorder, we will still reach the conclusion that many conditions are both mental and brain disorders.

The assumption that saying that a condition is a brain disorder has repercussions for psychiatric practice is implicitly present both in proponents and detractors from the 'mental disorders are brain disorders' view. For example, Miller (2010) strongly objects to the movement to classify mental disorders as brain disorders precisely because he believes this will lead to a prioritization of pharmaceutical treatment over talking therapies. The thought that how a condition is treated decides what kind of condition it is, is once again bolstered by appeal to the computer analogy. Hardware problems are the ones that are fixed mechanically, whereas software problems are fixed by programming. It is of course true, as Papineau (1994) points out, that we would be foolish to call a hardware engineer to fix a programming problem. But this isn't sufficient (and, to be fair, Papineau doesn't claim it is) to show that only problems which are addressed by fixing the hardware are hardware problems. To make matters more complicated, hardware problems can be 'fixed' by a software work-around, as was the case with patches that were developed for the Pentium bug.

Leaving aside this further wrinkle, even an appeal to the way a condition is treated will leave most psychiatric conditions somewhere at the intersection between being hardware and software problems. Kendler (2012) shows this for schizophrenia, major depression and alcohol addiction, where risk factors and points of intervention range from biological, psychological to higher order factors such as social or cultural ones. He concludes from this that we should embrace empirical pluralism regarding psychiatric illnesses and that we should adopt the biopsychosocial conception of psychiatric disorders.

The observant reader may wonder whether what I have been arguing for is just the biopsychosocial model without the label. However, this model is not what is at stake here. First of all, the biopsychosocial model is intended to apply both to our understanding of somatic medicine and psychiatry, as it is concerned with taking patients' perspectives, the way they experience their condition and the effect of social risk factors on their condition seriously (Engel 1977). These considerations apply whether you suffer from diabetes or depression. Choosing between a biomedical and a biosocial model therefore seems to be an orthogonal issue. Furthermore, the exact remit of the biopsychosocial model is controversial in and of itself (Davies and Roache 2017), so it isn't a useful concept for clarifying the already complicated issues at stake. Finally, I have not addressed social factors in my discussion at all, as I am focusing

on the mind/brain relation.⁸ In conclusion, looking at medical practice will in most cases give a mixed result as well, so that conditions cannot be clearly separated into mental disorders and brain disorders. To take but one example, we treat depression both pharmacologically and via talking therapy. What is more, given the fact that psychotherapy, too, changes the brain, it would be odd to say that a problem at the level of the brain cannot be addressed via the mental. The upshot of this is that even if we did think that it was the level of risk factors and treatment that decides whether we are dealing with brain disorders, many psychiatric illnesses would end up being both.

5 Conclusion

On the understanding of brain disorders I have proposed, a condition can be a brain disorder even if mental dysfunction is essential to it and even if it is best treated by psychotherapy. All that is required is an identifiable brain dysfunction, which will frequently rely on mental dysfunction for its characterization as pathological. While some may find the resulting understanding of brain dysfunction unpalatably thin and derivative, this worry can be mitigated by realising that in paradigmatic cases of brain disorder, too, we look at the effect on the organism as a whole in order to establish the presence of dysfunction and disorder. Given the fact that the brain is the organ of thought, it is unsurprising that we would look at the level of mental functioning to establish that certain brain differences are in fact pathological. Other brain disorders, such as Parkinson's disease and Alzheimer's dementia will be not at all or only partially dependent on mental dysfunction, as they are associated with motor problems (Parkinson's) or further problems (as in Alzheimer's, which eventually leads to death).

On this more inclusive understanding of brain disorders, the extension of the term 'brain disorder' would broaden to include those mental disorders where we can find brain dysfunction which is sufficient for the occurrence of the corresponding mental dysfunction. Finally, the account of brain disorders I propose does not put mental disorders on a more scientific footing by showing that they are in fact physical disorders. Given the relation between mental dysfunction and brain dysfunction, this would be a mistaken goal. However, the discovery of underlying brain dysfunction does support our categorization of mental dysfunction in those cases where we find systematic differences at the level of the brain that correlate with the psychological dysfunction we have posited.

Acknowledgements This paper is part of the project 'Mental Disorders, Brain Disorders and Moral Responsibility', generously supported by the Leverhulme Trust (Grant No. ECF-2015-493). An earlier version of this paper has been presented at the BSPS annual conference. I would like to thank Marion Godman, Birgit Beck, Jan-Hendrik Heinrichs, Hanna Pickard, Lisa Bortolotti and Lucy Campbell for comments

⁸ While I do not have space to address the social dimension adequately, I do want to make a brief comment. Saying that a condition is a brain disorder because it entails dysfunction of the brain does not entail that there are no psycho-social risk factors and causal contributors for the condition. Type 2 diabetes and obesity are physical conditions, but it is well established that in rich countries, these conditions are strongly linked to socioeconomic factors.

on earlier drafts of this paper. I would also like to thank the two anonymous reviewers at *Synthese* for their feedback. Special thanks go to David Jefferson for expert advice on hardware and software malfunction.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Blair, R. J. R. (2008). The amygdala and ventromedial prefrontal cortex: Functional contributions and dysfunction in psychopathy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1503), 2557–2565.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44(4), 542–573.
- Broome, M., & Bortolotti, L. (2009). Mental illness as mental: A defence of psychological realism. *Humana.Mente*, 11, 25–44.
- Cummins, R. C. (1975). Functional analysis. *Journal of Philosophy*, 72, 741–764. <https://doi.org/10.1086/289488>.
- Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1), 28–35. <https://doi.org/10.1002/wps.20087>.
- Davies, W., & Roache, R. (2017). Reassessing biopsychosocial psychiatry. *The British Journal of Psychiatry*, 210(1), 3–5. <https://doi.org/10.1192/bjp.bp.116.182873>.
- Engel, G. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, 196, 129–136.
- Garson, J. (2011). Selected effects and causal role functions in the brain: The case for an etiological approach to neuroscience. *Biology and Philosophy*, 26(4), 547–565. <https://doi.org/10.1007/s10539-011-9262-6>.
- Godfrey-Smith, P. (1993). Functions: Consensus without unity. *Pacific Philosophical Quarterly*, 74, 196–208.
- Graham, G. (2013a). *The disordered mind: An introduction to philosophy of mind and mental illness*. London: Routledge.
- Graham, G. (2013b). Ordering disorder: Mental disorder, brain disorder, and therapeutic intervention. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (pp. 512–530). Oxford: Oxford University Press.
- Graham, G. (2014). Being a mental disorder. In H. Kincaid & J. Sullivan (Eds.), *Classifying psychopathology*. Cambridge, MA: MIT Press.
- Graham, G., & Flanagan, O. (2013). *Psychiatry and the brain*. OUP Blog. <https://blog.oup.com/2013/08/psychiatry-brain-dsm-5-rdoc/>. Accessed 12 Jan 2018.
- Heinrichs, J.-H. (2014). On the relation of psychiatric disorder and neural defect. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.00040>.
- Holton, R., & Berridge, K. (2017). Compulsion and choice in addiction. In N. Heather & G. Segal (Eds.), *Addiction and Choice* (pp. 153–170). Oxford: Oxford University Press.
- Illness, L. w. I. (2016). *Post*. Retrieved from <https://www.facebook.com/LivingWithInvisibleIllness/photos/a.1576790999261551.1073741828.1576780732595911/1664596713814312/?type=3&theater>. Accessed 22 Jan 2016.
- Insel, T. (2013). *Transforming diagnosis*. Retrieved from <https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml>. Accessed 19 Aug 2017.
- Jefferson, A. (2014). Mental disorders, brain disorders and values. *Frontiers in Psychology*, 5, 130. <https://doi.org/10.3389/fpsyg.2014.00130>.
- Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Molecular psychiatry*, 17(4), 377–388.
- Margolis, E., & Laurence, S. (2014). Concepts. In E.N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2014 Edition). <https://plato.stanford.edu/archives/spr2014/entries/concepts/>. Accessed 17 Jan 2018.

- Miller, G. A. (2010). Mistreating psychology in the decades of the brain. *Perspectives on Psychological Science*, 5(6), 716–743. <https://doi.org/10.1177/1745691610388774>.
- Murphy, D. (2005). *Psychiatry in the Scientific Image*. Cambridge: MIT Press.
- Murphy, D. (2013). The medical model and the philosophy of science. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Sadler, G. Stanghellini, & T. Thornton (Eds.), *Philosophy and psychiatry* (pp. 966–986). Oxford: Oxford University Press.
- Nanay, B. (2010). A modal theory of function. *Journal of Philosophy*, 107(8), 412–431.
- Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, 58(2), 168–184.
- Papineau, D. (1994). Mental disorder, illness and biological disfunction. *Royal Institute of Philosophy Supplements*, 37, 73–82. <https://doi.org/10.1017/S135824610000998X>.
- Parnas, J. (2014). The RDoC program: Psychiatry without psyche? *World Psychiatry*, 13(1), 46–47. <https://doi.org/10.1002/wps.20101>.
- Pickard, H. (2009). Mental illness is indeed a myth. In L. Bortolotti & M. Broome (Eds.), *Psychiatry as cognitive science: Philosophical perspectives*. Oxford: Oxford University Press.
- Schaffner, K. (2013). Reduction and reductionism in psychiatry. In K. W. M. Fulford, D. Martin, G. T. G. Richard, G. George, Z. S. John, S. Giovanni, T. Tim, & F. S. Kenneth (Eds.), *The Oxford handbook of philosophy and psychiatry*. Oxford: Oxford University Press.
- Schramme, T. (2013). On the autonomy of the concept of disease in psychiatry. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2013.00457>.
- Szasz, T. (1960). The myth of mental illness. *American Psychologist*, 15, 113–118.
- Wakefield, J. C. (1992). Disorder as harmful dysfunction—A conceptual critique of DSM III-R's definition of mental disorder. *Psychological Review*, 99(2), 232–247.
- Wakefield, J. C. (2017). Addiction and the concept of disorder, part 2: Is every mental disorder a brain disorder? *Neuroethics*. <https://doi.org/10.1007/s12152-016-9301-8>.