

Brain Pathology and Moral Responsibility

Forthcoming in Matt King and Joshua May (eds): *Agency in Mental Disorder: Philosophical Dimensions*. Oxford University Press

Anneli Jefferson

Cases of mental illness and neurosurgeons who alter people's brains in ways that make them do unspeakable things are common in the philosophical literature on moral responsibility, often used to test the limits of responsibility. A common thought is that, if one's action is the result of something messing with one's brain, be it a brain disease or a nefarious neurosurgeon, this counts at least as a *prima facie* excuse or exemption from being blameworthy (or praiseworthy) for that action. A thought experiment employed by Lawrie Reznek (1997) and discussed by Neil Levy (2007b) illustrates the common intuition that brain disorders excuse from moral responsibility: a boy, Billy, is born with a slow-growing brain tumor and becomes aggressive and criminal as the result of the tumor. Levy claims that on finding out about Billy's tumor, we would not blame him for his immoral behavior. Here is the case:

As Billy grows, he develops a character which is formed, in part, by the presence of the tumor; it causes him to be aggressive and selfish. In his teens, he is involved in a string of increasingly serious crimes, culminating in a bungled bank robbery, hostage taking, and shoot out with police. Billy is fatally wounded. Now, it is clear that when his tumor is discovered at autopsy, we would cease blaming Billy for his vicious behavior. (Levy 2007, 134)

My question in this chapter is whether we should cease blaming people like Billy for vicious behavior. In what way, if any, do brain pathologies such as a tumors excuse immoral behavior? Time and again, brain pathology has been taken to undermine responsibility-relevant characteristics such as control, whether in the context of addiction (as in Harry Frankfurt's (1971) example of the addict whose desire for the drug is effective because he is physiologically addicted) or brain tumors which undermine people's self-control (Burns and Swerdlow 2003). I will argue that the inference from brain pathology to reduced or absent responsibility is very indirect. Brain pathology matters to moral responsibility only in so far as it underlies and provides further evidence for psychological dysfunction that is relevant to responsibility.

However, according to a common intuition, brain tumors also matter because of the way they *cause* problematic dispositions and behavior. I argue that generally, causal history does not matter for moral responsibility, rather, responsibility rests on psychological capacities, regardless of how these are caused. However, the way psychological capacities change in brain disorders does have implications for moral responsibility, but in a way that has not been appreciated to date. Both in the case of classic brain disorders, such as tumors, and in the case of psychiatric illness, what matters to moral responsibility is not just the fact that a person's intrinsic psychological capacities change. Rather, responsible agency can also be affected by the way these intrinsic changes affect relational aspects of moral responsibility: it affects how successfully an individual's moral agency can be supported by their social environment. The

resources we employ to find out how we should behave and to control our behavior are partly external, we rely on others to support our moral agency in numerous ways. When someone's moral psychology unexpectedly and sometimes drastically changes, both the individual and their social environment are unequipped to deal with these changes. This means that one's brain dysfunction can affect responsibility even if one's psychological capacities would still be sufficient for moral responsibility in the right kind of environment.

I provide a definition of brain pathology as well as a rough outline of the way mental illness (including brain pathology) can excuse in section 1. In section 2, I argue that the role of brain pathology for responsibility ascriptions should be an evidential one, knowledge of brain pathology can provide corroborating evidence for psychological dysfunction relevant to responsibility. I illustrate this with a case study. Finally, in section 3, I consider the claim that brain disorders matter for moral responsibility because they change an individual's moral psychology in a way that is beyond their control. While control over psychological changes is not an excusing factor, as I show, brain disorders may mitigate moral responsibility because they confront individuals with new psychological deficits or urges for which their previous moral education and existing external and internal moral resources have not prepared them.

1. Preliminaries

1.1. Why mental disorders or brain disorders might excuse

The idea that mental or brain disorders can provide a reason to excuse or exempt individuals from responsibility is commonplace in the moral responsibility literature, though the exact reasons why this should be the case are frequently unclear. While the aim of this paper is not to defend a specific theory of responsibility, it is necessary to make some minimal claims about what it takes to be morally responsible and why mental illness or brain disorder might impair one's moral responsibility. Rightly or wrongly, many philosophers have claimed that mental illness or brain disease undermine moral responsibility. This may happen in two key ways: when the ability to understand the nature and moral implications of one's actions is affected (and thus the ability to meet the knowledge or *epistemic condition* for moral responsibility), or when control of behavior is impaired, and one no longer meets the *control condition* for moral responsibility (cf. eg Brink and Nelkin 2013). Both conditions can be mapped onto a reasons-responsiveness account of moral responsibility, according to which an agent is responsible if they are responsive to moral reasons, where this involves both the ability to recognize moral reasons and to act on them (Fischer and Ravizza 1998).

The ability to recognize reasons is affected when psychiatric illness prevents individuals from understanding the permissibility of what they are doing because they are mistaken about external reality. For example, in the grip of psychosis patients might mistake non-threatening behavior as threatening,

believe humans are aliens or robots and therefore sincerely and non-culpably take themselves to be acting in self-defence.¹ Knowledge can be affected more narrowly if illness prevents someone from understanding the wrongness of their actions. Some authors have argued this is the case for psychopaths (Fine and Kennett 2004, Shoemaker 2011).

Psychiatric illness might also affect individuals' control over action, by making impulse control significantly more difficult, thereby diminishing moral responsibility without completely removing it. On some accounts, this is what happens in addiction.² Finally, mental illness can even lead to behavior that cannot be described as reasons-guided action at all, because it involves a reflex, sleepwalking, a compulsion, or similar cases in which control is undermined.

The question how frequently mental illness undermines individuals' capacity to recognise and respond to moral reasons for action is increasingly debated (King and May 2018, Arpaly 2005, Pickard 2015, Broome, Bortolotti, and Mamei 2010). For now, we can content ourselves with noting that in as far as one, or both, of the conditions of understanding and controlling one's actions are undermined by mental illness, moral responsibility is diminished. I have defined the capacity necessary for moral responsibility as reasons-responsiveness, which breaks down into a control condition and an epistemic condition. One other prominent account of moral responsibility defines responsible action in terms of deep, or real, selves (Frankfurt 1971): actions and desires have to be endorsed by an agent and reflect who they *really* are in order for the agent to be responsible for them. I will leave these positions to one side, as I think that—irrespective of whether one endorses these views—reasons responsiveness constitutes a necessary condition for responsible agency. I am also skeptical about the distinction between deep and non-authentic selves. Especially in the case of irreversible mental health conditions, it is morally problematic to claim that the disordered self is not the agent, as it forces us to discount the only self that the agent currently has (or is).

1.2. The role of the brain and the nature of brain dysfunction

In order to assess the relevance of brain pathology to moral responsibility, we need to clarify the notion of brain pathology and its relation to mental disorders.

Paradigm cases of brain pathology are neurodegenerative diseases, brain trauma, and brain tumors, where there is a clearly identifiable physiological problem with the brain. But many psychiatrists and neuroscientists also assume that most, if not all, mental disorders involve brain pathology and see identifying these putative pathologies as an important task of the brain sciences (Insel 2013, Cuthbert

¹ Though see (Broome, Bortolotti, and Mamei 2010) for some further complexities in these kinds of cases.

² See (Pickard 2015) for a convincing argument that in addiction, control is compromised but not undermined to the extent that responsibility is lacking.

2014). By contrast, others have argued for the claim that mental disorders and brain disorders are distinct categories and that a condition cannot be both a mental disorder and a brain disorder (Graham 2013). I hold the view that brain differences should count as dysfunctions if they reliably realize a psychological dysfunction. It is a further empirical question whether we will be able to find these kinds of brain anomalies for many mental disorders (Jefferson 2020); for example, people doubt that this will be possible for a condition like depression, which can take many different forms.

Whether one endorses an inclusive concept of brain disorder or not will depend on a number of issues in the philosophy of science which need not concern us here. The important thing is that appeals to brain differences feature regularly in arguments concerning self-control or moral understanding: in the case of addiction, people appeal to changes in the brain's reward system (Baler and Volkow 2006); in the case of psychopathy, they cite amygdala dysfunction as a factor that reduces responsibility (Glannon 2008). I will follow current practice and talk about such differences as dysfunctions, but one could in principle recast the argument in terms of brain differences that underly psychological dysfunction. For the discussion about moral responsibility what matters is the relationship between brain difference and psychological dysfunction, not whether we call this difference dysfunctional at the level of the brain. But it is important to recognize that not all brain differences or defects will ground mental dysfunction, and some mental dysfunctions will not be relevant for responsibility, because they affect areas of cognition and perception which are not relevant to moral judgment and decision-making (Jefferson and Sifferd 2018).

1.3. The relation between brain pathology, psychological dysfunction and moral responsibility

When do brain differences excuse? It needs to be the case that a certain type of brain difference realizes or causes a psychological problem that leads to diminished or absent responsibility. The relation between psychological dysfunction, brain pathology and moral responsibility can play out in (at least) three possible ways.

First, we could have psychological dysfunction which is relevant to moral responsibility but not associated with brain pathology. Assume, for example, that we cannot find any type of brain difference associated with a certain psychological dysfunction, and that there is nothing systematic to be said about how that dysfunction is realized in the brain. This is not an unlikely scenario, given that mental states can be realized by many different sorts of physical states (Papineau 1994, Schramme 2013), though the extent to which this is the case is an empirical question. Currently, many scientists believe depression is unlikely to be associated with specific brain differences (Radden 2018). In such cases, whatever token brain processes realize the dysfunctional mental states leading to diminished responsibility would still be causally relevant to any lack of responsibility. However, because there is no specific type of brain pathology underlying the psychological problems and there might have been a different neurobiological basis of such mental dysfunctions in another person (or within the same person at a different time), these are cases of causal relevance without any explanatory power. There is then no *type* of brain pathology that is relevant for

moral responsibility. All the explanatory work happens at the level of the mental and is based on an assessment of the individual's psychological dysfunctions.

Conversely, not every case of brain pathology will be associated with psychological problems relevant to moral responsibility. Some brain lesions, tumors, or cysts have no detectable effect at all. Others primarily affect motor functions or perceptual functions that are not relevant to moral judgment and action (e.g. motor neuron disease).

Finally, there will be cases where we can establish a relationship between specific brain dysfunction and psychological dysfunctions that disrupt capacities necessary for moral agency. For example, it has been suggested that functional differences in the amygdala and the orbitofrontal cortex of psychopaths show that they lack the empathy and impulse control necessary for full moral responsibility (Levy 2007b, Sifferd and Hirstein 2013). In order to establish whether brain pathology is relevant to moral responsibility, we thus need to a) show that a specific type of brain dysfunction underlies a specific psychological dysfunction; and b) show that the type of psychological dysfunction is relevant to moral responsibility and to the reasons that justify moral praise or blame for certain actions. As Nicole Vincent points out: "Neurological conditions do not undermine responsibility simply by virtue of being disorders, but rather they do so in virtue of the effect which they have on our mental capacities . . . which are required for moral agency" (Vincent 2008, 200). We may, for example, suffer from depression and low mood while still being able to recognize and react to moral reasons. The finding that someone suffers from a mental or a brain disorder needs to be supplemented by evidence that they exhibit psychological dysfunctions that are relevant to their moral responsibility, either across the board or in a specific situation.³

There may well be a *de facto* tendency to think that the mere presence of brain dysfunction in mental disorders shows that individuals are somehow ruled by brute physical mechanisms and are no longer reasons responsive or else lacking in control. But, as Nomy Arpaly (2005) has argued, this jump in reasoning is unjustified if there is no further evidence that the disorder has these effects. It follows that brain pathology should affect our moral responsibility judgments by informing our judgments of psychological dysfunction relevant to moral responsibility.

2. Brain pathology as evidence for mental dysfunction

When there is a link between specific brain anomalies and mental dysfunction, it is natural to think that our knowledge of brain dysfunction can inform our judgments of individuals' morally relevant mental dysfunctions, e.g. problems with impulse control or empathic deficits. For example, brain differences might be evidence for executive function problems, and this might be relevant for moral responsibility because an agent's ability to control their reactions or to attend to morally salient features of a situation

³ For detailed defences of the claim that we cannot simply move from a psychiatric diagnosis to a verdict on moral responsibility, see (King and May 2018) and (Jefferson and Sifferd 2018).

are implicated (for example, in the context of autism or psychopathy). Indeed, brain data are currently used as evidence for psychological deficits in legal defenses (Catley and Claydon 2015). A recent review article states that “In 2012 alone, over 250 judicial opinions (...) cited defendants arguing in some form or another that their ‘brains made them do it’”. (Farahany 2016, 486)

Legal responsibility is of course not the same thing as moral responsibility, but there is overlap between the deficits that provide candidate excusing conditions in the moral and the legal realm, such as understanding what one is doing and that it is (legally or morally) wrong. Neuroscience and knowledge of brain pathology can inform our assessment of legal and of moral responsibility by giving us more information on a certain condition and its psychological profile.

This can be illustrated with reference to a class of non-pathological cases. In deciding that teenagers should not be given the death penalty, the U.S. Supreme court appeared to be influenced by brain data. The evidence cited showed that teenagers’ brain development is not yet complete (Sifferd 2013), and they therefore tend to have less impulse control and planning abilities than adults. If insufficiently developed impulse control decreases moral responsibility, then the brain differences are further evidence that teenagers are not responsible to the same extent as adults. Similarly, brain evidence is used in the context of psychopathy to bolster the claim that psychopaths have responsibility-relevant affective deficits and problems with impulse control (Glannon 2008, Levy 2007b). However, Stephen Morse has prominently argued that the usefulness of these data for establishing psychological deficits is limited. Regarding the case of adolescents, Morse insists that we already knew that adolescents have worse impulse control, and that the brain findings do not add anything new.

“What did the neuroscientific evidence about the juvenile brain add? It was consistent with the undeniable behavioral data and perhaps provided a partial causal explanation of the behavioral differences. The neuroscience data was therefore merely additive and only indirectly relevant to the behavioral criteria for responsibility” (Morse 2011a, 853)

I concur with Morse on this point, but it is worth mentioning that neuroscientific evidence may have helped shift people’s thinking, by showing just how deep the differences between adults and adolescents go, as Katrina Sifferd points out (personal communication). So, while the neuroscientific data *should* not have made much of a difference in this case, given what we already knew, it is plausible that they did in fact. One might object to this on the basis that the brain difference gives us additional information in the following sense: it shows us why adolescents are more impulsive, namely because their brains aren’t fully developed. However, this fails to substantially change the evidential dynamic, for several reasons.

Evidence of brain pathology would rightly revolutionize our ascriptions of moral responsibility if knowledge of brain pathology could give us insights into mental deficits independently of what we could have gleaned through behavior or self-report. However, the way we currently establish the relationship

between psychological dysfunction and brain difference makes this goal unrealistic for the time being. To find out what constitutes normal and abnormal brain function, for example via imaging techniques, neuroscientists need to correlate brain data with clearly delineated psychological phenomena. In trying to establish, for example, which brain differences (if any) correlate with deficits in executive function, we need a clear psychological test for normal executive function. One can then look at brain function during tasks that test executive function. For example, in the Wisconsin Card Sorting Test, participants sort cards according to rules which change throughout the task and the speed with which they adjust to this is measured. Such tasks, along with neuroimaging, can establish brain differences between people who exhibit problems in executive function and those who do not. The individuation of relevant brain areas and functions is done by averaging over groups of subjects. This means that, as Morse (2017) points out, we rely on clearly identifiable mental phenomena and processes to find out whether there are related brain differences. Thus, we already need to be able to identify a psychological deficit in order to find its brain correlate (cf. Levy 2007a, p.149).

When establishing connections between data from neuroscience and psychological processes, neuroscientists need to navigate some well-known methodological pitfalls. Take for instance the process of reverse inference, where scientists attempt to draw conclusions regarding likely psychological processes from brain data, by appealing to known correlations between activation of a certain brain area and a certain psychological process. For example, we might infer that an individual finds a certain activity rewarding because a brain region of hers that has been correlated with reward in a previous study is active. However, this inference is only strongly supported if that area is *only* implicated in reward processing and in no other psychological functions (Poldrack 2006, 2011). Neuroscientists are well aware of this problem and take this into account when assessing the reliability of the inferences they draw from brain data. A further methodological issue is that we cannot straightforwardly draw inferences about individuals on the basis of group findings, because the effect of a certain brain finding averages over the group, which may be quite heterogenous internally. Therefore, a person may be an outlier in terms of brain function while still performing normally on a psychological measure.⁴

So at least for now, knowledge of brain dysfunctions will only yield corroborative evidence for the existence of specific mental dysfunctions. Such corroborative evidence will be especially useful in cases where behavioral evidence conflicts or we are not sure whether an individual is faking a mental illness.⁵ While evidence of brain difference or pathology is useful for aiding our understanding of mental health conditions and the way they affect reasoning and decision-making, it is only one piece in a much larger puzzle.

⁴ For a detailed discussion of problems in using brain imaging data as evidence in the legal context, see (Sinnott-Armstrong et al. 2008)

⁵ I thank Katrina Sifferd for pointing this out.

2.1 Assessing the evidential role – case studies

Where mental dysfunction mitigates responsibility and is associated with brain dysfunction, the brain dysfunction will be *a cause* of reduced responsibility because it realizes a psychological dysfunction. But even in cases where certain psychological dysfunctions are caused by, or are expressive of, brain difference, we *assess* individuals' moral responsibility at the psychological level. So, on an evidential level, the contribution of brain pathology is to corroborate the existence of psychological deficits or indicate that a certain individual is a likely candidate for a certain psychological dysfunction. This evidential role can, however, in practice be very important, as I will show below.

In the introduction, we encountered the example of Billy, the young criminal with the slow growing brain tumor. In the thought experiment, Levy stipulated that the tumor was the cause of Billy's aggressive and immoral behavior. In light of the above discussion about the evidential role of brain dysfunction, we might speculate that when an individual has a tumor of the location and size that Billy has, we can infer that their capacity for impulse control is compromised to such an extent that we could not expect the individual to be able to control their desires and behave morally. This would be an instance where brain pathology both causes and provides evidence for psychological dysfunction.

However, in real life cases, it is not straightforward to move from the knowledge that brain pathology caused certain psychological deficits to the claim that a person was not responsible, even if we know that certain types of brain pathology are normally associated with specific psychological problems. For example, we know that frontotemporal dementia is frequently associated with disinhibition (Zamboni et al. 2008). But moving from the knowledge that an individual has a brain disorder associated with specific psychological problems to the claim that they are therefore not fully responsible requires us to know the extent of the psychological deficits a person has *at the time* for which we are considering their responsibility.

Consider two real-life cases which illustrate this problem. First, Burns and Swerdlow (2003) describe the case of a man, who, following Morse (2011b), I will refer to as Mr. Oft (presumably short for 'orbito-frontal tumor'):

At age 40, Mr. Oft develops an interest in child pornography, starts secretly collecting it and making sexual advances to his pre-pubescent stepdaughter. After this is revealed to his wife, the child's mother, the man is removed from the home and takes part in an inpatient rehabilitation program as a condition for not being imprisoned for his actions. However, he does not successfully complete the rehabilitation program: "Despite his strong desire to avoid prison, he could not restrain himself from soliciting sexual favors from staff and other clients at the rehabilitation center and was expelled" (Burns and Swerdlow 2003, 437). Just before going to prison, Mr. Oft is admitted to hospital with a bad headache and undergoes neurological

examination because of reported balance problems. During his examination, he solicits favors from female medical staff and is unconcerned by the fact that he urinates on himself. He also shows impairment in motor tasks, such as writing and drawing a clock face. An MRI scan reveals a large right orbitofrontal tumor.

After the tumor is removed, both Mr. Oft's motor control and his behavior return to normal, and he returns to his family. Nine months later, the tumor regrows and headaches and interest in pornography recur. The tumor is once again operated upon. Burns and Swerdlow claim that because orbito-frontal lesions lead to a loss of impulse control, the patient "could not refrain from acting on his pedophilia despite the awareness that this behavior was inappropriate" (Burns and Swerdlow 2003), implying that the man was not responsible for his actions.

A further, less extreme, example which does not involve criminal behavior is that of brain scientist Barbara Lipska, who suffered from numerous brain tumors which were successfully treated. She subsequently describes the changes to her personality that these caused:

I didn't suddenly become someone else. Rather, some of my normal traits and behaviors became exaggerated and distorted, as if I were turning into a caricature of myself. (...) I had no time for anything – not even for the things that I really enjoyed, like talking to my children and my sister on the telephone. I would cut them off midsentence, running somewhere to do something of great importance, though what exactly, I couldn't say. I became rude, and snapped at anyone who threatened to distract me. (...) Strangely, I wasn't worried. Like so many patients with mental illness, whose brains I had studied for a lifetime, I was losing my grasp on reality. (Lipska 2016)

As in the case of Mr. Oft, Lipska's behavior changed back to normal once the tumors were successfully treated. It is clear that the tumors were causally involved in the behavioral changes both individuals experienced; it is also clear that the tumors corroborate the existence of psychological dysfunction that manifests itself in problematic behavior. It is less clear that knowing about the tumors helps us to decide at what stage, if any, the impairments were so substantial that moral responsibility was lost or reduced.

The case of Mr. Oft illustrates both the limits and the importance of brain data for assessing moral responsibility. While the brain tumor affected his control, it is possible that he was still fully responsible when he molested his stepdaughter, because control had not yet deteriorated significantly. Looking at a brain scan will not enable us to tell what level of impulse control Mr. Oft had, even when it is clear that his brain abnormality was causally responsible for his behavioral changes.

However, knowing there is a brain tumor which affects areas associated with executive function is still relevant. As Sifferd (2013) points out, absent information about brain pathology, Mr. Oft's behavior is far more difficult to place as showing problems in impulse control, because behavior such as seeking out sexual favors and collecting pornography, and even molesting one's stepdaughter, may occur in the absence of impulse control problems. It is important to bear in mind that despite quite erratic behavior,

Mr. Oft would have been imprisoned, had he not been admitted to hospital because of his headache and received neurological testing because of his balance problems. In other words, a tumor that did not cause the same kind of physical symptoms—headaches, balance problems—would likely have been discovered later and physicians would not have explored the possibility of psychological dysfunction in the same way. This means that, practically speaking, discovering brain tumors is sometimes of immense relevance to assessing the patient’s psychological capacity correctly.⁶

To summarize, knowledge of Mr. Oft’s tumor is crucial in establishing the causal origin of problems with impulse control and supports the psychological diagnosis, in some cases providing the trigger to consider the possibility of psychological dysfunction. Sometimes, brain pathology provides evidence that an individual is exempted or that their moral responsibility is reduced, because it causes relevant psychological deficits. Furthermore, knowing of this brain pathology will help us place troubling behavior in context and confirm the kind of dysfunction at issue. But important questions will remain unanswered. Knowing that Mr. Oft suffers from a brain tumor will not tell us at what stage he can no longer be said to be responsible, so it won’t help us retrospectively answer the question whether Mr. Oft was responsible when he made advances on his step-daughter.

3. The causal path and the relevance of changes in psychology

3.1 Lack of control over acquiring immoral dispositions and constitutive moral luck

We have seen that the existence of brain pathology can provide important information relevant to responsibility in cases where there is a known link between brain pathology and psychological functioning. However, there is a common intuition that causal history also matters to responsibility, and that the causal history of people with brain pathology can (partially) excuse. In his example of Billy, the young criminal with a brain tumor, Levy argues that the reason we should not blame Billy is because the causal path of how people acquire immoral dispositions matters to responsibility: “The agent’s causal history matters crucially to our assessment of his responsibility” (Levy 2007, p. 134). The underlying idea seems to be that we are not responsible for moral deficits acquired through physical illness. On Levy’s proposal, it is not just synchronic mental dysfunction that matters, but also the causal path that led to this dysfunction.

The idea that causal histories matter is a recurring theme in the philosophy of moral responsibility. In the context of brain pathology, the question that arises is whether the fact that an individual had no control over acquiring morally problematic characteristics excuses (partially or wholly). I will argue against the claim that causal histories mitigate responsibility when people are not in control of their personality changes. However, there is an important way in which personality changes do matter to moral

⁶ Thanks to Jan-Hendrik Heinrichs for pressing this point.

responsibility, because they undermine individuals' established ecological mechanisms of gaining moral knowledge and controlling their behavior, as I show in the next section.

If causal history is to matter to moral responsibility, then we should not focus on cases where a person suffering from brain pathology clearly lacks impulse control and an understanding of their action. In those cases, agents fail to meet criteria for full responsibility at the time of action and an excuse to causal history becomes unnecessary. While the causal history might contribute to the explanation of current deficits, the existence of current deficits is sufficient to excuse or mitigate responsibility.⁷

If causal history matters, it needs to add something. So the interesting cases are those where, looking at the individual's present psychological capacities, we would not judge them incompetent, but the causal history involving a brain pathology provides reasons to excuse them. For example, Mr. Oft may or may not have had pedophilic urges from the outset, but it appears that they might have become stronger or that his impulse control got worse. It is likely that there will have been stages of the brain disease where moral behavior became harder, but not prohibitively hard for the affected person. They would at this later time, call it t_2 , have the capacities we normally take to be sufficient for moral responsibility, even though they would not have the same capacities that they had before they became ill, at t_1 . In other words, the individual at t_1 and t_2 would both be considered responsible if you only look at the relevant time slice. Even though the one at t_2 does not have equally well-developed moral reasons-responsiveness as the one at t_1 (they might be more compulsive, less able to concentrate or more prone to anger), they both pass a threshold of responsiveness necessary for moral responsibility. (We will elegantly pass over the question where exactly that threshold lies.) At t_2 , the 'lack of psychological capacities for impulse control' excuse would not have applied if we just looked at the person's capacities at t_2 , and the person should be counted as responsible by synchronic criteria. However, they might count as having a (partial) excuse because of the way they came to the capacities they have at t_2 .

Walter Sinnott-Armstrong explicitly endorses the claim that "sometimes, the tumor adds force to the excuse by raising the threshold of control required for responsibility" (Sinnott-Armstrong 2012, 203). So even people whose capacities would normally count as good enough for being responsible, may no longer count as responsible. He points to the fact that the idea that causation matters appears in the Model Penal Code as well, which specifies that mental incapacity must be caused by a mental defect or disease. On this reasoning, if the cause of the reduced capacity is somehow external to the agent and beyond their control, this changes the threshold for responsibility. John Martin Fischer and Mark Ravizza, too, make the general

⁷ There are exceptions to the rule that current incapacity excuses. These are discussed in the debate about "tracing", which concerns the question how to best to deal with cases where cause harm while temporarily lacking self-control (for example, because they are drunk or high on drugs), but are responsible for voluntarily getting themselves into the current incapacitated state in the first place (see (King 2014) for discussion). These cases are the converse of the brain disorder cases I discuss below - rather than the causal path to current capacity being thought to reduce responsibility as in the brain tumor cases, in the tracing cases it increases it beyond the responsibility we would attribute on synchronous criteria alone.

claim that responsibility is a historical notion and that it is not just the current properties of an agent, but the way these were acquired, that decide whether the agent is morally responsible (Fischer and Ravizza 1998, 187)—an agent may look responsible by synchronic standards, but not be responsible because of the way they acquired their capacities. So the proposal to be considered is that historical facts about the way individuals came to have their current level of reasons responsiveness also matter to whether they are fully responsible.

Should we accept the claim that an agent who is responsible by synchronic criteria at *t* is not in fact fully responsible because they are not responsible for acquiring the psychological features that make them behave badly, because these resulted from brain pathology? Cases where agents are responsible by synchronic standards but have not acquired their current values and desires in the normal way have been considered in the context of scenarios in which people are manipulated by a nefarious person to achieve his own goals. Some authors deny that such manipulated individuals are responsible, even if they meet synchronic or current time-slice criteria for responsibility (Fischer and Ravizza 1998). Manipulation scenarios and cases of brain tumors or neuro-degenerative diseases share the feature that the agent's psychology is changed through brute processes which are external to their agency and non-transparent to them.⁸

One way of denying history sensitivity for these cases is to point out that everybody's moral character is shaped by many factors that are outside their control. We are all subject to constitutive moral luck in that, to a large extent, we do not control the environment and the genetic material that shape our personality. Our influence on the traits which lead us to behave in certain ways is extremely limited. Some people develop a bad character and are deficient in their ability to understand and respond to moral reasons because of their horrible parents, or because they never had a chance to learn certain values. Brain tumors which change our personality for the worse are just a more extreme case of this problem.⁹

But maybe there is something special about brain pathology that goes beyond the way luck normally affects our character and dispositions? Maybe some histories are special. The most obvious justification for such a claim would be that in the case of brain pathology, a person who used to have an intact moral character develops a problematic one because of a physical disease, without any voluntary contribution or moral slippage through embracing bad habits.¹⁰ However, falling into bad habits is likely caused by a combination of having certain predispositions which are determined by nature and nurture, and the circumstances one happens to be in. Just as one can be unlucky in being placed in a morally challenging

⁸ Manipulation cases also have the added feature that there is another agent responsible for the psychological changes in the subject, which may well contribute further to changes in our responsibility intuitions.

⁹ One might of course be tempted to draw the opposite conclusion and take this kind of case as a reason to reject responsibility altogether as so many relevant factors are ultimately out of our control. I take this to be an unattractive option and will not pursue it further.

¹⁰ In fact, personality change plays a major role in some arguments for the mitigated responsibility of individuals with brain tumors, cf. Reznick 1997 and Sinnott-Armstrong 2012.

environment, such as a totalitarian state, one can be unlucky in acquiring deviant impulses through a disorder. Similarly, the fact that someone is a pedophile and sexually attracted to children does not mean that they are not responsible for acting on this attraction, even though they may have no control over feeling this attraction in the first place, and they are certainly unlucky to have that disposition. If they have acquired normal self-control and reflection abilities, we can expect them to use these. The same applies to people who are naturally stingy, short-tempered, etc. So an appeal to the fact that agents had no control over acquiring certain dispositions will not help to distinguish between constitutive moral luck that we should accept and luck that undermines responsibility. In fact, the perceived impossibility of drawing a line between acceptable and unacceptable forms of moral luck is a reason why some authors argue that history sensitive compatibilism is not a tenable position (Levy 2009, Arpaly 2002).

3.2 The effect of psychological change on morally responsible agency

Lack of control over personality changes is not a good candidate mitigating condition. However, there is an important related phenomenon that does affect the psychological capacities necessary for moral responsibility and thereby provides some excuse. I believe that Sinnott-Armstrong is right to say that we are missing something important if we only look at psychological features such as self-control or reasoning. But what supplies an excuse in these cases is something different that hasn't been appreciated in the literature: what matters is how the *change* in capacities influences synchronic capacities, some of which rely on support from the social environment. Let me explain.

Individuals suffering from brain pathology started out with normally developed moral capacities, and these change because the illness affects their interests and concerns, impulse control, and moral judgment, as well as their awareness of these changes. In her memoir on living with a brain tumor, Lipska (2018) vividly describes her lack of insight into the changes in her personality, as well as the fact that her family felt overwhelmed and did not know how to deal with the new behavior. She became irritable, impulsive and demanding but did not notice the way these changes were distressing her family. Her family, in turn, was not able to call out the challenging behavior because they were unable to cope with the changes in someone who used to be considerate. It is plausible that over and above the impairment of impulse control, Lipska's responsible agency was affected by the fact that she was a) unaware of the changes in her behavior and attitudes, b) hadn't developed mechanisms to cope with the changes and that c) her family was too shocked and concerned by the changes to hold her accountable. Both she and her environment were insufficiently equipped to deal with the changes in her psychology due to her brain disease, and this further affects responsibility.

These problems can be illustrated more clearly with some further examples. As adults with developed moral capacities, we know our moral weaknesses and try to counterbalance them. Let us assume that Clara knows that she has a slightly flighty nature and will be unfaithful to her partner if separated from him for extended periods of time. In order to avoid hurting him and her relationship, she therefore avoids

long periods of separation. If, due to a brain tumor or a neurodegenerative condition, she all of a sudden finds her sexual urges much stronger, the measures she has put in place to remain faithful may well be insufficient. Furthermore, if she does not realize that there has been a drastic change in her psychology, she will not see the need to take steps to adjust to her stronger urges. We could speculate that something similar might have occurred in the case of Mr. Oft: he had acquired ways of dealing with unwelcome impulses of a certain strength through distraction, as part of his moral education. When the impulses became stronger due to the tumor, he was less able to control his behavior, because he hadn't learned to cope with impulses of that strength. In these ways, certain brain disorders can mitigate responsibility because such pathologies can present one with new challenges that one has not developed strategies for. Arguably, one of the reasons we hold children and teenagers to a different standard of responsibility is not just that they have less raw impulse control, they also have had less time to get to know themselves and develop ways to avoid temptation. They have not yet been taught to count to ten when they are angry, or to check their knee jerk reaction with a sympathetic interlocutor to see whether they are overreacting.

As can be seen especially vividly in the example of young people, our moral responsibility is socially supported or 'scaffolded'. McGeer and Pettit (2015) argue that other people's actual and imagined reactions to our moral and immoral actions play a role both in motivating us to do the right thing and in establishing, in a collaborative effort, what the right thing to do is. The extent to which our social environment contributes to our moral agency, in supporting both our sensitivity to moral reasons and our motivation to act on them, is increasingly recognized in the literature (Washington and Kelly 2016, Holroyd 2018). McGeer and Pettit make the plausible claim that part of what motivates us to behave morally is that we want to be able to justify our behavior to others. The fact that others call us out when we do wrong is one of the factors that keeps us on the straight and narrow. Our moral capacities are developed and scaffolded through the way we are embedded into a community that reacts to our moral and immoral behavior through praise, blame, punishment, etc. If an adult loses developed moral capacities, it will be harder for others to calibrate their responses to that person, which in turn makes it even more difficult for that person to respond to moral reasons. When people are baffled by our changed conduct, they sometimes do not call out inappropriate behavior, because they are at a loss how to respond to out-of-character behavior. This effect is likely to be aggravated if we *know* that the person suffers from a serious illness and the new behavior is caused by that illness.

Furthermore, just as we can make it easier to behave well by avoiding certain situations where we are tempted to behave badly in normal circumstances, we can enlist the help of others more directly in behaving well. We can get others to remind us to do things we know we should do, tell us not to send that e-mail to our obnoxious relative before having a night to cool down, etc. But, for this to work, the people who take on these roles in our social environment need to know what our weak spots are and *how* weak these are. So when our personality changes due to illness, these important relational factors supporting our responsible agency are endangered as well.

We need not adopt a history-sensitive view of moral responsibility in order to accept the claim that the psychological changes in some brain disorders provide a mitigating factor. What makes responsibility harder for those individuals who suffer from progressive brain pathology is the fact that their self-control, self-monitoring, and social support networks are not equipped to deal with the personality changes. The correct description would not be that somebody who was responsible by synchronic criteria while ill is not responsible because of the way they became ill. Rather, the problem arises from an overly narrow focus purely on intrinsic measures such as the strength of certain impulses or self-control (executive functions), even synchronically speaking. We need to also look at habits and mechanisms (both internal and external in terms of social scaffolding) that a person has developed in order to deal with familiar challenges and at the way these are supported by their social environment.

Washington and Kelly make a related point when discussing moral responsibility for implicit bias. In becoming aware of and controlling for implicit bias, we are heavily dependent on expert knowledge in our environment and on indirect measures to control it. They draw an analogy to blood pressure: “Moreover, no one can directly control her own blood pressure, or bring about an immediate and sustained change in it by direct act of will. To effectively control our blood pressure, most of us need to learn about and use the more roundabout, external methods that have been empirically verified” (Washington and Kelly 2016, 27-28). Just as we need to put external checks and indirect measures in place to monitor and control our blood pressure, we need to put indirect measures in place to assure self-control and sensitivity to moral reasons for action.

Let me give another example. Some people are prone to losing their temper easily when hungry. If they know this and their partner knows this, they will make sure that snacks are in supply at the right time. But if a newly acquired health condition means that this threshold for hunger-induced irritability has moved, your partner will not know when to bring out the cookies to keep you on an even keel and prevent intemperate bursts of rage. It is the fact that this system of support, accountability and control extending over more than just the individual has been disrupted that can provide further excuse or mitigation in the case of progressive brain diseases.

Importantly, this kind of mitigating or excusing condition does not only apply to paradigmatic brain diseases such as tumors. Similar considerations apply in other mental disorders, for example schizophrenia or psychosis. If individuals who suffer from these conditions start having unusual experiences and they and their social environment have not yet developed ways of categorizing and coping with them, this reduces responsibility. In disorders such as schizophrenia, social support of responsible agency may be further hindered by the fact that friends and family do not understand the nature and scope of the condition, and therefore find it hard to know how to tailor their ways of holding the affected individual responsible. While brain tumors provide a particularly vivid example and the progression is

different, similar considerations can apply when individuals develop mental disorders like schizophrenia or bipolar disorder.

4. Conclusion

I have argued that the primary role brain dysfunction or pathology should play in our practices of blame is to provide further evidence and explanation for responsibility-relevant psychological dysfunction. Translating brain data into relevant psychological dysfunction is not always straightforward. It is particularly difficult when a patient has a progressive brain disease and we are trying to establish the point where psychological functioning is sufficiently impaired for the patient's responsibility to be diminished or lost. I have considered whether the causal path by which people come to behave badly when they suffer from brain disorders matters to their responsibility. Specifically, I have considered whether the bar for excuse is lowered by the fact that the moral deficits are a result of brain pathology. I conclude that it is not the causal path that diminishes responsibility, but the way in which psychological changes undermine the normal mechanisms of control and moral feedback.¹¹

References

- Arpaly, Nomy. 2002. *Unprincipled Virtue: An Inquiry Into Moral Agency*: Oxford University Press.
- Arpaly, Nomy. 2005. "How it is not 'Just like Diabetes'. *Mental Disorders and the Moral Psychologist*." *Philosophical Issues* 15:282-298.
- Baler, R. D., and N. D. Volkow. 2006. "Drug addiction: the neurobiology of disrupted self-control." *Trends Mol Med* 12 (12):559-66. doi: 10.1016/j.molmed.2006.10.005.
- Brink, David O., and Dana K. Nelkin. 2013. "Fairness and the Architecture of Responsibility." *Oxford Studies in Agency and Responsibility* 1:284-313.
- Broome, Matthew R., Lisa Bortolotti, and Matteo Mameli. 2010. "Moral Responsibility and Mental Illness: A Case Study." *Cambridge Quarterly of Healthcare Ethics* 19 (2):179-187. doi: 10.1017/S0963180109990442.
- Burns, Jeffrey M., and Russell H. Swerdlow. 2003. "Right Orbitofrontal Tumor With Pedophilia Symptom and Constructional Apraxia Sign." *Archives of Neurology* 60 (3):437-440. doi: 10.1001/archneur.60.3.437.
- Catley, Paul, and Lisa Claydon. 2015. "The use of neuroscientific evidence in the courtroom by those accused of criminal offenses in England and Wales." *Journal of law and the biosciences* 2 (3):510-549. doi: 10.1093/jlb/lsv025.
- Cuthbert, Bruce N. 2014. "The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology." *World Psychiatry* 13 (1):28-35. doi: 10.1002/wps.20087.
- Farahany, Nita A. 2016. "Neuroscience and behavioral genetics in US criminal law: an empirical analysis." *Journal of law and the biosciences* 2 (3):485-509. doi: 10.1093/jlb/lsv059.

¹¹ I would like to acknowledge support by the Leverhulme Trust through an Early Career Fellowship entitled 'Mental Disorders, Brain Disorders and Moral Responsibility'. This paper was presented at the work in progress seminar at the University of Birmingham (UK) and at the workshop 'Responsibility and Mental Disorder' at the University of Alabama at Birmingham; and I received insightful comments from both audiences. I thank Josh May, Matt King, Katrina Sifferd and Jan-Hendrik Heinrichs for helpful feedback on earlier drafts of this chapter.

- Fine, Cordelia, and Jeanette Kennett. 2004. "Mental Impairment, Moral Understanding and Criminal Responsibility: Psychopathy and the Purposes of Punishment." *International Journal of Law and Psychiatry* 27:425-443.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Vol. 61: Cambridge University Press.
- Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1):5-20. doi: 10.2307/2024717.
- Glannon, Walter. 2008. "Moral responsibility and the psychopath." *Neuroethics* 1 (3).
- Graham, George. 2013. *The Disordered Mind: An Introduction to Philosophy of Mind and Mental Illness*: Routledge.
- Holroyd, Jules. 2018. "Two Ways of Socializing Moral Responsibility: Oppression, Politics and Moral Ecology." In *Social Dimensions of Moral Responsibility*, edited by Katrina Hutchison, Catriona Mackenzie and Marina Oshana, 137-162. New York: Oxford University Press.
- Insel, Thomas. 2013. "Transforming Diagnosis." NIMH, accessed 07.06.2017. <https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml>.
- Jefferson, Anneli. 2020. "What does it take to be a brain disorder?" *Synthese* 197 (1):249-262. doi: 10.1007/s11229-018-1784-x.
- Jefferson, Anneli, and Katrina Sifferd. 2018. "Are Psychopaths Legally Insane?" *European Journal of Analytic Philosophy* 14 (1):79-96.
- King, Matt. 2014. "Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility." *European Journal of Philosophy* 22 (3):463-482. doi: doi:10.1111/j.1468-0378.2011.00502.x.
- King, Matt, and Joshua May. 2018. "Moral Responsibility and Mental Illness: a Call for Nuance." *Neuroethics* 11 (1):11-22. doi: 10.1007/s12152-017-9345-4.
- Levy, Neil. 2007a. *Neuroethics, Challenges for the 21st century*. Cambridge: Cambridge University Press.
- Levy, Neil. 2007b. "The responsibility of the psychopath revisited." *Philosophy, Psychiatry, and Psychology* 14 (2):pp. 129-138.
- Levy, Neil. 2009. "Luck and History-Sensitive Compatibilism." *The Philosophical Quarterly* 59 (235):237-251. doi: 10.1111/j.1467-9213.2008.568.x.
- Lipska, Barbara. 2016. "The Neuroscientist who lost her mind." *The New York Times*. Accessed 14.2.2019. <https://www.nytimes.com/2016/03/13/opinion/sunday/the-neuroscientist-who-lost-her-mind.html>.
- Lipska, Barbara. 2018. *The Neuroscientist Who Lost Her Mind: A Memoir of Madness and Recovery*: Bantam Press.
- McGeer, Victoria, and Philip Pettit. 2015. "The Hard Problem of Responsibility." In *Oxford Studies in Agency and Responsibility*. Oxford: Oxford University Press.
- Morse, Stephen. 2011a. "Avoiding Irrational NeuroLaw Exuberance: A plea for Neuromodesty." *Mercer Law Review* 62:837-859.
- Morse, Stephen. 2011b. "Lost in Translation? An Essay on Law and Neuroscience." In *Law and Neuroscience: Current Legal Issues Volume 13*, edited by Michael Freeman, 529-562. Oxford University Press.
- Morse, Stephen J. 2017. Neuroethics:Neurolaw. In *Oxford Handbooks Online*.
- Papineau, David. 1994. "Mental Disorder, Illness and Biological Disfunction." *Royal Institute of Philosophy Supplements* 37:73-82. doi: doi:10.1017/S135824610000998X.
- Pickard, Hanna. 2015. "Psychopathology and the Ability to Do Otherwise." *Philosophy and Phenomenological Research* 90 (1):135-163. doi: 10.1111/phpr.12025.
- Poldrack, Russell A. 2006. "Can cognitive processes be inferred from neuroimaging data?" *Trends in Cognitive Sciences* 10 (2):59-63. doi: <https://doi.org/10.1016/j.tics.2005.12.004>.

- Poldrack, Russell A. 2011. "Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding." *Neuron* 72 (5):692-697. doi: 10.1016/j.neuron.2011.11.001.
- Radden, Jennifer. 2018. "Rethinking disease in psychiatry: Disease models and the medical imaginary." *Journal of Evaluation in Clinical Practice* 24 (5):1087-1092. doi: 10.1111/jep.12982.
- Reznek, Lawrie. 1997. *Evil or Ill? Justifying the Insanity Defence*. London: Routledge.
- Schramme, Thomas. 2013. "On the autonomy of the concept of disease in psychiatry." *Frontiers in Psychology* 4 (457). doi: 10.3389/fpsyg.2013.00457.
- Shoemaker, David W. 2011. "Psychopathy, Responsibility, and the moral/conventional distinction." *The Southern Journal of Philosophy* 49:99-124. doi: 10.1111/j.2041-6962.2011.00060.x.
- Sifferd, Katrina, and William Hirstein. 2013. "On the Criminal Culpability of Successful and Unsuccessful Psychopaths." *Neuroethics* 6 (1):129-140.
- Sifferd, Katrina L. 2013. "Translating Scientific Evidence into the Language of the 'Folk': Executive Function as Capacity-Responsibility." In *Legal Responsibility and Neuroscience*, edited by Nicole A. Vincent. Oxford University Press.
- Sinnott-Armstrong, Walter. 2012. "A CASE STUDY IN NEUROSCIENCE AND RESPONSIBILITY." *Nomos* 52:194-211.
- Sinnott-Armstrong, Walter, Adina Roskies, Teneille Brown, and Emily Murphy. 2008. "Brain Images as Legal Evidence." *Episteme* 5 (3):359-373. doi: 10.3366/E1742360008000452.
- Vincent, Nicole A. 2008. "Responsibility, Dysfunction and Capacity." *Neuroethics* 1 (3):199-204. doi: 10.1007/s12152-008-9022-8.
- Washington, Natalia, and Daniel Kelly. 2016. "Who's Responsible for This?: Moral Responsibility, Externalism, and Knowledge about Implicit Bias." In *Implicit Bias and Philosophy, Volume 2*. Oxford: Oxford University Press.
- Zamboni, G., E. D. Huey, F. Krueger, P. F. Nichelli, and J. Grafman. 2008. "Apathy and disinhibition in frontotemporal dementia: Insights into their neural correlates." *Neurology* 71 (10):736-742. doi: 10.1212/01.wnl.0000324920.96835.95.