Journal of Cognition and Neuroethics

Is Neuroscience Relevant to Our Moral Responsibility Practices?

Joseph M. Vukov
Fordham University

Biography
Joe Vukov is a PhD candidate in the philosophy department at Fordham University. His research interests lie primarily in philosophy of mind, metaphysics, and bioethics.

Publication Details
Journal of Cognition and Neuroethics (ISSN: 2166-5087). November, 2014. Volume 2, Issue 2.

Citation
Vukov, Joseph M. 2014. "Is Neuroscience Relevant to Our Moral Responsibility Practices?" *Journal of Cognition and Neuroethics* 2 (2): 61–82.

Is Neuroscience Relevant to Our Moral Responsibility Practices?

Joseph M. Vukov

Abstract

Some psychologists and philosophers have argued that neuroscience is importantly relevant to our moral responsibility practices, especially to our practices of praise and blame. For consider: on an unprecedented scale, contemporary neuroscience presents us with a mechanistic account of human action. Furthermore, influential studies – most notoriously, Libet et al. (1983) – seem to show that the brain decides to do things (so to speak) before we consciously make a decision. In light of these findings, then – or so some have argued – we ought to revise our practices of praise and blame. In the current paper, I argue that this conclusion is unwarranted. The reason is that the argument for it depends on controversial non-empirical premises, premises we need not accept. I suggest, however, that neuroscience does bear on our moral responsibility practices in one important, if less revisionary, way. In particular, neuroscience offers a new kind of evidence for determining when agents should be held exempt from our normal moral responsibility practices.

Keywords

Moral responsibility; neuroscience; incompatibilism; Libet; conscious control

Is Neuroscience Relevant to Our Moral Responsibility Practices?

Some psychologists and philosophers have argued that neuroscience is importantly relevant to our moral responsibility practices, especially to our practices of praise and blame. Specifically, they have argued for the following theses:

i. On an unprecedented scale, contemporary neuroscience presents us with a mechanistic account of human action. But if our actions can be accounted for mechanistically, this gives us reason to revise the ways in which we hold agents morally responsible.

ii. Influential studies – most notoriously Libet et al. (1983) – show that the brain decides to do things (so to speak) before we consciously make a decision. But since conscious control is necessary for holding agents morally accountable, these studies suggest the need for revising our moral responsibility practices.

iii. Still other studies have discovered neural correlates for specific psychological abnormalities. Neuroscience thus offers a genuinely new kind of evidence for

deciding when particular individuals meet criteria for exemption from our moral responsibility practices.

The worries presented in (i-iii) therefore make it seem that neuroscience is importantly relevant to our moral responsibility practices. I will argue, however, that neuroscience is in fact relevant these practices only in the way spelled out in (iii). More specifically: I will argue (in Section 1) that (i) is an important consideration only if we assume a number of controversial metaphysical premises. And since metaphysical premises cannot be supported empirically, neuroscience cannot by itself raise the worry presented by (i). The worry presented by (ii), in turn, is a concern only if we take a lack of conscious control to undermine our moral responsibility practices. But, or so I will argue (in Section 2), a lack of conscious control doesn't undermine our moral responsibility practices. This leaves (iii), then, as the only way neuroscience bears on our moral responsibility practices (Section 3). And while (iii) is not as revisionary as either (i) or (ii), recognizing this specific role neuroscience can play still has important implications.

1. Neuro-Revisionism

1.1 The Argument for Neuro-Revisionism

A number of authors have argued that the mechanistic picture of human action emerging from contemporary neuroscience should lead us to revise our moral responsibility practices (Farah & Heberlein 2006; Wright 1994; Greene & Cohen 2004). Let's call anyone who argues for this conclusion a *Neuro-Revisionist*. Why should we accept Neuro-Revisionism? Greene & Cohen (2004) articulate a standard version of the argument – it can be reconstructed as follows:

P1. Our actions can be accounted for mechanistically.

P2. If our actions can be accounted for mechanistically, then there is no free will.

C1. Therefore, there is no free will.

P3. If there is no free will, our moral responsibility practices should be revised.

C2. Therefore, our moral responsibility practices should be revised.

Let's consider why we might be tempted to accept each premise:

P1, first of all, claims that our actions can be accounted for *mechanistically*. I'll follow Bechtel and Richardson's (1993) definition of a 'mechanistic explanation' as any explanation that "propose[s] to account for the behavior of a system in terms of the functions performed by its parts and the interactions between these parts" (Bechtel and Richardson 1993, 17). And indeed, new work in neuroscience seems to give just such an explanation of many of our actions. Consider, for instance, the story we are now able to tell about the simple action of reading aloud a written word:

> Light reflects off the page and strikes the retina. Different retinal cells react to various wavelengths of light, which in turn leads to different patterns of firing in the ganglion cells. These cells then innervate the lateral geniculate nucleus of the hypothalamus, which in turn sends a tract to the primary visual cortex, V1. V1 then innervates the angular gyrus, which decodes the visual information and transmits this information to Broca's area via the arcuate fasciculus. Broca's area, in turn, creates a motor plan for speaking the word and transmits this plan to the primary motor cortex, where the motor plan is implemented. (Breedlove et al. 2010)

This story, it seems, thus accounts for the action exclusively in terms of neurological function; it accounts for the action *mechanistically*. And while were are not yet able to tell a story to this degree of detail about complex human actions – actions such as falling in love, composing a poem, or playing a game of chess – advances in neuroscience suggest that such a description is possible, if not inevitable.

Turn next to P2. Greene & Cohen defend this premise by appeal to an intuitive understanding of free will (which is, as we will see, problematic – but more on this later). Specifically, they argue that "people regard actions only as fully free when those actions are seen as robust against determination by external forces" (Greene & Cohen 2004, 1780). My actions are intuitively free, in other words, only when they have their source in *me and me alone*. On Greene & Cohen's account, then, actions performed with intuitive free will are importantly different from actions that can be accounted for mechanistically. When we account for an action mechanistically, after all, we account for it not in terms of an agent's volition, but rather in terms of the operations of sub-personal mechanisms. If my actions can be accounted for mechanistically, then, they could not have their source in *me* – they could not be intuitively free. Equivalently then, if our actions can be accounted for mechanistically, there are not performed freely.

Last, consider P3. Greene & Cohen defend this premise by observing that our current moral responsibility practices are largely founded on a presumption of free action. Consider, for instance, our current practices of punishment – these practices, Greene & Cohen argue, are founded on the assumption that "people…*deserve* to be punished, and that is why we punish them" (Greene & Cohen 2004, 1776). Intuitively, however, a person cannot *deserve* to be punished for an action that was not freely performed.[1] Insofar then as our current practices of punishment are founded on a presumption of free will, it follows that if there is no free will, we must revise these practices.[2]

And from this, Neuro-Revisionism follows – recent discoveries in neuroscience provide us with reason to revise our moral responsibility practices. If we accept P1-P3, we must therefore accept that "for ethics, the only alternative we can see is a shift to a more utilitarian approach [to our moral practices]" (Farah & Heberlein 2007, 46), that "advances in neuroscience are likely to change the way people think about human action and…responsibility" (Greene & Cohen 2004, 1784).

1.2 Response to the Argument for Neuro-Revisionism

Above, I outlined the argument for Neuro-Revisionism. In the current section, I hope to show that the argument's conclusion is unwarranted. The reason is that the argument depends on the truth of a number of controversial metaphysical premises, premises we need not accept.

To clear up some terminology as I will be using it: *determinism* claims that given the state of the world at any given time and the laws of nature, there is only one possible resultant state of the world at any given later time.[3] *Incompatibilism* is the thesis that free will is not compatible with determinism; incompatibilism, therefore, claims that if determinism is true, our actions cannot be free.[4] *Compatibilism*, on the other hand, claims that even if determinism is true and our actions are thus causally determined, it is still

1. For a similar and more thoroughly defended argument for this position, see Pereboom (2013).

2. Greene & Cohen make some obviously controversial claims in the course of defending P3. But as nothing they say here is relevant to my current purposes, I allow their claims to remain unchallenged.

3. This definition may be contended, of course. For my purposes, however, it is sufficient to have a working definition of determinism. And I take it that this definition – while not beyond reproach – is by no means unorthodox either.

4. For some influential incompatibilist accounts, see Chisholm (1964), van Inwagen (1975), and Kane (1996).

possible for these actions to be free.[5] Importantly, then, compatibilists and incompatibilists do not disagree about the truth or falsity of determinism; rather, the positions disagree about the status free will would have were determinism true.

With this terminology in hand, let's turn back to Greene & Cohen's argument for Neuro-Revisionism. Specifically, consider P2 of their argument – this premise claims that 'if our actions can be accounted for mechanistically, then there is no free will.' As we saw, Greene & Cohen defend this premise by appealing to an intuitive understanding of free will. Intuitively free actions, they claim, are importantly different from actions that can be accounted for mechanistically. At first blush, this reasoning may seem sound. In fact, however, it depends on two suppressed metaphysical premises.

The first of these premises forms the foundation for Greene & Cohen's claim that "people regard actions only as fully free when those actions are seen as robust against determination by external forces" (Greene & Cohen 2004, 1780). For this claim, as should now be apparent, is not innocent. Compatibilists, after all, *do* regard free actions as compatible with determinism – compatibilists can therefore argue that actions have their source in *me* even when those actions are determined.[6] In the course of arguing for P2, Greene & Cohen must therefore assume:

> **Incompatibilism**: If our actions are determined, then there is no free will.

Greene & Cohen, to their credit, acknowledge that their argument for P2 assumes **Incompatibilism**. They urge, however, that contemporary neuroscience reveals our *true* intuitions – they argue that the mechanistic picture of human action presented by neuroscience supports:

> ...a powerful moral intuition that…compatibilist philosophies sweep under the rug…[P]eople regard actions only as fully free when those actions are seen as robust against determination by external forces. But if determinism…is true, then no actions are truly fee because forces beyond our control are always sufficient to determine behavior. Thus,

5. For some influential compatibilist accounts, see Ayer (1954), Frankfurt (1971), Watson (1996), Strawson (1963), Wolf (1990), Fischer (1987; 1994), Fischer & Ravizza (1998).

6. Although they need not argue in this way – it is possible to be a compatibilist (and so affirm that our actions are free) and yet deny that we are the source of our actions.

> intuitive free will is [incompatibilist], not compatibilist. (Greene & Cohen 2004, 1780)[7]

According to Greene & Cohen, in other words, when we carefully consider the mechanistic picture painted by neuroscience, we intuitively see that free will finds no place in it.

This argument for **Incompatibilism** won't work, and for a number of reasons. *First*, Greene & Cohen's claim that our raw moral intuitions are decidedly incompatibilist seems to be false on purely empirical grounds, or – at the very least – more empirically complicated than they make it seem. A series of recent studies have found that untutored intuitions are in many cases *compatibilist*. Nahmias et al. (2005), for instance, found that people tend to judge agents morally responsible for actions even when these actions are construed as determined. Similarly, Nichols and Knobe (2007) found that people's intuitions tend to be compatibilist as long as moral scenarios are posed in a concrete, emotional manner. These findings, of course, are not beyond reproach – Nahmias et al. (2007), for instance, found that our intuitions *do* tend towards incompatibilist when scenarios are couched in mechanistic neural terms. This particular study therefore supports Greene & Cohen's hypothesis. Taken as a whole, however, recent empirical work suggests that our raw moral intuitions are much more nuanced that Greene & Cohen assume.

Suppose though that our intuitions *were* incompatibilist in just the way Greene & Cohen assume. Even then, this does not give us good reason to accept **Incompatibilism**. For while consistent incompatibilist intuitions may count as *prima facie* evidence in favor of **Incompatibilism**, intuitions – especially moral ones – are notoriously theory-laden; they often rest on unfounded assumptions, confusions, and equivocations. Most compatibilist arguments, furthermore, consist precisely in diagnosing the faulty reasoning lying at the heart of incompatibilist intuitions. So then, by treating these intuitions as decisive rather than *prima facie* evidence in favor of incompatibilism, Greene & Cohen ignore the standard compatibilist reply. And in ignoring this reply, they fail to offer an argument that any compatibilist would accept.

Now, let me be clear about what I am arguing here: I am not, in the first place, claiming that moral intuitions are irrelevant to our moral reasoning; clearly, they are. And I am also not suggesting that **Incompatibilism** is false; it could very well be that our incompatibilist intuitions are perfectly accurate. Rather, I am pointing out that these intuitions cannot – as Greene & Cohen suppose – be taken as decisive in the debate between compatibilists and incompatibilists; they should rather be an occasion for this

7. This claim is echoed in Kane (1999) and Strawson (1986).

debate. Insofar as the argument Greene & Cohen give for Neuro-Revisionism depends on **Incompatibilism**, it therefore depends on a metaphysical premise they do not adequately defend.

And things get worse. The reason is that **Incompatibilism** does not entail P2 – P2 claims that 'if our actions can be accounted for *mechanistically*, then there is no free will' while **Incompatibilism** claims that 'if our actions are *determined*, then there is no free will.' To move from **Incompatibilism** to P2, Greene & Cohen must therefore assume the following thesis:

> **Mechanism entails Determinism**: If our actions can be accounted for mechanistically, then our actions are determined.

But it is by no means obvious that **Mechanism entails Determinism** is true. The reason is simple: it is not obvious that just because we can account for an action mechanistically that the mechanisms in play are *themselves* deterministic. It is possible, for instance, that our actions could ultimately be explained in terms of the behavior of *indeterministic* mechanisms – and indeed, this is precisely the position taken by event-causal libertarians (Wiggins 1973; Ekstrom 2000; Kane 1996). In taking on **Mechanism entails Determinism**, Greene & Cohen therefore find themselves suppressing yet another controversial and inadequately defended metaphysical premise.

The upshot of this is that Greene & Cohen's argument does not establish Neuro-Revisionism – Neuro-Revisionism follows from their argument only if we accept both **Incompatibilism** and **Mechanism entails Determinism**. But determining the truth of *these* theses is a distinctively philosophical project – these theses, after all, are distinctively metaphysical, and there is no clear way in which neuroscience could be of use answering them. In the end, then, the argument for Neuro-Revisionism is a poor one, and we need not accept its conclusion. If Neuro-Revisionists are to succeed in showing that neuroscience is relevant to our moral responsibility practices, they must first address philosophical questions about free will, determinism, mechanism, and moral responsibility – the very questions philosophers have focused on from the beginning.

2.0 Libet et al. (1983), the Principle of Conscious Control, and Moral Responsibility Skepticism

2.1 The Argument from Libet for Moral Responsibility Skepticism

I argued in Section 1 that the mechanistic picture of action presented by neuroscience is not capable of leading us to revise the way in which we hold agents morally accountable. Or it is not capable of doing so, at any rate, without some hefty philosophical lifting. But there is another way to raise the worry that findings in neuroscience give us reason to revise our moral responsibility practices. Furthermore, this way of raising the worry does not depend on any metaphysically-contentious claims concerning the relationship between mechanistic explanation, determinism, and free will. It rather points to specific findings in neuroscience – most centrally, those of Libet et al. (1983) – and argues that these findings universally undermine moral responsibility. I'll call anyone who argues along these lines a *Moral Responsibility Skeptic*.

Moral Responsibility Skepticism lurks in a great deal of the literature surrounding Libet's findings (Libet 1983; Libet 2011; Banks & Isham 2011; Hallett 2011; Pockett 2004, Roediger et al. 2008; Spence 2009; Wegner 2002). As Bayne (2011) points out, however, full-fledged arguments for the position rarely boil to the surface – Moral Responsibility Skeptics rarely articulate precisely *how* Libet endangers moral responsibility. In this section, then, I will consider a generic version of the argument for Moral Responsibility Skepticism, one that – while generic – seems to lie at the base of the more particular worries raised by Libet and his followers. The argument proceeds as follows:

> P1: For any action, *a*, and any subject, S, S can be held morally responsible for *a* only if *a* is caused by the conscious decision of S.
>
> P2: There is no action, *a,* such that *a* is caused by the conscious decision of a subject.
>
> C1: Therefore, there is no action, *a*, such that there is some subject who can be held morally responsible for *a*.

Let's work through the premises. Call P1 the Principle of Conscious Control [PCC]; according to PCC, if an agent does not cause an action by his or her conscious control, that agent cannot be held morally responsible for the action. Velleman (1992) offers a reason for why we might adopt something like PCC. In the course of cashing out a scenario – one in which he finds himself yelling at a friend without having made a conscious decision to do so – Velleman reasons:

> …viewing the decision [to yell at my friend] as directly motivated by my desires, and my behaviors as directly governed by the [unconscious]

> decision…leads to the thought that as my words become more shrill, it was my resentment speaking, not I. (Velleman 1992, 465)

According to Velleman, in other words, events that result from unconscious decisions cannot be understood as genuine actions at all. This is because, on Velleman's view, insofar as a decision is made unconsciously, it is not genuinely *one's own*. But if an event is not an action at all, it is certainly not the sort of thing for which agent could be held responsible. Hence, if Velleman is right, we seem to have good reason to suppose that conscious control is a necessary condition for holding agents morally responsible.

Velleman's considerations aside, there is an intuitive plausibility to PCC. The paradigmatic actions for which we hold agents morally responsible, after all, are those which seem to be caused by the conscious decision of an agent; conversely, the paradigmatic actions for which we don't hold agents morally responsible are those which do not seem to be caused by the conscious decision of an agent. Consider: we hold agents morally responsible for lying, murdering, theft, donating to charity, and so on; we don't hold agents morally responsible for twitches, schizophrenic episodes, compelled misdeeds, and so on. And it seems that the latter actions differ from the former precisely in that the former are caused by the conscious decision of the agent in a way that the latter are not. More, of course, needs to be said about PCC. But the motivation behind adopting P1 should be clear.

On then to P2. This is the premise Libet et al. (1983) purports to confirm, so it's worth reviewing Libet's findings: in the electroencephalography [EEG] study, participants were instructed to relax and then – whenever they chose – to self-initiate "quick, abrupt flexion of the fingers and/or wrists of [the] right hand" (Libet et al 1983, 625). Participants were also asked to report the moment at which they first felt the desire to perform this movement. The study found that "with few exceptions, onset of [cerebral activity] occurred before reported awareness time by substantial amounts of time" (Libet et al. 1983, 634). Libet et al. (1983) thus concludes that:

> …the brain evidently 'decides' to initiate or, at the least, prepare to initiate the act at a time before there is any reportable subjective awareness that such a decision has taken place. It is concluded that the cerebral initiation of even a spontaneous voluntary act, of the kind studied her, can and usually does begin unconsciously. (640)

Subsequent studies have echoed Libet's findings (Soon et al. 2008; Haynes 2011; Haggard & Eimer 1999; Keller & Heckhausen 1990; Lau et al. 2004). Soon et al. (2008) is especially

noteworthy: in this functional magnetic resonance imaging [fMRI] study, participants' neural activity was monitored during simple motor decisions. The study found neural activations predictive of the decision up to *10 seconds* before the decision was reported as having been made consciously. Soon et al. (2008) thus conclude, following Libet, that "a network of high-level control areas can begin to shape an upcoming decision long before it enters awareness" (543). Taken together, these studies thus suggest that in simple motor actions, a decision can *seem* to be caused by a conscious decision even while it is in fact caused by some earlier neural activation. P2 then generalizes these findings to include all actions.

And this is all the Moral Responsibility Skeptic needs. For from P1-P2, C1 follows straightforwardly: no one can be held morally responsible for anything.

2.2 Against Moral Responsibility Skepticism

I want to resist the argument for Moral Responsibility Skepticism. There are two general strategies for doing so. First, one could maintain that P2 is false. Adopting this strategy thus means claiming that there are some actions that are directly caused by the conscious decisions of agents. This position remains viable, even in the face of contemporary neuroscience. For Libet's findings simply do not confirm the totalizing claim made by P2, the claim, that is, that there are *no* actions that are caused by conscious decisions. At most, Libet's findings confirm that there are some, relatively simple motor decisions that are not caused by conscious decisions. And the findings may not even show this – Libet's study has been challenged on an empirical level, and these contrary findings need to be taken into account before adopting anything like P2.[8]

Still, the strategy of rejecting P2 outright has its weaknesses. In the first place, while Libet and his followers draw inarguably hasty conclusions, they aren't wrongheaded in worrying that the study has serious implications. Here's why: I take it that our best evidence for not-P2 is that it *seems to us* as if our actions are often caused by conscious decision; our best reason to deny P2, in other words, is our phenomenology of decision making. But if Libet's findings are correct, this phenomenology is *just plain wrong* at times; we can *feel* as if our conscious decision is causally-efficacious even when that decision has already been made on the neural level.[9] So then, while Libet's findings may

8. See, especially, Trevena & Miller (2010) amd Travena & Miller (2002).

9. See, though, Horgan (2011), who discusses the phenomenology of decision making in relation to Libet's findings.

not work as evidence for P2, they do undermine our best evidence for not-P2. Furthermore – and perhaps even more worrisome – Libet's findings suggests that empirical work in neuroscience eventually *could* confirm P2. So, rejecting P2 outright is a weak strategy in that it is open to empirical falsification.

In what follows, then, I target P1 – PCC. More specifically, I will argue that one can be held responsible for actions over which one does not have conscious control and that, therefore, PCC is false.

2.2.1 Counterexamples to PCC

According to PCC, we can be held responsible only for actions that are caused by a conscious decision. Any action which is not caused by a conscious decision but for which we still hold the agent morally responsible will thus function as an effective counterexample to the premise. But it turns out that finding such counterexamples is not very difficult. Consider, for instance, the following:

> **Gangster Wally**: Wally is a gangster, and has lived the gangster life so long that he acts like a gangster *without even thinking about it*. So, when Jerry – who runs the local roulette table – isn't looking, Wally switches out the normal dice with weighted dice out of habit, without consciously noticing what he is doing.
>
> **Saintly Wally**: Wally is walking through a poor section of town when a woman approaches him and asks for a meal. *Without thinking* and purely out of habit, Wally hands her the bag of Taco Bell he has just purchased.

Now, I take it that we want to hold Wally morally responsible for the actions described in each of these scenarios – it seems fully appropriate to blame Gangster Wally and to praise Saintly Wally.[10] However: *ex hypothesi* neither Gangster Wally nor Saintly Wally caused their respective actions by a conscious decision. The lack of conscious control in these scenarios, then, does not seem to affect the fact that we want to direct praise and blame at Wally. But if this is the case, the scenarios serve as counterexamples to PCC –

10. In fact, my own intuitions suggest that Wally is *more* praiseworthy and blameworthy in these scenarios than he would have been had he performed the actions by consciously deciding to do so. For isn't there something *especially* repugnant about the fact that Gangster Wally acts as he does without even thinking, and something *especially* praiseworthy about the fact that Saintly Wally acts as he does without even thinking? I won't belabor the point, as it is tangential to my immediate purposes, but it is, I think, worth thinking about.

they provide cases in which we hold agents morally responsible for actions even when these actions are not caused by a conscious decision.

2.2.2 Responses and Replies

Proponents of PCC might respond to this argument in a number of ways. First, they might argue that Wally cannot accurately be described as *acting* in these scenarios. Remember Velleman's position: since events that are not under one's conscious control are not *one's own*, these events cannot be considered actions. And certainly, if the events described in the scenarios above are not Wally's actions, he cannot be held morally responsible for them.

Now, a thorough reply to Velleman's position would involve a discussion of what does and does not count as an action. And the current paper is hardly the forum for *this*. Still, I want to suggest that while Velleman may accurately identify some strong sense of action, there is also a weaker sense of action, and that this weaker sense is also morally-relevant. For consider:[11] *first*, I take it that common sense counts behaviors caused by unconscious decisions as morally-relevant actions.[12] I take it, for instance, that the following would count as morally-relevant actions on an everyday understanding: my instinctually shoving an elderly lady in a rush to the subway; my habitually greeting my coworkers when I arrive in the morning; my deciding on a whim to ignore a colleague in the hallway. Insofar then as common sense treats such behaviors as morally-relevant actions, the burden lies with Velleman to show where common sense goes wrong. *Second*, it does not follow from a decision's being unconscious that the decision cannot be understood in terms of reasons. Saintly Wally is a good example of this. For it isn't that his decision to give away his Taco Bell is irrational or mysterious; Saintly Wally, I take it, clearly does have reasons for his behavior and could articulate these reasons if he wanted to. It's just that these reasons remain unconscious and unarticulated. But if Saintly Wally's behaves as he does *for a reason*, this counts in favor of his behavior counting as a morally-relevant action on many theories.[13] *Third*, many significant moral decisions are made unconsciously. As Arpaly (2003) suggests, many of us have "made a drastic career change, left a marriage, a

11. These arguments that follow largely follow those put forward by Arpaly (2003).

12. Arpaly (2003) points out – correctly, I think – that moral philosophy can skew our understanding of what counts as an action. The typical agent of moral philosophy, after all, is a conflicted deliberator. But this obscures the fact that *most* of our actions are carried out habitually, not consciously or deliberatively.

13. Scanlon (2008), among others, argues for the relevance of reasons to action. See, for instance, pages 122–131.

church, or a cult, or otherwise made a hard choice" without consciously deciding to do so; we have, rather, *found ourselves* making or having made these decisions (Arpaly 2003, 4). But certainly, if anything should count as an action for which we can be held responsible, important moral decisions such as these should. And if these decisions count as morally-relevant actions, then morally-relevant actions can be caused unconsciously.

Suppose that the proponent of PCC finds these arguments convincing. There is a still another strategy available to her. In particular, she can reason as follows: sure, we want to hold Wally morally responsible in the scenarios presented above. But it isn't that we hold him morally responsible for *what he does*. Rather, we hold him morally responsible for *who he is*. When we praise Saintly Wally and blame Gangster Wally, in other words, we are evaluating Wally as, respectively, a good guy and a bad egg. And this blame or praise is in turn justifiable just in case the actions that led Wally to become who he is were the result of a conscious decision.[14] When we hold Wally responsible in the scenarios presented above, we are thus *ultimately* holding Wally responsible for actions that were the result of a conscious decision; it's just that these actions happened in the past. But if this is the case, the scenarios do not work as effective counterexamples to PCC.

In response: if we suppose that Wally is morally responsible *solely* for past actions, we thereby commit ourselves to the position that Wally's present actions are irrelevant to our praise and blame. But this is problematic. For suppose that Gangster Wally hadn't switched out the dice. If our blame of Gangster Wally is directed exclusively at his past actions, then the Gangster Wally who *doesn't* switch out the dice would be just as blameworthy as the Gangster Wally who does – the two Gangster Wallys, after all, would have performed identical past actions. But this seems obviously wrong. For it seems obvious that the Gangster Wally who switches out the dice is more blameworthy than the one who does not. And the explanation for this also seems obvious: when we blame Gangster Wally, part of what we are doing is blaming him for his *present* action. It is therefore problematic to claim that we praise and blame Wally solely for those actions that led Wally to become who he is.

Proponents of PCC, however, could advance still a third objection. In particular, they could emphasize just how *intuitive* PCC is. PCC, after all, seems able to account for why we don't hold people accountable for twitches, schizophrenic episodes, compelled misdeeds and so on – it seems, after all, that we do not hold people responsible for such actions precisely because these actions are not caused by a conscious decision. At the very least,

14. So, if Wally *didn't* consciously choose to become the way he is, we *wouldn't* be justified in holding him morally responsible. Wolf (1990)'s figure Jojo offers an occasion for thinking about this.

this objector could continue, one should be able to account for why we often make what seem to be fully appropriate counterfactual claims of the form: "had S caused *a* by way of conscious decision, then S would have been morally responsible *a*." But it isn't clear, or so the objection goes, how one can hold on to such claims if one rejects PCC. By rejecting PCC, in short, one loses sight of the obvious: conscious control is importantly relevant to moral responsibility.

I sympathize with this worry, as I do suspect that conscious control is importantly relevant to our moral responsibility practices. But I don't think the objection is damaging to my position. This is for two reasons: *first*, all I have argued here is that conscious control is not a *necessary* condition for holding an agent morally responsible. There is thus nothing I have said that would make it problematic to contruct an account according which conscious control (along with some other conditions, no doubt) is *sufficient* for holding an agent morally responsible. And such an account of moral responsibility, it seems, could more than adequately deal with the intuitive pull of counterfactuals of the form above. *Second*, nothing I have said entails that conscious control does not importantly affect our moral responsibility practices. It is perfectly compatible with everything I have said, for instance, that Gangster Wally would have been more (or less) morally responsible for switching out the dice had he done so consciously.

In short: on a complete account of our moral responsibility practices, it may well be that conscious control is not only relevant, but central. But there is nothing I have said that is incompatible with such an account; all I have tried to show here is that there are actions which are not caused by a conscious decision but for which an agent can still be held morally responsible. For if this is the case, PCC is false, and the argument from Libet's findings presented in section 2.1 does not go through.

3.0 Neuroscience as Evidence for Exemptions

Given what I have argued above, one might wonder: are there *any* ways in which recent advances in neuroscience are importantly relevant to our moral responsibility practices? In the current section, I want to suggest that there are, but that neuroscience finds its proper place in these practices only after some serious philosophical work has been completed. More specifically, I will suggest that neuroscience can be used as evidence for deciding when individuals meet philosophically-defined criteria for moral exemption.

3.1 Neuroscience as Non-Sufficient Evidence for Moral Exemption

Jay Wallace (1994) argues that we should, in certain circumstances, exempt particular individuals from our more responsibility practices, especially our practices of praise and blame. More specifically, Wallace argues that we should exempt any agent from moral responsibility who lacks "the power to grasp moral reasons and the power to control their behavior in accordance with them" (Wallace 1994, 162). Wallace then gives some examples of agents who seem to meet this criterion: young children; individuals afflicted with mental illness; agents under hypnosis; and so on. According to Wallace, these individuals either cannot fully grasp moral reasons or cannot control their behavior in accordance with them, and should not, therefore, be subject to normal moral practices of praise and blame (Wallace 1994).

Now, there are other accounts of moral exemption, and taking sides in *this* debate is not among my purposes here. Rather, I want to point out two features of Wallace's methodology that I find helpful. *First*, Wallace seems right that it is important for us to exempt at least some individuals from our normal practices of praise and blame – clearly, we do not want to blame and praise children and those with schizophrenia in the same way we praise and blame normally-functioning adults. When we praise and blame normally functioning adults, after all, we think of this as a form of deserved condemnation; when we praise and blame children or those with schizophrenia, on the other hand, we think of this as a form of education, protection, adjustment, or regulation. *Second* – and as Wallace also sees – if we want to exempt certain individuals from moral responsibility, it becomes an important philosophical project to determine what the criteria for exemption are. It becomes an important project, in other words, to determine some principled way of deciding when an agent should be held exempt from moral responsibility. This is no small project. And it is, it must be emphasized, a distinctively *philosophical* one. For it seems obvious that any account of moral exemption will have to reference standards of rationality, the appropriateness of our moral responsibility practices, the nature of moral reasons, our ability to grasp these reasons, or some such criteria. And it isn't immediately clear that empirical findings have much to say about any of this.

Once we determine the criteria for moral exemption, however, empirical input becomes important. For suppose that, i.e., we adopt Wallace's criteria for moral exemption – suppose that an inability to grasp moral reasons is sufficient for exemption from moral responsibility. And further suppose that anyone diagnosed with schizophrenia is unable to grasp moral reasons. How then do we determine which individuals are schizophrenic? Clearly, by way of empirical evidence: behavioral analyses, case histories, expert testimony, and so on. And here too is where contemporary neuroscience finds a place of relevance in regard to our moral responsibility practices. Neuroscience has, after all, successfully

located specific neural correlates of schizophrenia, as well as neural correlates of other potentially exempting disorders (Hyde & Weinberger 1990; Torrey et al. 1994; Suzuki et al. 2002; Thompson et al. 2001; Bremner et al. 1995; Gilbertson et al. 2002). So, if we can determine that an individual has one of these neural correlates, this gives us evidence for thinking that the individual should be diagnosed with the relevant disorder. And this, in turn – again, assuming that a diagnosis of the relevant disorder is sufficient evidence that philosophically-determined criteria for moral exemption have been met – could give us reason to exempt the individual from our normal moral responsibility practices.

A word of caution, however: both psychologists and neuroscientists consistently warn how easy it is to become enamored with neuroscientific evidence.[15] For it is all too easy to forget that psychological disorders are defined *behaviorally*; they are not defined in neuronal terms – on its own, then, neuroscientific evidence can never offer conclusive evidence for the diagnosis of any disorder. If it's not clear why this is the case, think about it this way: an individual who exhibits cortical activation in the area of primary motor cortex that is mapped to hand movement but who is *not moving her hand* simply cannot be described as moving her hand. Similarly, an individual who exhibits the neural correlates of schizophrenia but who is behaviorally normal simply cannot be described as schizophrenic. This thus places an important limitation on the relevance of neuroscientific evidence to our moral responsibility practices. In particular, even if being diagnosed with a certain disorder is sufficient for being held exempt from our moral responsibility practices, and even if there is neural evidence that can help make this diagnosis, this evidence can never be taken as sufficient for a diagnosis. Therefore, the neural evidence relevant to moral exemption should never be treated as sufficient evidence for moral exemption.

So, let's take stock. Neuroscience, it turns out, *is* relevant to our moral responsibility practices. In particular, neuroscientific evidence is relevant to our moral responsibility practices as long as the following conditions are met:

> **Criterion Relevancy** – the evidence is framed as providing reason that some philosophically-defined criterion for moral exemption has been met.
>
> **Non-Sufficiency** – the evidence is treated as *non-sufficient* evidence for moral exemption.

15. See, for instance, Morse (2006) and Racine et al. (2005).

These are strict conditions. They may, in fact, lead one to conclude that neuroscience isn't all that relevant at all to our moral responsibility practices, after all. And in a way, this is the thesis of the current paper. Contrary to what some have argued and to the way it may seem, neuroscience *isn't* very relevant to our moral responsibility practices; it finds its proper role only after a significant amount of philosophical work has been completed. And even then, its role is rather humble.

4. Conclusion

In this paper, I have tried to show how neuroscience is and is not relevant to our moral responsibility practices. More specifically: I have argued that – contrary to what some have supposed – the mechanistic explanations of human action offered by neuroscience are not capable on their own of leading us to revise our moral responsibility practices (Section 1). Next, I argued that while findings such as those of Libet et al. (1983) may *seem* to universally exempt us from moral responsibility, universal exemption does not in fact follow from these findings (Section 2). I concluded, however, that neuroscience is relevant to our moral responsibility practices as long as it is properly situated, as long as it is used as non-sufficient evidence that philosophically-defined criteria for moral exemption have been met (Section 3).

References

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Ayer, A. J. 1954. *Philosophical Essays*. New York: Macmillan.

Banks, William P. and Eve A. Isham. 2011. "Do We Really Know What We are Doing? Implications of Reported Time of Decision for Theories of Volition." In *Conscious Will and Responsibility*, edited by Walter Sinnott-Armstrong and Lynn Nadel, 47–61. Oxford: Oxford University Press.

Bayne, Tim. 2011. "Free Will and the Sciences of Human Agency." In *Free Will and Modern Science*, edited by Richard Swinburne. Oxford: Oxford University Press.

Bechtel, William and Robert E. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press.

Breedlove, S.M., N.V. Watson, and M.R. Rosenzweig. 2010. *Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience* (6th ed.). Sunderland, MA: Sinauer Associates, Inc. Publishers.

Bremner, J.D., P. Randall, T.M. Scott, R.A. Bronen, et al. 1995. "MRI-Based Measurement of Hippocampal Volume in Patients with Combat-Related Posttraumatic Stress Disorder." *American Journal of Psychiatry* 152 (7): 973–981.

Chisholm, Roderick. 1964. "Human Freedom and the Self." In *Free Will,* edited by Robert Kane, 47–58. Malden: Blackwell.

Fischer, John Martin. 1987. "Responsiveness and Moral Responsibility," In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand Shoeman, 81–106. Cambridge: Cambridge University Press.

Fischer, John Martin. 1994. *The Metaphysics of Free Will*. Oxford: Blackwell Publishers.

Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Ekstrom, Laura W. 2000. *Free Will: A Philosophical Study*. Boulder: Westview.

Farah, Martha J. and Andrea S. Heberlein. 2007. "Personhood and Neuroscience: Naturalizing or Nihilating?" *The American Journal of Bioethics* 7 (1): 37–48.

Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (1): 5–20.

Gilbertson, M.W., M.E. Shenton, A. Ciszewski, K. Kasai, et al. 2002. "Smaller Hippocampal Volume Predicts Pathologic Vulnerability to Psychological Trauma." *Nature Neuroscience* 5 (11): 1242-1247.

Greene, J. & J. Cohen. 2004. "For the Law, Neuroscience Changes Nothing and Everything." *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* 359 (1451): 1775–1785.

Hallett, Mark. 2011. "Volition: How Physiology Speaks to the Issue of Responsibility." In *Conscious Will and Responsibility*, edited by Walter Sinnott-Armstrong and Lynn Nadel, 61–69. Oxford: Oxford University Press.

Haggard, P. and M. Eimer. 1999. "On the Relation Between Brain Potentials and the Awareness of Voluntary Movements." *Experimental Brain Research* 126 (1): 128–133.

Haynes, John-Dylan. 2011. "Beyond Libet: Long-Term Prediction of Free Choices from Neuroimaging Signals." In *Conscious Will and Responsibility*, edited by Walter Sinnott-Armstrong and Lynn Nadel, 85–96. Oxford: Oxford University Press.

Horgan, Terry. 2011. "The Phenomenology of Agency and the Libet Results." In *Conscious Will and Responsibility*, edited by Walter Sinnott-Armstrong and Lynn Nadel, 159–172. Oxford: Oxford University Press.

Hyde, T. M. and D. R. Weinberger. 1990. "The Brain in Schizophrenia." *Seminars in Neurology* 10 (3): 276–286.

Kane, Robert H. 1999. "Responsibility, Luck, and Chance: Reflections on Free Will and Determinism." *Journal of Philosophy* 96 (5): 217–40.

Keller, I. and H. Heckhausen. 1990. "Readiness Potentials Preceding Spontaneous Motor Acts: Voluntary vs. Involuntary Control." *Electroencephalography and Clinical Neurophysiology* 76 (4): 351–361.

Lau, H. C., R.D. Rogers, P. Haggard, and R.E. Passingham. 2004. "Attention to Intention." *Science* 303 (20): 1208–1210.

Libet, Benjamin. 2011. *Do We Have Free Will?* In *Conscious Will and Responsibility*, edited by Walter Sinnott-Armstrong and Lynn Nadel, 1–10. Oxford: Oxford University Press.

Libet, B., C. Gleason, E. Wright, and D. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act." *Brain* 106 (3): 623-664.

Morse, Stephen J. 2006. "Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic Note." *Ohio State Journal of Criminal Law* 2 (397): 397–412.

Nahmias, Eddy A., Stephen G. Morris, Thomas Nadelhoffer, and Jason Turner. 2005. "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility." *Philosophical Psychology* 18 (5): 561–584.

Nahmias, Eddy, D. Justin Coates, and Trevor Kvaran. 2007. "Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions." *Midwest Studies in Philosophy* 31 (1): 214–242.

Nichols, Shaun and Joshua Knobe. 2007. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *Noûs* 41 (4): 663–685.

Pereboom, Derk. 2013. "Free Will Skepticism, Blame, and Obligation." In *Blame: Its Nature and Norms*, edited by D. Justin Coates and Neal A. Tognazzini, 189–206. Oxford: Oxford University Press.

Pockett, Susan. 2004. "Does Consciousness Cause Behaviour?" *Journal of Consciousness Studies* 11 (2): 23–40.

Racine, E., O. Bar-Ilan, and J. Illes. 2005. "fMRI in the Public Eye." *Nature Reviews Neuroscience* 6 (2): 159–164.

Roediger, H. L., M. K.Goode, and F.M. Zaromb. 2008. "Free Will and the Control of Action." In *Are we Free? Psychology and Free Will*, edited by J. Baer, J. C. Kaufman, and R. F. Baumeister, 205–225. Oxford: Oxford University Press.

Scanlon, Thomas. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Belknap Press.

Soon, C.S., M. Brass, H.J. Heinze, and J.D. Haynes. 2008. "Unconscious Determinants of Free Decisions in the Human Brain." *Natural Neuroscience* 11 (5): 543–545.

Spence, Sean. 2009. *The Actor's Brain: Exploring the Cognitive Neuroscience of Free Will*. Oxford: Oxford University Press.

Strawson, Peter F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1-25.

Strawson, Galen. 1986. *Freedom and Belief*. Oxford: Oxford University Press.

Suzuki, M., S. Nohara, H. Hagino, K. Kurokawa, et al. 2002. "Regional Changes in Brain Gray and White Matter in Patients with Schizophrenia Demonstrated with Voxel-Based Analysis of MRI." *Schizophrenia Research* 55 (1–2): 41–54.

Thompson, P.M., J.N. Giedd, R.P. Woods, D. MacDonal, et al. 2000. "Growth Patterns in the Developing Brain Detected by Using Continuum Mechanical Tensor Maps." *Nature* 404 (6774): 190–193.

Torrey, E.F., A.E. Bowler, E.H. Taylor, and I.I. Gottesman. 1994. *Schizophrenia and Manic Depressive Disorder*. New York: Basic Books.

Trevena, J.A. and J. Miller. 2002. Cortical Movement Preparation Before and After a Conscious Decision to Move. *Consciousness and Cognition* 11 (2): 162–190.

Trevena, Judy and Jeff Miller. 2010. "Brain Preparation Before a Voluntary Action: Evidence Against Unconscious Movement Initiation." *Consciousness and Cognition* 19 (1): 447–456.

van Inwagen, Peter. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27: 185–199.

Velleman, D.J. 1992. "What Happens When Someone Acts?" *Mind* 101 (403): 461–481.

Wallace, R. Jay. 1996. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.

Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 227–248.

Wegner, Daniel M. 2003. *The Illusion of Conscious Will*. Cambridge: MIT Press.

Wiggins, David. 1973. "Towards a Reasonable Libertarianism." In *Essays on Freedom of Action*, edited by Ted Honderich, 31. New York: Routledge.

Wolf, Susan. 1990. *Freedom within Reason*. Oxford: Oxford University Press.

Wright, R. 1994. *The Moral Animal: Evolutionary Psychology and Everyday Life*. New York: Pantheon.