PhiMiSci
Philosophy and the Mind Sciences

Neuroethics and agency

Some comments on Joshua May's *Neuroethics*

Adina L. Roskies[a] (aroskies@ucsb.edu)

Abstract
Neuroethics provides a sober and thoughtful introduction to major issues in neuroethics. This discussion offers a measured assessment and critique of the book.

Keywords
Agency · Neuroethics · Review

This article is part of a symposium on Joshua May's book "Neuroethics: Agency in the Age of Brain Science" (OUP, 2023), edited by Carolyn Dicey Jennings.

Neuroethics, by Joshua May, is a good introduction to issues in neuroethics. It assumes neither background in philosophy nor in neuroscience, and thus is suitable as a text for an undergraduate course. It covers a lot of ground, and its organization and length make it fairly easy to adapt to a quarter or semester course structure. Although I have quibbles about the breeziness of some of the discussion, on most issues May ultimately espouses a position I find to be measured and reasonable. That alone is sufficient to garner kudos from me in a field in which sensationalism seems to be relatively common and largely uncriticized.

The book is subtitled "Agency in the age of brain science", and much of the book centers on human agency, and the way that neuroethical issues revolve around and challenge human agency and our conceptions of it. Framing neuroethics in the context of human agency is a promising strategy. Indeed, many of the central issues in neuroethics involve questions about the nature and limits of human agency, and the way in which the brain, as the seat of the self or the core of the agent, demands analysis that goes beyond the standard fare of bioethical argumentation. May discusses agency in several places, including introducing what he calls the "corporate model of agency", on which agency is distributed among many component parts. However, one never gets a clear picture of what *human* agency amounts

[a] Department of Philosophy, UCSB.

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838
 https://philosophymindscience.org ISSN: 2699-0369

to – i.e. what the relevant parts or functions are – although arguably at least some of every chapter concerns something that has to do with human agency. By the end of the book I found it hard to decide whether organizing around agency was a substantive move or merely a conceit.

The book is subdivided into four sections, Autonomy, Care, Character and Justice, and these are used to organize the ten chapters – one introductory and one summative, sandwiching eight which focus on substantive neuroethical issues. This format enables it to be easily integrated into a college-level course. It seems to be written with this in mind, for other aspects also make it accessible. For example, the book does not presuppose any significant understanding of the brain or knowledge of philosophy, and most of the chapters start with a case study, a real story that is relevant to one or several neuroethical issues under discussion, which may engage the average reader and illustrate how neuroethical issues can arise in real life. What often follows in each chapter is a somewhat simplified sketch of an area, seemingly prompting the reader to consider an extreme position (albeit a position that someone has defended). Then more detail is provided, and arguments offered that walk the reader back to a position that is more nuanced than the one originally sketched. Indeed, May acknowledges late in the book that "All neuroethical analyses require nuance"(May, 2023, p. 257).

May talks about a philosophical toolkit and a map of neuroscience as being essential for neuroethics, and he tries to provide both. In an introductory section on ethics, May provides very short and certainly contestable descriptions of the three main classes of normative ethical theories (consequentialism, deontology and virtue ethics). These paragraph-long descriptions of normative theory are too brief to do them any justice, which May recognizes. However, he justifies the sketchy treatment because he claims "we will not be relying much on them to resolve practical issues in the ethics of neuroscience…[as they are] just as controversial as any issue in neuroethics. It won't help to trade one controversy for another." (May, 2023, p. 21). It certainly would be nice if it were possible to remain theory-neutral when doing neuroethics, simplifying the task and increasing the probability of consensus, so I was looking forward to seeing how this could be accomplished. My hopes were dashed a mere two chapters later, when May assesses the ethics of brain manipulation in terms of costs and benefits, never acknowledging the implicit adoption of one or another of these controversial theories. Unflagged episodes of reliance on normative theories continue throughout the book. Although May does not adopt utilitarianism wholesale (for example, on p. 21 he acknowledges the importance of informed consent, which may be thought of as a deontological principle, but also may also emerge from other approaches), it would help to be clear about the normative assumptions that underlie all our neuroethical deliberations, and what justifies them. In general, May seems to argue in line with a standard bioethical principlist approach (Beauchamp & Childress, 2019), yet does not tell us why we should assume that this framework is the right one for the arguably novel questions raised by neuroethics. May similarly claims to be able to sidestep other contentious moral

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838

issues, such as issues of moral realism/objectivism, relying on (what he indicates is) an unproblematic notion of moral truth. But in ethics there is no free lunch, no normative conclusions without normative premises. I wish this were addressed more straightforwardly.

In the first chapter May also provides a very short introduction to brain science. As with the philosophy, he reviews basic neuroscience. The scientific content of *Neuroethics* is quite accessible to nonexperts, but in its simplicity is sometimes misleading and often problematically incomplete. This is a worry, as I believe that good neuroethics depends on a careful and deep understanding of the neuroscience, and I worry that a model of neuroethics that signals that it is OK to gloss or approximate the science will ultimately lead us astray. There are certainly other works in neuroethics that make egregious errors due to lack of depth of understanding of what the science shows, for example claiming that neuroscience allows 'mind-reading' or that it currently threatens mental privacy. In contrast to these overstated claims, however, May is cautious and measured in his neuroethical analyses, often filling out the scientific landscape more fully as discussion of various topics proceeds. As May himself puts it, the book is an "opinionated introduction to neuroethics…I ultimately defend my own views on these issues, but I aim to provide a comprehensive and balanced discussion."(May, 2023, p. 9).

In what follows I briefly discuss the content in a selection of the chapters. There is no particular rhyme or reason to the selection, but each of these illustrates things that I both like and dislike about the book.

Chapter 2 tackles one of the central topics in neuroethics: free will and responsibility. It begins with the case of a man who murders his wife, and then is found to have a large tumor. Did he act freely? Did his brain make him do it? If it did, do any of us act freely? May starts the discussion by acknowledging that there are different conceptions of free will. He notes that regardless of that diversity most people, including philosophers, think that free will involves choice, control, and/or coherence. While avoiding any precise formulation of free will, he asserts that we have an idea of what free will requires, and then formulates a potential argument sketch for how neuroscience can show that we lack it. He then argues that neuroscience does not show determinism to be true, and that experimental philosophy has shown that people's conception of agency does not necessarily require physicalism to be false. While I strongly agree with both conclusions, they seem too easily reached, especially given the depth of the problems and the rich philosophical literature that has addressed them. In the following section, May discusses the argument that our conscious will is inefficacious (or as he says, epiphenomenal), a form of argument based on the famous Libet studies that still seem to make their way to the public consciousness despite decades of compelling reasons to discount them. Again, May – correctly, to my lights – concludes that these studies do not support the conclusion that we lack free will, but I think he is too uncritical of the science in the studies that purport to show this, and he doesn't clearly diagnose the errors in scientific interpretation that lie at the core of this doctrine. For ex-

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838

ample, in addition to the criticisms that May offers, the purported finding that our brain shows signals reflecting a decision prior to awareness has been shown to likely arise from a mistaken interpretation of the data. Given neurocomputational models of decision-making, what seems to be a brain signal preceding decision appears to be an artifact that arises from the design and analysis of any Libet-style experiment which selects epochs of brain activity based on their culmination in a decision (Schurger et al., 2012). Thus, the signal most likely reflects not the outcome of a decision, but the decision process that produces the decision to act (Schurger et al., 2021). In accepting too uncritically some of these results, May gives too much credibility to Libet-style threats to agency. He concludes the chapter with an interesting "corporate model" of human agency, one which accommodates the notion that we often act on auto-pilot, and argues that normally we are free enough to justify our normal practices of moral responsibility, even if we are often less-than-fully free. As is often the case, I agree with May's conclusion, but less so with how he arrives at it.

In chapter 5 May discusses addiction. Addiction was historically taken to be a moral failing, but many neuroscientists have argued recently that it is a brain disease. May describes the important distinction made by Berridge and colleagues (2008) between liking and wanting, the neural circuitry underlying these forms of motivation, and the basic idea of dopamine as a reward and learning signal in the brain. Nonetheless, May argues against the dominant model of addiction that conceptualizes it as a brain disease, preferring instead to consider it a disorder (note that the already DSM-5 categorizes it as such). The motivation for this is not entirely clear, as it doesn't seem to me that these categories are either perfectly well defined nor mutually exclusive, but several strands provide clues. One is that May seems to think diseases but not disorders preclude control, and he notes that people have some (at least temporary) control over their drug-taking, although he later acknowledges that not all diseases imply lack of control. He argues, too, that cravings are not compulsions, and that drugs work in virtue of the appeal of their effects. Here more detail in understanding the neurochemistry of dopamine as a learning mechanism could be helpful and could ground a more nuanced discussion, as the drug-induced reinforcement of neural pathways seems to me to deviate from normal processes of reinforcement learning. May also seems to argue that if addiction were a brain disease, then neural dysfunction would be necessary *and sufficient* conditions for addiction, but the fact that environmental factors play a role in addiction shows that not to be the case (May, 2023, p. 135). It is not clear, however, why one should construe the requirements for brain diseases that way, given that some diseases, such as diabetes, involve both choice and environmental factors, yet we still categorize them as diseases. These kinds of considerations seem to undermine the motivation for needing to determine whether addiction is a disorder rather than a disease. Nor is it clear why the frequent comorbidity of mental illness with addiction does not better support the brain disease model.

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838

It seems, ultimately, that May's argument rests most strongly on the claim that the neurobiology of addiction is continuous with normal functioning: It is not different in kind from the neurobiology of motivation and choice more generally, and he argues instead that cravings in addicts just differ in degree from normal cases of temptation. If that is true, he argues, then addicts are also responsible for their using behaviors. However, arguably most diseases are continuous with normal functioning in many ways, and as mentioned, it is not clear to me that reinforcement by the drug is equivalent to reinforcement from normal kinds of feedback mechanisms, even if the motivational pathways are the same ones operative in normal reward learning. Moreover, there is abundant evidence showing widespread neural differences between addicts and non-addicts with regard to cravings (Ray & Roche, 2018). Does this suffice to make addiction a disease rather than a dysfunction? To my mind, not that much rides on the answer. None of these arguments seems to undermine the brain disease model unless one makes strong and contentious assumptions about what the term "disease" implies, such as that diseases are always extreme and indicative of categorical differences, or that they are incompatible with responsibility judgments. And if one adheres to the capacitarian model of agency that May seems to endorse early in the book, the point is moot: what matters is the degree to which the changes wrought by addiction impede a person's ability to refrain from using. Perhaps the best argument May offers is a pragmatic one: He thinks the "disease" label is not helpful, and it effectively excuses behavior that is to some degree culpable. He suggests that conceptualizing it as a disorder allows for agency and empowerment of people with addiction, and treatment and compassion on the part of society. Conceptualized as a disorder, addiction does not negate responsibility, although, as Hanna Pickard (2017) argues, it may mitigate blameworthiness.

Of all the chapters in this book, Chapter 8 on motivated reasoning is the least canonical – it does not treat a topic like cognitive or moral enhancement, brain manipulation, or mental privacy, that is recognized as one of the core issues in neuroethics. Instead, this chapter concerns the ordinary phenomenon of reasoning to a conclusion that may not be warranted by the data. The chapter is kicked off by a case study from brain science, the case of people with split brains (severed corpus callosa) – the result of surgery to treat intractable epilepsy. These patients are interesting because they seem to have two separate consciousnesses that are not integrated as they are in normal brains. One documented phenomenon is that when information is shown to one hemisphere of split-brain patients, only that hemisphere has access to the content. When the right hemisphere generates a response, the left hemisphere, which is the hemisphere that houses most linguistic processing, is not privy to the information. If prompted to explain the action, the left hemisphere will confabulate a reason, one that clearly does not accurately reflect the causal processes that drove the behavior being explained.

The fact that we are organisms that will try to rationalize our behaviors if called upon to seems only weakly linked to the main topic of this chapter, which is mo-

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838

tivated reasoning and how it affects scientific integrity and trustworthiness of science. Criticism of scientific methodology seems only marginal as a neuroethical topic, although such criticisms definitely play an important role in the evaluation of scientific claims in neuroethical analysis. What worries me more is that it is a fine line to walk to try to explain the pitfalls in scientific inquiry without unintentionally having a naive reader instead conclude that they should devalue and/or mistrust science. The sciences of the brain are still in their infancy, and understanding the brain is a problem as difficult and fundamental as exploring the far reaches of space. May, for example, criticizes neuroimaging studies for "notoriously small sample sizes," without explaining why small samples abound – e.g. the high cost of imaging, both in money and time. The effect is that low sample sizes seem corner-cutting and unjustified, rather than driven by pragmatic and often unavoidable constraints. At the same time, there is little acknowledgement of how difficult brain science is or discussion of how contextual an information processing engine the brain is, and why, given the brain's complexity, individual differences, and the impact of context, we might find poor replicability even if all of the science is done well.

To be sure, science is full of questionable incentive structures, as is every other human endeavor, but it is still the best way to find out about the natural world. Although the chapter ends by saying this, I worry that it gives an overall impression of science as craven and corrupted by self-interest. Pages are devoted to discussion of p-hacking and statistical significance, but conflicts of interest on the part of corporations only get a passing mention, and the chapter doesn't include a discussion of several other problematic issues, such as sensationalistic media coverage and poor scientific literacy, both of which grease the wheels of scientific misinformation, or of the (perhaps too recent for coverage) growing and very disruptive impacts of AI in producing fake scientific papers. To some degree, in each of his chapters, May provides the scaffolding for discussion, but a lot needs to be filled in by additional reading and/or discussion. This chapter, more than others, calls for interpretation and guidance for students on the part of the instructor.

In the final chapter May revisits the project of neuroethics, reflecting on the themes of the book. Throughout the book May makes the case that abnormal minds tend to differ from "normal" ones in degree and not of kind, and that atypical minds are more like neurotypical ones than unlike them. He also emphasizes that human agency is less dependent on conscious processes, but is also more flexible and adaptable, than we might naively think. Consideration of the diversity of human agency, our constantly evolving neurotechnologies and knowledge of the brain, the situatedness of cognition, and the complexity of situations that people may encounter, argues for treating neuroethical problems on a case-by-case basis, rather than making sweeping moral proclamations.

Most of what he concludes seems like good advice: "Be neither alarmist nor incredulous", and "Resist dubious dichotomies." The real difficulty is to know when something is dubious or when one is being overly worried or failing to adequately

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838

understand the subject matter. Addressing this difficulty, I think, requires a more in-depth knowledge of both philosophy and neuroscience than can be provided in an introductory textbook.

References

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (Eighth edition). Oxford University Press.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2008). Dissecting components of reward: "Liking," "wanting," and learning. *Current Opinion in Pharmacology*, *9*(1), 65–73. https://doi.org/10.1016/j.coph.2008.12.014

May, J. (2023). *Neuroethics: Agency in the age of brain science*. Oxford University Press. https://doi.org/10.1093/oso/9780197648087.001.0001

Pickard, H. (2017). Responsibility without blame for addiction. *Neuroethics*, *10*(1), 169–180. https://doi.org/10.1007/s12152-016-9295-2

Ray, L. A., & Roche, D. J. O. (2018). Neurobiology of craving: Current findings and new directions. *Current Addiction Reports*, *5*(2), 102–109. https://doi.org/10.1007/s40429-018-0202-2

Schurger, A., Hu, P., Pak, J., & Roskies, A. L. (2021). What is the readiness potential? *Trends in Cognitive Sciences*, *25*(7), 558–570. https://doi.org/10.1016/j.tics.2021.04.001

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, *109*(42), E2904–13. https://doi.org/10.1073/pnas.1210467109

Roskies, A. L. (2025). Neuroethics and agency: Some comments on Joshua May's *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.11838