



The promise and perils of moral enhancement with neurotechnology

Daniel D. Moseley^a  (moseleyd24@ecu.edu)

Abstract

Joshua May's *Neuroethics: Agency in the Age of Brain Science* is a landmark achievement in the field. I discuss May's chapter on the topic of moral enhancement. May examines whether it is morally permissible to go beyond treatment and enhance our moral capacities and character via direct brain interventions. May argues for a permissive stance in which the safe experimentation by individuals of some currently available forms of direct brain interventions are permissible. I contend that May's discussion overlooks some central distinctions and arguments in the literature on moral enhancement. Specifically, I argue that the discussion would benefit from: a closer look at the Bioethics literature on the treatment/enhancement distinction, the relevant sense of 'moral' in this debate, arguments in favor of a moral duty to develop and widely deploy neurotechnologies that could be used for moral enhancement, and examining how radical enhancements, transhumanism and moral status issues are central to the topic.

Keywords

Human enhancement · Moral enhancement · Moral status · Neuroethics · Post-persons · Transhumanism

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.

1 Introduction

Joshua May's *Neuroethics: Agency in the Age of Brain Science* is a landmark contribution to the field. May nicely delineates the central issues emerging from current neuroscientific research and the real-world ethical problems encountered in neuroethics. The book provides a useful conceptual framework for organizing and addressing the variegated questions that one encounters in neuroethics. Each chapter merits its own commentary, and those responses will probably lead to innovative

^a Brody School of Medicine at East Carolina University.



contributions to the field. In this brief essay I focus on Chapter 7: *Moral Enhancement*. I'll begin with a summary of the main arguments.

May examines whether we can and should morally improve the brain with the use of current and emerging neurotechnologies. He points to examples of electromagnetic neuromodulation, psychopharmacology and psychedelic drugs that are currently being used for the purpose of moral improvement. He asks whether it is morally wrong to go beyond clinical treatments and enhance our moral character with various types of neurotechnologies. May contends that we should be cautiously optimistic about the prospects of using neurotechnologies for the purpose of moral enhancement. To defend and unpack this view, he examines arguments that are critical of the use of neurotechnologies for moral enhancement purposes. Those arguments appeal to considerations of authenticity, freedom, character, fairness and safety. He claims that the arguments to support these concerns proceed from weakest (authenticity) to strongest (safety). He argues that the state mandated use of neurotechnologies to promote moral enhancement is morally wrong, but it is morally permissible for individuals to safely experiment with psychedelic drugs and neurotechnologies, because that experimentation may lead to moral insights and morally praiseworthy transformations. May admits that these moral gains from experimentation may be modest, but they might lead to widespread cultural, social and moral progress.

I am sympathetic to May's overall cautious optimism about the potential of safely using neurotechnologies for moral enhancement purposes, especially the prospects for safe and responsible use of psychedelics to lead to moral improvements. However, I am less optimistic than he is about the prospects of the other neurotechnological interventions effectively and safely improving our moral character. I articulate those points of disagreement and some additional considerations on the topic in the following sections.

2 The distinction between treatment and enhancement

May quickly dismisses the view that the distinction between treatment and enhancement is morally significant. He writes, "Healthcare providers often draw a sharp distinction between *treatment* for ailments and *enhancement* beyond normal capabilities" (2023, p. 178). The discussion of the treatment/enhancement distinction would benefit from greater nuance. The moral relevance of this distinction is not just an issue in the clinical context about payment for services (e.g., whether insurance should pay for treatments but not pay for enhancements) and clinical interventions; May suggests that the distinction is mainly used in debating whether insurance should pay for treatments but not enhancements. However, the distinction does important work in policy debates about the just allocation of limited health research funding. The debate about enhancement is also not just confined

to people in the clinical and health research domain, but a central topic in Bioethics. In the context of that debate, the term ‘enhancement’ does not have its ordinary usage. In this debate an enhancement is not just an improvement in wellbeing, it is a use of biomedical technology to improve human form or functioning beyond what is necessary to restore or sustain health (Juengst & Moseley, 2019). May discusses moral improvements and improvements in well-being, but those are not the same as moral enhancements.

There are at least two principled reasons why May does not follow the standard definition of ‘enhancement’ in Bioethics. One of the most common reasons for rejecting a principled distinction between treatment and enhancement is a commitment to welfarism (the view that well-being is the only value) or some other type of consequentialist moral theory, such as utilitarianism (Crisp, 2021; Sinnott-Armstrong, 2023). Perhaps May has implicit welfarist or consequentialist commitments. That interpretation fits well with May’s suggestion that safety considerations have the most weight and deontological considerations about authenticity and freedom have the least weight.

A second reason why May might not use the standard definition of ‘enhancement’ is that elsewhere in the book he endorses *the cognitive continuity thesis*: “Neurotypical individuals and people with psychopathologies are more alike psychologically than they are unlike; their psychological differences are primarily a matter of degree, not kind” (2023, p. 109). Since this thesis challenges the assumption that neurotypical and neurodiverse individuals can be clearly categorized, it calls into question the notion of a universal baseline of normal human health and functioning. Thus, the standard definition of ‘enhancement’ is problematic. However, May says that the cognitive continuity thesis is consistent with both the medical model of disability (the view that disabilities are inherently undesirable and appropriate targets of treatment) and the social model of disability (which holds that disabilities are based in the failure of societies to accommodate people who are labeled as disabled). Moreover, these are models of disability, not general models of disease, disorder, injuries and other types of ill-health. May’s view is thus, consistent with a distinction between levels of well-being below and above normal health, and that is all that the standard definition of enhancement requires. Given that the cognitive continuity thesis does not provide a reason to reject the distinction between normal health and levels of well-being above that baseline, it does not provide a conceptual problem for the standard definition.

When discussing the notion of a *moral* enhancement, it is important to clarify (1) what does ‘moral’ refer to in this context and (2) what is the baseline of health (or normality) that must be exceeded for an improvement to count as an enhancement?

3 What is a moral enhancement?

In elucidating the notion of a moral enhancement, May suggests that it involves an improvement in moral behavior that will proceed by neurological means that target certain psychological states or processes (2023, p. 181). What are those states and processes? May provides the following formula for the types of behavioral change that would count as a moral improvement:

Moral judgment + moral motivation → moral behavior

The model represents the idea that moral behavior is generated by beliefs about appropriate ways to act *and* motivation to act in those ways (2023, p. 181). He says this model of moral behavior is limited for at least two reasons. First, it does not indicate the role of willpower and self-control in moral behavior. However, he says that the moral motivation component does capture some degree of the role of self-control in moral behavior. The second limitation is that the model is underspecified: it does not show whether the moral behavior is motivated by a general desire to do what is right or a more specific desire or motivation (2023, p. 182). I think the main problem with this approach to explaining moral improvement is the glaring circularity. It is not *aesthetic*, *prudential*, or *political* or some type of non-moral judgments and motives that explain moral behavior, it is *moral* judgements motives. However, the issues at hand are (1) what is a moral enhancement and (2) how does one distinguish moral enhancements from other types of enhancements?

Mark Rowlands has plausibly argued that most discussions of moral enhancement have not addressed these questions and contends that one needs to use a substantive conception of what morality is to have an intelligible and consistent conception of moral *improvements* and *enhancements* (2018). He argues that order to know whether our moral capacities are improved by improving certain human capacities, one needs to have a clear grasp on the relevant notion of a moral improvement. A Kantian might argue that moral improvement is a matter of improving our capacity to evaluate our own motives and reasons for action. An Aristotelian virtue theorist might argue that moral improvement is a matter of improving the capacities that underly practical wisdom. A Humean sentimentalist might argue that a moral improvement must involve an improvement of one's capacities for empathy or sympathy (Rowlands, 2018). These suggestions are oversimplifications of what a moral enhancement would be according to each of these meta-ethical theories, but they show the inadequacy of assuming that there is a straightforward and univocal sense of 'moral improvement'. These considerations also problematize the notion of a moral baseline that must be set for a moral improvement to count as an *enhancement* (in the standard sense). The Kantian, Aristotelian, and Humean accounts of the moral baseline diverge. Rowlands provides a sketch of how one might use each of those accounts to specify what a moral enhancement would be.

Another approach to avoiding the thorny issue of what counts as a *moral* improvement is to focus on particular examples that capture what matters in the debate. Thomas Douglas examines the nature of moral enhancement by focusing on two central examples: (1) biomedical interventions that target strong aversions to certain racial groups and (2) biomedical interventions that target impulsive, violent aggression (Douglas, 2008). This approach can ask the relevant questions for practical ethics: would it be wrong to use a biomedical intervention (psychopharmacology, psychedelic drugs, non-invasive neuromodulation, electroconvulsive therapy, deep brain stimulation) to make someone less racist or less prone to violent acts.

Along these lines, one might argue that the demand for May to provide a clear conception of moral improvement is unfair: it is clear from the chapter that what May has in mind when he speaks of moral enhancements. He is concerned with very general pro-social attitudes and traits. The types of improvement that he has in mind are the ones listed on Figure 7.1 (2023, p. 184), which include traits like empathy, generosity, and compassion. He regularly refers to these sorts of pro-social attitudes and general traits when discussing the potential of moral enhancement in the near future, and whichever normative ethical view (or metaethical view) one adopts, an increase in general pro-social attitudes and traits is surely morally favorable.

The problem with this reply is nicely illustrated by John Harris's recent discussion of the fallacies involved in appealing to *pro-social attitudes* as evidence of moral improvement (Harris, 2016). He provides many arguments to illustrate this fallacy. One argument is that describing a psychological state as "pro-social" means that it expresses a concern for the well-being of others with whom that person is confronted. However, that attitude of concern (empathy and compassion for certain individuals or for a certain in-group) may actually be anti-social in a larger context when account is taken of the greater good that might have been done by choosing a different strategy that would do more good for more people (Harris, 2016, p. 82). For instance, a person who has "pro-social" attitudes towards their family, local community and nation may lack any concern for, and be completely unwilling to assist, needy people outside of their sphere of concern. Harris contends that enhancing pro-social attitudes with biomedical technologies is morally equivalent to promoting deep-seated irrational prejudices, such as racism or sexism.¹

For many years Harris has defended the view that there is a duty to use biomedical enhancements to improve human capacities. However, when it comes to the topic of moral enhancement, he takes a restrictive approach. He is skeptical of the ability of genetic technologies, nanotechnologies and robotics in application to the brain to target our distinctive capacities for moral judgment and moral action. May

¹ Another point that Harris makes to expose this fallacy is that the term 'pro-social' often means *prudential* or *biologically adaptive*, which is usually contrasted with being *morally right* (Harris, 2016, pp. 129–131). Hence, one cannot infer a moral improvement only based on an improvement of pro-social attitudes.

discusses Harris (2011), which was an earlier argument in which Harris mainly focuses on our right to do what is morally wrong as the basis for rejecting moral enhancement. May presents important objections to that argument (2023, pp. 189–190). However, Harris (2016) provides more arguments in favor of the “right to fall” and provides consequentialist reasons to defend a precautionary stance towards moral enhancement.

4 Are we unfit for the future?

Another question that May does not consider in the chapter but has received a lot of attention in the recent literature is whether there is a moral requirement to develop and widely deploy neurotechnologies that could be used for moral enhancements. One of the main factors driving this recent spike in interest in the topic is Ingmar Persson and Julian Savulescu’s provocative book *Unfit for the Future* (2012). They argue that the survival of our species depends upon the urgent development and widespread use of moral enhancement technologies. They argue that collective action problems in the world cannot be resolved by traditional means or by cognitive enhancements. Rather, they argue that moral enhancements will be required to establish enough cooperation among societies to solve the problems of war, overpopulation, climate change, pandemics, and other social ills that may drive human beings into extinction. They contend that the widespread use of medical technologies for the purposes of moral enhancement is necessary to save human life on earth, and (on welfarist/utilitarian/consequentialist grounds) argue that there is a moral duty to develop and deploy these moral enhancements.

In a similar vein, May writes, “We might not be equipped to handle these social problems without direct, intentional manipulation of ourselves to be better. Conventional forms of moral improvement are limited ... In short, moral education can only go so far, and super-technologies could breed supervillains” (2023, p. 180). Persson and Savulescu argue that moral education has failed us and if we use cognitive enhancements to improve our condition before using moral enhancements, malicious actors could use their enhanced knowledge to become supervillains. So, they maintain that we have a moral duty to quickly develop and widely disseminate moral enhancement technologies. They maintain that doing so will be the first step towards ensuring our survival. The second step would be to develop and widely disseminate cognitive enhancements so that the super-smart people would already be morally good. This is one familiar and influential stance in the moral enhancement literature that May does not consider. Susan Levin points out that the moral enhancement project that Persson and Savulescu defend is based on the claim that the main threats facing humanity are due to a species-wide psychological flaw: inadequate prosocial motivation. She also contends that arguments in favor of the permissibility of using psychopharmacology to manipulate oxytocin and serotonin levels are used as a “practical proof of concept” to show that low-level moral enhancements are possible, and we urgently need to develop more effective

neurotechnology and genetic interventions to increase prosocial attitudes and behavior (Levin, 2021, pp. 85–130). If May shares the consequentialist commitments of Persson and Savulescu, then his arguments might be a first step on the slippery slope to defending a moral duty for radical moral enhancements.

5 Radical enhancements, post-persons and moral status

May observes that “Our topic of moral enhancement is nested within a broader debate about all biomedical interventions that allow one to perform beyond normal” (2023, p. 178). Nicholas Agar points out that the relevant technologies that are usually the subject of debate in this context are GNR: ‘G’ for genetic interventions, ‘N’ for nanotechnology, and ‘R’ for robotics (Agar, 2010, p. 6). Although May focuses on psychopharmacology and neuromodulation technologies, another avenue for future research could focus on how GNR neurointerventions on the current technology horizon could be used for the purpose of moral enhancement. The potential development of GNR neurotechnologies raises serious concerns about turning human beings into a different type of entity: post-humans or post-persons. These questions are often framed as issues of moral status in the enhancement literature (Buchanan, 2011, Chapter 7). Agar describes these type of extreme interventions as *radical enhancements* (Agar, 2010). May briefly discusses Elon Musk’s Neuralink and CRISPR gene editing technologies but sets them aside because “they are less likely to support *moral* enhancement, at least in the foreseeable future.” Instead, May focuses on the more pedestrian cases of using psychopharmacology and current neuromodulation technologies (deep brain stimulation, transcranial stimulation, etc.) to build the case for general optimism about using neurotechnologies for moral enhancement. By setting aside radical enhancements, and debates for and against transhumanism and a post-person future, May ignores one of the more challenging safety concerns and theoretical challenges posed by GNR neurotechnologies. The possibility of radical enhancements leading to the emergence of post-persons is one of the most interesting and challenging ethical issues in the debates about moral enhancement. Agar has defended a very permissive stance towards enhancement uses of technologies in his previous work, but Agar (2010) defends a restrictive stance towards radical enhancements, largely due to safety concerns about radical enhancements. May does not examine the issues of moral status and radical enhancement, but the issues do take on greater urgency for his stance on moral enhancement if he does have the utilitarian commitments that are lurking in the background. Utilitarianism prioritize maximizing overall well-being. If radical enhancements create new types of entities (e.g., cyborgs or A.I. with general intelligence), then the utilitarian needs to find out whether these entities have well-being that needs to be included in the utilitarian calculus. Radical enhancements might produce vastly different capacities for pleasure, suffering or rational agency.

Levin perceptively argues that the philosophical grounds for transhumanism, especially implicit utilitarian commitments, are more widely shared in the literature than is commonly acknowledged (Levin, 2021, pp. 131–169).

6 Concluding remarks

I have only scratched the surface of the fascinating arguments raised in the chapter on moral enhancement and pointed to several areas for promising further inquiry. *Neuroethics* is a rich and important accomplishment, and I plan to take up other arguments from May's excellent book in future work.

Acknowledgments

I am grateful to Carolyn Dicey Jennings and two anonymous referees for their valuable feedback on earlier drafts. I also wish to thank Joshua May for writing this stimulating book and for discussions with him and our "Author Meets Critics" co-panelists, Zina Ward and Tyler Fagan, at the 2024 meeting of the Alabama Philosophical Society. I would also like to thank my colleagues Maria Clay, John Collins and Mike Veber for helpful discussions. Any mistakes or faults with this paper are my own.

References

- Agar, N. (2010). *Humanity's end: Why we should reject radical enhancement*. Bradford.
- Buchanan, A. E. (2011). *Beyond humanity? The ethics of biomedical enhancement*. Oxford University Press.
- Crisp, R. (2021). Well-being. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Douglas, T. (2008). Moral enhancement. *Journal of Applied Philosophy*, 25(3), 228–245. <https://doi.org/10.1111/j.1468-5930.2008.00412.x>
- Harris, J. (2011). Moral enhancement and freedom. *Bioethics*, 25(2), 102–111. <https://doi.org/10.1111/j.1467-8519.2010.01854.x>
- Harris, J. (2016). *How to be good: The possibility of moral enhancement*. Oxford University Press UK.
- Juengst, E., & Moseley, D. (2019). Human enhancement. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Levin, S. B. (2021). *Posthuman bliss? The failed promise of transhumanism*. Oxford University Press.
- May, J. (2023). *Neuroethics: Agency in the age of brain science*. Oxford University Press.
- Persson, I., & Savulescu, J. (2012). *Unfit for the future: The need for moral enhancement*. Oxford University Press UK.
- Rowlands, M. (2018). What is moral enhancement? *Royal Institute of Philosophy Supplement*, 83, 5–18. <https://doi.org/10.1017/s1358246118000267>
- Sinnott-Armstrong, W. (2023). Consequentialism. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Open Access

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

