PhiMiSci
Philosophy and the Mind Sciences

Response to commentators of *Neuroethics*

Joshua May[a] (joshmay@uab.edu)

Abstract
In this reply, I address questions and concerns about *Neuroethics* raised by a formidable cast of commentators: Amanda Evans, Tyler Fagan, Jonathan Pugh, Daniel Moseley, and Adina Roskies. Their insightful commentaries cover most, if not all, topics within the book, namely: neuroscientific challenges to free will, motivated reasoning and the replication crisis, the brain disease model of addiction, moral responsibility and autonomy in mental disorder, valid consent for brain interventions, and enhancement of moral and cognitive capacities through neurotechnologies.

Keywords
Addiction · Brain interventions · Enhancement · Free will · Mental illness · Moral responsibility · Motivated reasoning

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

Tina Fey once said of writing comedy sketches: "It will never be perfect, but perfect is overrated. Perfect is boring on live TV." *Neuroethics* isn't *Saturday Night Live*, but it too is imperfect, partly because it aims to draw in a broad audience. It also advances some controversial views that call for scrutiny, and I'm deeply grateful for the opportunity to do that in this forum. It's an honor to have such thoughtful commentators raise incisive questions and concerns. I've profited greatly from grappling with them individually and altogether in this symposium.

In what follows, I respond to questions and make (some) concessions. The discussion is organized by topic, rather than by commentator, since there is some overlap in their concerns and fruitful connections to be made. I'll begin with free will and motivated reasoning, shift to addiction and other disorders, then wade into consent to brain interventions, and conclude with enhancement and transhumanism.

[a] University of Alabama at Birmingham.

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

2 Free will and scientific evidence

Neuroscience is often thought to reveal that free will is a farce, with the most notorious threat being evidence that our conscious choices are mere byproducts of unconscious brain processes. On this view, our actions are like those of the sleepwalker who runs primarily on autopilot. Although **Adina Roskies** agrees with me that such conclusions aren't well supported by the evidence, she worries that I concede too much to Benjamin Libet and other free will skeptics, that I'm "too uncritical of the science."

In the book, I do get into several prominent concerns with the interpretation of the science, writing that criticisms "abound" (May, 2023, p. 50). There's the legitimate objection that we can't generalize from experiments involving simple arbitrary movements to significant decisions that compare weighty reasons. Opting to move your hand (or not) in an experiment is quite different from deciding to exact revenge or to cheat on a test. I also discuss the critique Roskies highlights: that the relevant brain activity might not represent an unconscious decision but instead background activity in the motor cortex. Even if the unconscious activity is an urge or inclination, that is far from an unconscious *decision*. In the end, I concede only that a great deal of cognitive science suggests that unconscious brain processes do influence our actions. And that evidence goes well beyond the Libet-style studies that Roskies repudiates. I reference studies on the bystander effect, somatic markers, automatic mimicry, and so on—while duly noting some concerns about replication.

To the extent that there is a concession, it was largely strategic. Even if we grant the provocative interpretation of the studies, skeptical conclusions about free will don't follow. After all, the "corporate model" of agency I go on to defend denies that free will always requires conscious decisions. Like a corporation, we can act freely and be held responsible for decisions even if the CEO isn't fully aware of what led to them.

3 Motivated reasoning in neuroscience

Interestingly, Roskies worries that elsewhere I'm too critical of neuroscience generally. In the chapter on motivated reasoning, I argue that brain science suggests people regularly engage in self-deception, including scientists who are, after all, humans too. Conflicts of interest and problematic incentives generate a surprising amount of fraud and questionable research practices that contribute to a replication crisis in many areas of science. Roskies worries that I give the impression that we "should devalue and/or mistrust science" because it's "craven and corrupted by self-interest." Now, as Roskies notes, I explicitly urge readers to draw the opposite conclusion, arguing that "science remains one of the best tools" for procuring knowledge, and I provide "several reasons to be optimistic about its fundamental

May, J. (2025). Response to commentators of *Neuroethics. Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

soundness" (May, 2023, p. 224). So I do hope that most readers will not get the opposite impression.

Yet we should be cautious about undermining faith in science. A small portion of the American public has become less confident in the scientific enterprise since the COVID-19 pandemic (Tyson and Kennedy, 2024), which is disappointing and dangerous. The public should generally rely on science even if not blindly trust every bit of it (Machery, forthcoming). However, it's reassuring that polls indicate most of my fellow Americans (over 75%) express confidence in science (Tyson and Kennedy, 2024) and believe it should be funded by the federal government (National Science Foundation, 2024).

Does discussing motivated reasoning in neuroscience risk eroding this confidence? The book's discussion represents my firm belief that maintaining that confidence—and in some cases winning it back—requires transparency, humility, and a clear-eyed reckoning. Soul searching within psychology and neuroscience has already led to more transparency, open data, and reproducibility (Gilmore et al., 2017).

Perhaps scholars should go soul searching in private without airing their dirty laundry. Yet some readers of *Neuroethics* are scientists or will become practitioners someday, and all of them are aiming to learn about ethical issues in neuroscience. As Roskies says elsewhere in her commentary, "good neuroethics depends on a careful and deep understanding of the neuroscience." I could have spared discussion of the replication crisis, underpowered studies, p-hacking, and overblown conclusions from findings. Such omissions, however, would be a dereliction of duty. Imagine a practicing scientist or ethicist who looks back and says, "Why didn't I learn about any of this in college?"

We must confront these topics head on, not only with fellow scholars but also students and the public. Compare the unethical experiments we drum into students of medical ethics. I teach bioethics every year and many of my students have already heard about the infamous Tuskegee syphilis study, which occurred not far from Birmingham. From 1932–1972, government-funded scientists studied the progression of syphilis in African American residents of Tuskegee while deceiving them about the nature of the research and remaining tightlipped about the development of an effective antibiotic treatment. The ordeal was exposed by an intrepid journalist, and eventually the U.S. government acknowledged its mistake, publicly apologized, and compensated victims. Despite this breach of trust, the majority of Black Americans express confidence in science's ability to act in the public interest (Tyson and Kennedy, 2024). I believe it's likewise important to openly acknowledge limitations and problematic practices within neuroscience, especially when the public already hears stories in the media about controversies, such as fabricated data in Alzheimer's research and hasty approval of drugs (e.g., Piller, 2025; Hiltzik, 2021).

Frank discussions can backfire. Some of my students who were taught about Tuskegee in high school came away with misconceptions about the incident—most

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

notably that the researchers actively infected the residents with syphilis. Although some students get the wrong idea, it remains imperative that we devote ample time to discussing shortcomings in science, both to improve scientific practice and public trust in it. Despite my best intentions, some readers might develop a negative view about neuroscience specifically or science generally. However, if the alternative is to shield readers from the pitfalls of scientific practice, I'd rather risk some misinterpretations than downplay such important matters or perpetuate ignorance of them. While imperfect, I believe sunlight remains the best disinfectant.

4 Addiction and disorder

A candid reckoning is also warranted for the neuroscience of addiction, specifically. For decades now, researchers and health authorities have operated with a brain disease model, which according to one of its founders posits that "prolonged use" of drugs or alcohol throws a "metaphorical switch in the brain" that turns voluntary behavior into "compulsive drug seeking and use" (Leshner, 1997, p. 46). Research within this paradigm has cost millions if not billions in taxpayer dollars and potentially overpromised on its ability to tackle the public health crisis of addiction, which has only grown since the model was developed in the 1990s.

Roskies doesn't think the brain disease model is necessarily flawed. In particular, she is not convinced that hyperactivity of the mesolimbic dopamine pathway in addiction is a mere difference in degree from normal reward learning. I'm not fully convinced either. The idea that addicts experience some dysfunction in relevant brain circuits is the strongest piece of the brain disease model. Now, whether the relevant dysfunction is best thought of as a disease caused primarily by the ingestion of drugs—that's another matter. As I say in the book, "even if addiction involves dysfunction, it is not only or primarily *drug-induced* brain dysfunction" (p. 136). So, contra Roskies, I don't think my rejection of the brain disease model "rests most strongly on the claim that the neurobiology of addiction is continuous with normal functioning." My goal was merely to show that this pillar of the brain disease model isn't ironclad and that the other pillars are even shakier.

The weakest pillar, in my opinion, is the claim that addiction results primarily from the *ingestion of drugs*. There is overwhelming evidence that many people become addicted to drugs—or behaviors like sex and gambling, for that matter—when they are despondent. People often turn to these substances to quell anxiety and silence negative thoughts, which arise from genetic predispositions to mental unwellness and environmental factors like poverty, unemployment, social isolation, and other increasingly frequent motifs of modern life. Of course, environmental factors and cooccurring mental disorders do manifest in the brain (as stress, anxiety, depression, impulsivity, and so on), but the problem for the brain disease model is that these are pre-existing conditions that do not result from *ingestion of drugs* and their effects on the brain. I can't emphasize this enough. The proponents of the brain disease model are not merely claiming that the brain is involved in addiction

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

(even stalwart dualists—from Descartes to David Chalmers—would recognize that). The power and potential of the brain disease model is that it's supposed to illuminate how ingesting these drugs alters the brain in a way that causes compulsive use and relapse (recall Leshner's switch). No doubt the brain disease model can allow that other pre-existing factors like trauma contribute to addiction, but these aren't just background risks—they are arguably the primary causes of addiction, and the brain disease model downplays, if not ignores, them.

We can, however, account for these pre-existing conditions and avoid the other problems with the brain disease model by opting for what I call a *disorder* model. Roskies worries that this is a distinction without a difference, or is at any rate undermotivated. Even if other factors—such as poverty, unemployment, and prior trauma—are important causes of addiction, Roskies objects: "some diseases, such as diabetes, involve both choice and environmental factors, yet we still categorize them as diseases." I completely agree. As I wrote in the book, researchers "actively investigate potential environmental causes" of genuine diseases like "cancer, heart disease, and Parkinson's" (pp. 138–139). Roskies suggests that I may think that "diseases but not disorders preclude control," but that's not the concern. To be clear, I agree genuine diseases can involve control and social determinants. The brain disease model just downplays them—not because it has the word "disease" in the name but because that is how its proponents have conceptualized the model for decades. The boundaries between the words "disease" and "disorder" in general may be hazy, but the brain disease model is crystal clear.

The disorder model—or whatever we should call it—provides a very different approach. It denies that addiction chiefly involves the flipping of a switch in the brain due to the ingestion of addictive drugs. Instead, it focuses on how addictions (including behavioral ones) arise from pre-existing conditions and dispositions within individuals. More than a merely semantic difference, this approach urges us to tackle addiction primarily by addressing poverty, trauma, isolation, access to highly addictive substances, and so on. It also suggests that we're unlikely to discover a pill that solves the problem of addiction by, say, altering the mesolimbic dopamine pathway. Drugs like buprenorphine can be tools for handling withdrawal and reducing the risk of overdose, at least for opioid use disorder. And there is some preliminary evidence suggesting that GLP-1 receptor agonists, like Ozempic, can dampen cravings for alcohol (e.g., Hendershot et al., 2025). However, if the reason someone abuses opioids or alcohol is that they're unemployed, disconnected, anxious, and depressed, then reducing their cravings for the drug won't get to the root of the problem. Ozempic might be marketed as magic, but it doesn't provide a fulfilling job, strong friendships, or the skills necessary to tame negative thought-patterns. No doubt, a pill might help, but a disorder model doesn't regard brain-based interventions as the holy grail in the fight against this public health crisis.

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

5 Responsibility in mental disorder

Having a mental disorder, including substance use disorder, often seems to imply that one is less responsible for conduct that results from symptoms. However, my nuanced view (initially developed in earlier work with Matt King) flatly denies that there is any general connection between moral responsibility and mental disorder. **Tyler Fagan** accurately and charitably expresses the ethical upshot of this picture: "Because we are already inclined to assess the responsibility of neurotypical individuals on a case-by-case basis… we should take the same case-by-case approach to people with psychopathologies."

Yet Fagan worries that too much nuance might neglect useful generalizations that can be made carefully. It's all fine and good to ignore a close friend's diagnosis of schizophrenia and evaluate their responsibility by attending to the particulars of their personality and circumstances. But what about when we lack such an intimate relationship? Fagan argues that in a courtroom, for instance, it can be useful to know that the defendant is schizophrenic and draw some inferences about their culpability on that basis.

That's fair enough, as far as it goes. But what this shows is just that generalizations are sometimes useful in the absence of information, which holds outside of psychopathology. If a blind date stands you up and you learn they're battling cancer, you should reasonably withhold blame based on that information alone. Yet, if a neurotypical friend blows off your dinner party, you might still blame them despite their cancer diagnosis because you know full well that their absence is unfair retaliation for last week's office drama. The point is that, in the absence of any other information, categories and stereotypes can be useful. So I wouldn't say that Fagan is pointing to a special connection between moral responsibility and mental disorder per se, but rather the usefulness of relying on imperfect generalizations in some contexts.

I'd just caution that people too easily overgeneralize and typecast those with atypical brains. Psychiatric and neurological categories aren't just heterogenous; most neurotypical people have little experience or understanding of them. Fagan imagines an example in which someone succumbs to peer pressure and cheats on his taxes. A neurotypical person might be fully accountable for tax fraud, despite facing powerful social pressure, while a person with intellectual disability might be less blameworthy for the same action. However, like autism, intellectual disability is often too oversimplified. As the saying goes, "If you've met one autistic person, you've met one autistic person." Similar caution is familiar from racial or ethnic stereotypes. Even if all you know about a person is that they're Asian or Australian, that might provide some evidence about their accent or food preferences, but we rightly warn against relying on stereotypes. Even if they can provide some information in theory, in practice generalizations are often more ethical trouble than they're worth.

May, J. (2025). Response to commentators of *Neuroethics. Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

6 Autonomy in mental disorder

Our focus so far has been on ordinary blame (or praise) and legal culpability, but I appreciate **Jonathan Pugh's** call to consider medical contexts in more depth. Evaluating responsibility and autonomy case-by-case is the norm for neurotypical patients, but Pugh points out that this isn't always carried over to people with mental health conditions. If I'm understanding him correctly, in the UK neurotypical patients are allowed to make decisions about their own medical care unless that is overridden by an evaluation that the individual lacks decision-making capacity. Yet people diagnosed with a mental disorder can be involuntarily committed without a determination of decisional capacity. Pugh is spot on in predicting that I'd be morally opposed to this practice.

Pugh is also right that I understand the nuanced view to have implications not just for moral responsibility but also for autonomy in this way. In the book, I describe Elyn Saks, a renowned law professor with schizophrenia, but I didn't get into all the relevant details of her case. In her memoir, she writes vividly about the value of being treated as an autonomous person (in the UK no less):

> In England, treatment recommendations were always just that—recommendations. To leave a hospital, to stay in it, to take medications, to participate in group activities or not—they never forced any of it on me, and each time the decision was mine. Even at my craziest, I interpreted this as a demonstration of respect. When you're really crazy, respect is like a lifeline someone's throwing you. Catch this and maybe you won't drown. (Saks, 2007, pp. 79–80).

Eventually, with the aid of antipsychotic medication, Saks was able to realize the importance of exercising her own agency to improve her situation, which meant getting released from an institution.

Yet Pugh worries that responsibility and autonomy are quite separable, especially when it comes to Hanna Pickard's (2013) notion of "responsibility without blame." What is the analog in medical contexts, Pugh asks, when presumably the focus is on autonomy and decision-making capacity, not blame or punishment? The relevant sense of "responsibility" here just is autonomy or decision-making capacity—i.e., patients have "knowledge of what they are doing, and can exercise choice and at least a degree of control over the behavior" (Pickard, 2013, p. 1141). When a patient is responsible without blame, that's just to say they are autonomous and accountable, though it would be inappropriate to overtly blame them or moralize.

Of course, there's a sense in which the individual remains blameworthy. Our linguistic intuitions just need to tolerate sentences like: "He's blameworthy for falling asleep at the wheel, but don't blame him right now… not in the hospital for goodness' sake." In this way, as Fagan suggests in a slightly different context, it seems that *being worthy* of blame doesn't imply that it's appropriate to *actively blame* right now or possibly ever (see also Pickard, 2013, p. 1142; King, 2023, p. 119).

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

How does this apply in medical contexts? Perhaps only when accountability and blame are relevant, as it is for psychiatrists and counselors. When aiming to resolve mental maladies, clinicians often need to hold patients accountable for their choices but without communicating blame or similar hostility. For Pickard, this means a therapist or social worker can and should treat patients as autonomous people capable of directing their own lives, provided they keep blame in check, precisely because letting judgment rip would undermine therapeutic goals. Distinguishing responsibility from blame in this way furnishes healthcare providers with the conceptual tools to respect the patient's autonomy while withholding judgment and other subtle forms of counterproductive sanction.

Of course, blame isn't always forbidden, and it can be difficult to distinguish from merely holding accountable. Sometimes patients benefit from what I like to think of as the "Dr. Phil approach," which forgoes niceties to deliver the hard truth, the tough love. Consider, Orlando, a resident of the Broadway Housing Communities in New York City, who struggles with addiction and bipolar disorder. Orlando is grateful for having met friends who treat him like a regular person, even if that involves moralized judgment. About his friend Yvonne, he says:

> She'll smoke a cigarette and we'll talk, you know? She'll tell me straight up "Look, you're screwing up. I don't want to hear it." (Miller and Spiegel, 2016, 38:01)

In this way, being treated as a person with accountability and sanction can sometimes better support mental health by discouraging a passive attitude toward one's condition that can impede improvement or integration into the community. So I agree with Pugh that there are "moral costs associated with the mitigation of blame for somewhat morally responsible individuals." Importantly, though, often the relationship of friend is required for blaming to work. Not every physician or social worker can play Dr. Phil, and not every patient will respond well to moralizing friends. As usual, it all depends on the case.

7 Brain interventions and valid consent

Respecting a patient's autonomy requires letting them make an informed choice about which treatment to start—or refuse. In these situations, risk-benefit analyses are paramount, especially for serious brain interventions, like deep brain stimulation (DBS) or psychedelics, which can alter a patient's personality and values. Like **Pugh**, I do think these changes are possible and that we should take seriously qualitative evidence about patient experiences. Nevertheless, I remain doubtful these raise special ethical problems, since quite drastic psychological changes due to transformative experiences are common in the lifespan. We don't think there are special ethical problems with going to college, bootcamp, or getting married, so why think they're a special problem with brain interventions?

May, J. (2025). Response to commentators of *Neuroethics. Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

Amanda Evans reads me as a bit too optimistic about DBS for neurological and psychiatric conditions. Interestingly, a neurologist who read the book thought I was unfairly pessimistic about the therapeutic value of DBS! The truth is, despite being cautiously optimistic throughout much of the book, I'm least optimistic about DBS for psychiatric disorders, largely due to the limitations I describe in the book. Rather than being concerned about personality changes, though, I'm quite concerned about insufficient knowledge of neurobiological mechanisms combining with problematic incentives in medical research. I open the precis of this symposium with the case of James Fisher who says DBS has helped him stay clean. I believe we should welcome these successes, but it's not yet clear whether DBS will be an effective treatment for many cases of substance use disorder (and when it is effective other support systems must be in place too). There's danger in overstating our understanding of how brain interventions work when they do and whether the benefits are likely to outweigh costs.

Pugh argues that special concerns about valid consent do arise for brain interventions because these patients can be particularly vulnerable and prone to impairments in decision-making. DBS, which requires brain surgery, is usually a last resort for those who are desperate for relief from debilitating ailments, like Parkinson's and major depression. Physicians should realize, Pugh says, that these neurological conditions, combined with desperation, can lead to distorted reasoning or misconceptions about the treatment's risks and likely benefits. I agree, though in keeping with the discussion of agency in mental disorder, I'd caution against jumping to conclusions and encourage case-by-case evaluation. To me, that's true of other medical contexts too. Patients can be desperate for an experimental cancer treatment or major cosmetic surgery to improve their dating prospects. So I don't see a special problem with acquiring valid consent here.

Evans worries that, unlike Parkinson's and other neurological disorders, we should be more concerned about using invasive brain interventions like DBS for psychiatric conditions in particular. She highlights anorexia nervosa, which is often diagnosed in young adulthood and "not considered to be an irreversible or progressive condition." I appreciate the call for nuance here and agree that invasive interventions might not be appropriate for certain cases. However, the ethical issues don't clearly arise from the interventions being on the *brain* (thus sparing the parity principle). In the case of anorexia, DBS might be too aggressive and ultimately coerce a young adult into surgery that doesn't align with their authentic self. Yet ethical issues like coercion remain pressing with anorexia even when the interventions aren't neurobiological, such as force-feeding or wilderness therapy. Even if I'm cautiously optimistic about brain interventions for both neurological and psychiatric conditions, that's not meant to be a blanket optimism in all cases. It would depend on the condition and whether the risks are outweighed by the potential benefits, as judged by a particular patient's values.

But how can valid consent be acquired when the brain intervention will likely yield a transformative experience that significantly changes the patient's values?

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

Both **Pugh** and **Evans** worry that I too quickly presume an answer to the very problem Laurie Paul originally raised: If an experience will transform a person's values, how can one rationally decide whether to go for it based on what one values presently? In the book, I did skate over this issue, because I do think it's orthogonal. Paul's idea is only that you can't make a "rational" choice in the narrow sense of following the rules of decision theory. Yet even Paul believes we can and should make these decisions—about whether to attend college, join the military, become a parent—we just need to realize that they are a bit more like leaps of faith. These are still perfectly autonomous choices, in my opinion. As the free will chapter suggests, making free choices isn't the same as being *rational*.

Pugh suggests an interesting asymmetry between ordinary transformative experiences and medical ones. Aren't psychological changes resulting from puberty and college more gradual and outside a professionalized relationship of care? I'm not so sure. Many life experiences are sudden or involve a relationship of care. As the parent of a teen, I can confirm that some of the drastic psychological changes wrought by puberty are quite sudden! Transformations in my students' personalities and values may be gradual, but I do occupy a professional relationship with them as a professor. And many transformations following medical treatments are gradual too. A change in view from DBS or chemotherapy can take months or years. So I'm not sure these features will ground an asymmetry between ordinary transformative experiences and medical ones.

Besides, even if transformative medical experience were different, we shouldn't thereby take them to raise special problems for consent. Otherwise, much of medicine would have to change. Obstetricians would have to question whether their patients can validly consent to giving birth—or terminating a pregnancy, for that matter. Oncologists would have to pay special attention to whether their patients can provide valid consent for chemotherapy or surgery. I take this to be a *reductio ad absurdum* that reveals no special problem for obtaining valid consent to brain interventions.

Importantly, that doesn't mean valid consent is irrelevant or that only risk-benefit analyses matter. **Roskies** reads me as presuming a utilitarian or some other form of consequentialist view here (and elsewhere). However, as someone who isn't very sympathetic to utilitarianism as an ethical theory, I see wellbeing as one among many moral considerations. Factoring in a decision's effects on well-being isn't only for utilitarians. Consider the standard "switch" version of the trolley dilemma. Deontologists and virtue ethicists can agree that it's permissible to sacrifice one person to save five, since one ought to choose the option that will minimize harm when all options respect people's rights (Foot, 1967). Similarly, my point wasn't that consent and vulnerability are irrelevant to brain interventions, but rather that in the absence of such concerns a proper assessment of risks and benefits for patients remains (particularly given our limited knowledge of the brain and incentives in medical research for investigating benefits more than harms).

May, J. (2025). Response to commentators of *Neuroethics. Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

Ultimately, there may be little substantive disagreement here. Healthcare providers should always be sensitive to patients' vulnerabilities and decision-making capacities, whether in cases of bodily illness, neurological disease, or psychiatric disorders. My concern is only to avoid overinflating concerns about personality change or impaired autonomy in mental or neurological disorder, which often stem from a misconception of the mind that the book aims to dispel. Far from being regularly rational or joyful or stable, the neurotypical mind is filled with motivated reasoning, negative thoughts, cognitive biases, and transformative experiences—not unlike those minds considered "disordered" (for more on these parallels, see May, forthcoming).

8 Moral enhancement

For these reasons and more, there is plenty of room for improvement in mental health and moral character. Can neurotechnologies help enhance our moral capacities without being unsafe, unfair, or otherwise unethical? In the book, I defend a form of biomedical moral enhancement through individual experimentation with devices like neurofeedback and substances like psychedelics, provided they are sufficiently safe and used to facilitate the development of valuable knowledge and skills.

Daniel Moseley worries that we can't reliably tell what's a genuine moral improvement. He's right that I aim to rely not on a definition of moral improvement but instead relatively uncontroversial examples (à la Kumar and Campbell, 2022), such as terrorism, human trafficking, and the oppression of religious and sexual minorities. Most people (including moral theorists) would agree that being less racist, sexist, and fanatical is a moral improvement. Whether they're distinctively moral improvements, rather than political or social ones, seems like a semantic issue that needn't detain us. We also don't need to settle whether utilitarianism or virtue ethics is the one true moral theory to know that less slavery and discrimination in the world is moral progress. Sure, some people will disagree, but some people think the Earth is flat. Advancing a defensible view doesn't require the absence of any detractors.

Still, some relevant changes aren't necessarily improvements. Becoming more empathic seems morally admirable but could just amount to hoarding one's compassion and resources for friends and family, rather than the most impoverished people (or animals) suffering from harsh treatment and preventable disease. While I take the point, I'm not sure it matters greatly for my aim, which is to determine whether biomedical moral enhancement is permissible. We can just stipulate whatever really is moral improvement using some working examples, and then focus on the potential ethical issues, such as fairness and safety.

Compare an analogy to better health, not morality: Is it ethical to take a pill that will improve one's health, such as GLP-1 drugs like Wegovy or Zepbound? Ethicists, and ordinary people, have worried that using these new drugs to lose

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

weight is unsafe due to side-effects or unfair to those who don't have access. Yet we needn't settle on the one true theory of health to assess fairness and safety. The one true theory isn't irrelevant; it's just not always necessary. Even if we can't be certain that taking a pill to reduce obesity is any healthier, we can grant it for the sake of argument—that is, for the sake of determining whether, supposing it's an improvement in health, taking the pill would be unfair or unsafe overall. The same goes for moral enhancement: assuming that being less self-centered is a moral enhancement, would taking a pill to become less self-centered be inauthentic, unfair, or unsafe? My answer in the book is *not necessarily*.

On my view, to avoid the pitfalls of eugenics, any project of moral enhancement must be pluralistic. As Moseley highlights, in many cases people can reasonably disagree about what's right and wrong. However, just as people can and should freely choose whether to buy a book on how to be an anti-racist or whether to hire a life coach to be less self-centered, people can and should be able to pursue these goals by directly manipulating their brains, provided it's sufficiently safe. Moseley suggests that my discussion of enhancement might rely on some "implicit welfarist or consequentialist commitments." However, as with my discussion of risk-benefit analyses for brain interventions, the focus on safety here is incidental. Safety is one among many moral considerations and it just so happens to be the most pressing of the others in this context.

All this talk of moral improvement might not seem to address *enhancement* specifically. Moseley worries that I too quickly dismiss the treatment/enhancement distinction, partly because I don't seem to "follow the standard definition" of enhancement as improvements that go beyond "what is necessary to restore or sustain health." Although I probably should have explicitly defined enhancement, I did intend to implicitly rely on the standard definition. I characterized enhancements briefly as improvements that go "beyond normal capabilities" to make one "better than well" (May, 2023, p. 178). That characterization seems compatible with the standard definition, and at any rate none of my arguments rely on rejecting it.

To be clear, I believe there is a distinction between treatment and enhancement. I only mean to caution against taking this distinction to be all that informative with well-defined boundaries. Like the categories *bald* and *not bald*, there can be differences between treatment and enhancement, but I emphasized that the distinction is "blurry at best" and the categories lie "on a spectrum" (May, 2023, p. 196). Moseley is right that I think the boundaries are fuzzy partly because the line between "normal" and "abnormal" brains isn't bright—and indeed morally fraught, as the neurodiversity movement has taught (Chapman, 2023; May, forthcoming). In all these cases, the distinctions can be made, but the boundaries are unclear and unprincipled, so we should be wary of relying on them to do much ethical work.

May, J. (2025). Response to commentators of *Neuroethics. Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

9 Transhumanism

One clear example of an enhancement, in the sense of going beyond what's necessary to restore or sustain health, is modifying our brains to become superhuman. **Moseley** calls for more discussion of such "radical enhancements" that would use gene editing or brain implants to make people supremely logical, selfless, just, or whatever else we desire.

The book does set aside radical enhancement rather quickly. That's partly because, even if unethical, I don't think it's all that imminent, since it won't happen on a large scale in the absence of public support or consumer interest. As just one data point, a Pew survey of over 10,000 Americans found that only 13% thought it's a "good idea for society" to pursue "computer chip implants in the brain that would allow people to far more quickly and accurately process information" (Rainie et al., 2022). And that's just for cognitive enhancement, not an attempt to transform the human species into Vulcans.

On these grounds alone, I'm not sure I agree with Moseley that radical enhancement is "one of the most interesting and challenging ethical issues" here. Nevertheless, some of the wealthiest and most powerful people in the world are taking it seriously, especially in the age of AI. Neuralink may be a fairly small operation now, but the company has already implanted its brain-computer interfaces into several human patients with some impressive results. The Neuralink home page states that its ultimate aim is to "unlock human potential" and redefine "the boundaries of human capabilities." Such ambitions do deserve detailed ethical scrutiny.

My own views on transhumanism through radical enhancement are mixed. Some find the idea of post-humanity alarming in principle, as it threatens a distinctively "human essence" that gives us dignity and rights (Fukuyama, 2004). However, as someone who rejects speciesism, I find myself agreeing with transhumanists like Nick Bostrom (2005) who are wary of arbitrarily privileging humans. Gorillas and other non-human animals matter, and I'm glad Sapiens evolved from Heidelbergensis, though not necessarily that this happened at the expense of the Neanderthals. So is it morally acceptable for a new species to exist that goes "beyond human" through enhancement? I don't see why not—provided it doesn't lead to the decimation and oppression of neurotypical or neurodivergent humans (or the monkeys used in relevant research).

But there's the rub. In theory, radical enhancement could be defensible, but in practice it seems unlikely and often stems from unbridled ambition and arrogance in the face of limited knowledge about how the brain works, what will happen in the future, and how to ethically traverse it when we get there. In addition to safety and efficacy, going "beyond normal" presumes a clear conception of what a "normal" human mind is like and how to go beyond it. If *Neuroethics* achieves its main goal, readers will balk at projects that lack humility and fail to appreciate the complexity and continuity of human agency.

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

10 Conclusion

It's an exciting time for neuroethics as a field. Decades ago, the famed neurosurgeon Wilder Penfield proclaimed: "The brain is the organ of destiny. It holds within its humming mechanism secrets that will determine the future of the human race." Perhaps that's right, but we're far from being privy to those secrets. Both neuroscience and ethics have much work to do.

These commentaries have certainly enriched the discussion of neuroethics, which is crucial as we enter a new era in which artificial intelligence is poised to advance our understanding of the brain and how to manipulate it. Although we should be wary of neurohype, advances in neuroscience and AI will continue to challenge our understanding of ourselves and the minds we create. More than ever, rigorous science must coincide with careful philosophical reflection on what it means to be free and how to use novel technologies responsibly.

References

Bostrom, N. (2005). In defence of posthuman dignity. *Bioethics*, *19*(3), 202–214. https://doi.org/10.1111/j.1467-8519.2005.00437.x

Chapman, R. (2023). *Empire of normality: Neurodiversity and capitalism*. Pluto Press.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, *5*, 5–15.

Fukuyama, F. (2004). Transhumanism. *Foreign Policy*, *144*, 42–43. https://doi.org/10.2307/4152980

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*(1), 5–18. https://doi.org/10.1111/nyas.13325

Hendershot, C. S., Bremmer, M. P., Paladino, M. B., Kostantinis, G., Gilmore, T. A., Sullivan, N. R., Tow, A. C., Dermody, S. S., Prince, M. A., Jordan, R., McKee, S. A., Fletcher, P. J., Claus, E. D., & Klein, K. R. (2025). Once-weekly semaglutide in adults with alcohol use disorder: A randomized clinical trial. *JAMA Psychiatry*, *82*(4), 395–405. https://doi.org/10.1001/jamapsychiatry.2024.4789

Hiltzik, M. (2021). The FDA's hasty approval of an unproven Alzheimer's drug is bad news for everyone. *Los Angeles Times*. https://www.latimes.com/business/story/2021-06-10/fda-alzheimers-drug-aduhelm

King, M. (2023). *Simply responsible: Basic blame, scant praise, and minimal agency*. Oxford University Press.

Kumar, V., & Campbell, R. (2022). *A better ape: The evolution of the moral mind and how it made us human*. Oxford University Press.

Leshner, A. I. (1997). Addiction is a brain disease, and it matters. *Science*, *278*(5335), 45–47. https://doi.org/10.3389/fpsyt.2013.00024

Machery, E. (forthcoming). Science without trust. *Philosophy of Science*. https://philsci-archive.pitt.edu/24681/

May, J. (2023). *Neuroethics: Agency in the age of brain science*. Oxford University Press.

May, J. (forthcoming). Neurodiversity with nuance. *Neuroethics*. https://philpapers.org/rec/MAYNWN

Miller, L., & Spiegel, A. (2016). The problem with the solution (season 2, episode 3). Invisibilia [audio podcast]. https://www.npr.org/programs/invisibilia/

National Science Foundation. (2024). *Science and technology: Public perceptions, awareness, and information sources* (tech. rep. No. NSB-2024-4). Science and Engineering Indicators 2024. Alexandria, VA. https://ncses.nsf.gov/pubs/nsb20244

Pickard, H. (2013). Responsibility without blame: Philosophical reflections on clinical practice. In K. W. M. Fulford et al. (Eds.), *Oxford handbook of philosophy and psychiatry* (pp. 1134–1152). Oxford University Press.

Piller, C. (2025). The devastating legacy of lies in Alzheimer's science. *The New York Times*. https://www.nytimes.com/2025/01/24/opinion/alzheimers-fraud-cure.html

Rainie, L., Funk, C., Anderson, M., & Tyson, A. (2022). *AI and human enhancement: Americans' openness is tempered by a range of concerns*. Pew Research Center. https://www.pewresearch.org/science/2022/03/17/ai-and-human-enhancement-americans-openness-is-tempered-by-a-range-of-concerns/

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199

Saks, E. R. (2007). *The center cannot hold: My journey through madness*. Hachette.

Tyson, A., & Kennedy, B. (2024). *Public trust in scientists and views on their role in policymaking*. Pew Research Center. https://www.pewresearch.org/science/2024/11/14/public-trust-in-scientists-and-views-on-their-role-in-policymaking/

May, J. (2025). Response to commentators of *Neuroethics*. *Philosophy and the Mind Sciences*, *6*. https://doi.org/10.33735/phimisci.2025.12199