Précis of Neuroethics

Joshua May

To be published with commentaries in *Philosophy & the Mind Sciences*

(see the call for commentaries, due August 31, 2024)

Abstract: The main message of *Neuroethics* is that neuroscience forces us to reconceptualize human agency as marvelously diverse and flexible. Free will can arise from unconscious brain processes. Individuals with mental disorders, including addiction and psychopathy, exhibit more agency than is often recognized. Brain interventions should be embraced with cautious optimism. Our moral intuitions, which arise from entangled reason and emotion, can generally be trusted. Nevertheless, we can and should safely enhance our brain chemistry, partly because motivated reasoning crops up in everyday life and in the practice of neuroscience itself. Despite serious limitations, brain science can be useful in the courtroom and marketplace. Recognizing all this nuance leaves little room for anxious alarmism or overhype and urges an emphasis on neurodiversity. The result is a highly opinionated tour of neuroethics as an exciting field full of implications for philosophy, science, medicine, law, and public policy.

Word count: 4,256 (5,323 w/references)

Key words: free will, responsibility, brain interventions, mental illness, addiction, psychopathy, moral cognition, motivated reasoning, neurolaw, neurodiversity

Introduction

Imagine you've been grappling with addiction to prescription drugs for over 20 years. During that time, you've survived several overdoses that were nearly fatal, and you've tried multiple rehabilitation programs with no relief. Would you try an experimental treatment that involves cutting open your skull and implanting permanent electrodes deep into your brain that deliver electrical stimulation? That was the decision faced by James Fisher, a patient at the Rockefeller Neuroscience Institute of West Virginia University. At just 36 years old, James decided to undergo neurosurgery to regain control of his life from benzodiazepines. A month after the operation, James is sober and reports, "I'm just not depressed, not anxious, not irritable" (NBC News 2021).

Fischer's story isn't covered in *Neuroethics* (May 2023a), but I ground the discussion in real cases like these. The stories, mostly of patients and defendants, raise ethical issues about a new frontier of brain-based technologies but also about what neuroscience is telling us about ourselves. Is addiction a brain disease that prevents users from being responsible for their relapse? Are any of us truly free if our choices are driven largely by unconscious brain processes? Which brain processes are responsible for knowing right from wrong, and are they absent in psychopathy? Is brain science developed enough to

warrant the use of invasive and expensive techniques on patients, defendants, and consumers? These are questions of neuroethics, which have come into their own since the turn of this century (Illes 2005; Levy 2007; Farah 2010; Racine 2010; Glannon 2013; Roskies 2021). A main theme of the book is that neuroscience forces us to reconceive (not reject) human agency as diverse and flexible even if not wholly conscious and reliable.

These aren't merely academic matters. Over 1 million people have died from drug overdoses since 1999 in the United States alone (NPR 2021). In the late 1990s, there were fewer than 20,000 overdose deaths in America; in recent years rates have climbed to over 100,000 lives—of friends, family members, coworkers—lost each year (NIDA 2023). Yet there is moral danger in being overzealous. We now look back at the days of lobotomies with disdain for the brazen use of brain surgery on vulnerable patients who put too much trust in a new technique touted by numerous physicians. Will our descendants look back on us and recoil at the way we are using neurosurgery and psychedelics to treat problems that are only partly neurobiological and largely social? The book develops a *nuanced neuroethics* that cautions against anxious alarmism and overzealous hype.

Overall, *Neuroethics* provides a highly opinionated tour of this intellectual terrain. The goal is to reveal the field's many implications for philosophers and scientists but also clinicians, lawyers, and policy makers. In what follows, I highlight the opinionated parts: the book's contributions to the literature, including the overarching view of human agency that it develops.

From Free Will to Brain Interventions

A better understanding of the human brain naturally raises the question of whether we really have free will. The sciences of the mind do seem to show that many of our actions are driven largely by unconscious brain processes. We've all had the experience of running on autopilot only to discover that we've engaged in a complex series of actions, such as navigating traffic, without consciously thinking about it. At the same time, we do experience conscious decision making, especially when it matters. We weigh the pros and cons of whether to take a new job, to get married, or to lie on a tax form.

Yet some evidence suggests that these conscious considerations might just come along for the ride while unconscious thoughts and desires really run the show. Study participants do exhibit brain activity indicating a choice among options well before an intention or decision becomes conscious (Libet 1985; Wang et al. 2017). It's not a stretch to think that these situations are more common than we think. But then where is free will? Philosophers have often worried that we're not free if our actions are preordained by prior events, but matters might seem even worse if our choices are determined by processes of which we're utterly *unaware*. Free will might be an illusion, as declared by prominent scientists and commentators, such as Michael Gazzaniga (2012), Sam Harris (2012), and Robert Sapolsky (2023).

Chapter 2 (Free Will) argues that we can view neuroscience as exposing a need to revise, not reject, our conception of freedom and responsibility. We often think of ourselves—the source of our wills—as our conscious minds, like a CEO who directs a corporation (Kennett & Fine 2009; Hirstein et al. 2018). However, it's a mistake to identify a corporation with its executive. Tim Cook plays a special role in directing Apple, but he's only one small part. My *corporate model of agency* suggests that human agents too are a complex amalgam, with executive functions playing only one role among many. It's not a threat to Apple's agency to learn that many of its decisions are driven by factors unknown to its executive. Similarly, it's not necessarily a threat to our freedom and responsibility to learn that many unconscious forces direct our actions. Neuroscience forces us to reconceptualize human free will, not necessarily reject

it, much in the way that physics forced us to reconceptualize solidity upon learning that a bowling ball is mostly empty space.

A more complicated picture of neurotypical agency also forces us to reconceptualize patients with brain disorders. Ethicists often raise concerns about novel brain interventions, like deep brain stimulation, transcranial magnetic stimulation, and psychedelics. Unlike antibiotics or gallbladder removal, intervening on the seat of the self risks altering a patient's identity or hampering their autonomy. However, **Chapter 3** (Manipulating Brains) argues that these concerns are often overblown, given that agency and autonomy are dynamic and flexible.

The key is to recognize the parallels between brain interventions and ordinary life. I draw on L. A. Paul's (2014) notion of transformative experience to show that these are frequent throughout the lifespan. Patients who receive deep brain stimulation or take psychedelics might have their personalities altered, but so do all of us when we go through puberty, higher education, military service, parenthood, and the loss of loved ones. Put in this perspective, brain interventions don't seem to pose a special ethical problem. Indeed, with the rise of mental illness, addiction, and neurodegenerative disorders like Alzheimer's, we shouldn't be overly alarmist about tools that might substantially reduce suffering and even save lives.

One special concern does remain for intervening on patient's brains, one that's often overlooked. Since we know so little about how the brain produces thoughts, emotions, consciousness, and other mental states, there is a risk of medical hubris. Drawing on concerns about the effectiveness of medical interventions generally (e.g., Stegenga 2018), I suggest that we should be especially cautious in neuroscience. It's not even clear why deep brain stimulation helps movement disorders like Parkinson's, let alone why it is sometimes effective for major depression or anorexia nervosa. These make risk-benefit analyses more complicated and prone to error. Nevertheless, provided physicians and other health professionals are honest about the limitations, we should err on the side of allowing patients to exercise their autonomy to choose whether to alter their brains directly, just as they do so indirectly throughout life.

Mental Disorders, Including Addiction

Setting aside ethical issues with brain interventions, questions remain about agency for patients with mental disorders. Mental illness can involve distorted thoughts, perceptions, and desires, making it unclear whether a given patient is fully responsible for any illicit behavior or autonomous enough to refuse treatment. Although it's important to emphasize that people with mental illness are not inherently immoral or dangerous, questions of autonomy relate to multiple ethical issues, including criminal liability, paternalistic treatment, and societal stigma (Kennett 2007; Morse 2011).

Often theorists assume that mental disorder is thereby an excusing condition, but matters are far more complicated than this naïve view allows. In **Chapter 4** (Mental Disorder), I build on some prior work to develop a nuanced view of agency in mental health (King & May 2018). Some ethicists have recognized that certain neurological conditions might impair one's agency while others don't (e.g., Shoemaker 2015; Sripada 2022), but even this is too coarse-grained. Diagnostic categories in psychiatry and neurology are extremely heterogeneous. Symptoms of schizophrenia, for example, often flare up intermittently, and many patients exercise great autonomy while maintaining professional jobs and relationships with friends and family. If someone with schizophrenia commits a crime or fires off some insensitive remarks, their condition doesn't immediately tell us anything about their responsibility or

autonomy. Indeed, the neurodiversity movement has shown that not all people with neurological conditions should be regarded as ill. At least in some cases, autism, dyslexia, schizophrenia, deafness, and other conditions should be regarded as mere differences, not deficits (Chapman 2019; Dwyer 2022). I argue that we must look case-by-case, action-by-action to see whether certain symptoms do compromise one's mental health or agency.

This nuanced approach parallels how we treat people generally. Being neurotypical doesn't mean that one is always fully autonomous or responsible for one's actions. Mitigated blame and punishment are common in the law and ordinary life. Consider the many times we excuse someone's hurtful outburst or inconsiderate planning once we learn that they were unusually sleep deprived, starving, living paycheck-to-paycheck, distracted by an unruly toddler, or overwhelmed by caring for a dying parent. Whether one is mentally well or autonomous all depends on one's circumstances, not whether one can be categorized as neurotypical, as neurodivergent, or as having a mental disorder. This emphasis on the continuity between the neurotypical and atypical might best promote social integration and combat stigma.

A similar approach is taken to substance use in **Chapter 5** (Addiction). The dominant view of addiction among health professionals and many ethicists is now the "brain disease model." On this view, based on decades of research, addiction is a disease like diabetes or hypertension, except that it primarily results from dysfunction in brain circuits that directly impair self-control (e.g. Koob & Volkow 2010; Charland 2002). The basic idea is that illicit drugs (and alcohol) produce abnormally high levels of dopamine in the reward center of the user's brain which over time lead to significant, if not permanent, changes that lead to irresistible cravings and ultimately relapse. The hope is that this model will help to reduce stigma, mitigate blame, and develop effective treatments. Neuroscience research on addiction is plentiful and promises to dramatically improve recovery.

The brain disease model, however, struggles to capture the nuances in addiction seen in various clinical and research settings. Many addicts do exert control over cravings, even without treatment. Remarkably, some addicts do maintain deliberate abstinence and remain sober, often after coming into middle age, getting married, having children, or otherwise "maturing out" (e.g., Lewis 2015; Pickard 2017).

Another issue is that addiction doesn't appear to be primarily a matter of dysfunction of brain circuits *due to ingestion of drugs*. Many factors other than drug consumption contribute to addiction, including financial instability, lack of social support, trauma, and other mental maladies like depression and impulsivity due to borderline personality disorder (e.g., Hart 2013; Pickard 2022). James Fisher, like many addicts, highlights how drugs helped him cope with depression and anxiety. Of course, mental disorders involve brain differences, but the point is that these are often prior conditions not due to the ingestion of drugs, as the brain disease model contends. Now, even genuine diseases have social determinants, so it might make sense to consider addiction a disease, even if it's not primarily due to taking drugs and doesn't always involve irresistible cravings. Yet it's even unclear whether the disease *label* is fitting given that addiction involves normal brain mechanisms, particularly reinforcement learning and homeostasis (Lewis 2015; Hall, Carter, & Forlini 2015; Pickard 2022).

I argue that here too we need to inject nuance into the discussion. Sometimes addiction can involve cravings that are practically irresistible, other times not. Sometimes addicts should be held less accountable for relapse or stealing money to pay for a hit, other times not. It might even make sense to consider some cases of addiction as a disease of the reward system. Perhaps some patients like James will benefit from brain stimulation of the reward center (though another patient receiving the same DBS

treatment in the same clinic did relapse). Given the many other determinants of addiction, however, it will likely be better to see variation as the norm, rather than focus on extreme cases when coming up with models and labels for addiction in medicine, the law, and public policy. We can still at the same time emphasize that addicts need care and support to combat this serious public health problem.

What in the world is addiction if not a brain disease? We needn't go back to considering it a simple moral failing. Rather, I argue that it's a disorder. Its categorization as substance use disorder in diagnostic manuals does an excellent job of recognizing that addiction can be a problem but that it needn't be primarily a brain disease due to ingestion of drugs that invariably causes irresistible desires. As with major depression and schizophrenia, there are no doubt neurobiological elements but there are also environmental and social causes that are often more important to address. Here too we can learn from the concept of neurodiversity and the social model of disability. Often mental health struggles have less to do with the individual's neurobiology considered in isolation and more to do with environmental conditions like stigma, discrimination, isolation, poverty, alienation, and unemployment.

Moral Knowledge and Enhancement

So far, neuroscience has already forced us to rethink human agency as diverse and flexible. The dichotomy between the neurotypical and atypical becomes blurred in addiction, mental illness, and neurological disorders. Nuance is also required in understanding moral cognition. **Chapter 6** (Moral Judgment) argues that we often can trust gut feelings in ethics and politics, for they are essential to rational choice.

Some evidence suggests that gut feelings are crucial for moral knowledge. After all, psychopaths seem to lack proper knowledge of right from wrong and their condition is characterized by impaired emotions like compassion, shame, and remorse (Nichols 2004; Glenn & Raine 2014). Yet it's becoming clear that reason and emotion are deeply intertwined in the brain. Adult patients with damage to the ventromedial prefrontal cortex likewise have attenuated gut feelings which impairs their ability to make appropriate decisions (Damasio 1994; Roskies 2003). Even those with psychopathy, which have dysfunction in this same area as well as the amygdala, often behave quite irrationally, making poor decisions that get them into trouble with the law (Maibom 2005). Gut feelings help us to pick up on relevant information and quickly decide among myriad options, which explains why emotional and rational deficits go hand-in-hand (May 2018; May 2023b).

Of course, gut feelings can lead us astray in ethics. We can be biased by love, self-interest, ideology, and more. These same mechanisms might also drive moral intuitions that tell us the ends don't always justify the means. Utilitarians have long argued that these deontological judgments of commonsense morality are based on unreliable gut feelings (Singer 2005; Greene 2014). Some neurobiological evidence does suggest that these intuitions are driven by automatic emotional heuristics, but I argue that these cannot be easily dismissed as unreliable. Again, gut feelings help humans to appreciate morally relevant information and make sound decisions. Even if automatic, they are far from inflexible, irrational, or untrustworthy. Contemporary moral disputes may be complicated and novel, but gut feelings prove to be quite flexible and adaptable to new circumstances (Railton 2017), not ancient relics of our hominid brains.

Nevertheless, there is much room for moral improvement. The world would be a better place with less violence, domination, oppression, exploitation, and discrimination. By understanding how morality arises in the brain, **Chapter 7** (Moral Enhancement) asks whether we can and should improve it by

directly altering electrochemical signals. Far from science fiction, several forms of brain stimulation, neurofeedback, and pharmaceuticals are able to modulate aggression, compassion, and even moral insight. Classical psychedelics like psilocybin and mescaline can lead to greater compassion and dampened self-interest through the experience of ego-dissolution. These effects are now well-documented in controlled studies (e.g. Forstmann et al. 2020), not just anecdotes, cultural traditions, or folklore.

The question is whether we ought to go beyond treatment to enhance our moral characters through direct brain manipulation. Ethicists have primarily raised concerns about safety, fairness, freedom, authenticity, and resilience (e.g. Sandel 2004; Sparrow 2014). These are legitimate concerns, but concerns alone aren't decisive arguments, as Allen Buchanan so nicely puts it (2011: 144).

In particular, these various qualms aren't knockdown arguments against biomedical moral enhancement carried out freely by individuals exploring their own conception of the good life. We can rightly object to state-mandated moral enhancement by lacing the water supply with oxytocin. It's much harder to condemn individuals who choose to safely experiment with psychedelics or wearable devices that provide feedback on one's brain waves, similar to smartwatches that track heart rate and sleep patterns. As with any commercial device, safety must be carefully monitored, but we should be less concerned when the drugs or devices are not habit forming and their effects are reversible. Psychedelics and brain stimulators won't lead to utopia, but they might facilitate moral insight and moral transformations when combined with reasoning (Earp, Douglas, Savulescu 2018). Although modest, these individual explorations can fuel cultural evolution and pockets of moral progress.

Neuroscience and Society

Many ethicists are rightly concerned that neuroscience isn't ready for mainstream use outside of medicine. The discipline and its technologies are still relatively new, and we know so little about how the brain gives rise to specific thoughts, preferences, and emotions. Even worse, there is a replication crisis in science, including the social sciences, and neurobiology isn't immune to questionable research practices driven largely by bad incentives (Ioannidis 2005).

Chapter 8 (Motivated Reasoning) discusses how brain science reveals the pervasive influence of self-deception and rationalization on our everyday choices but also on scientific practice. Motivated reasoning crops up in many forms, from wishful thinking about one's chances on the job market to willful ignorance about one's carbon footprint (Kunda 1990; May 2018: ch. 7). These are familiar human foibles well-supported by scientific evidence, but what's particularly fascinating is how these patterns of reasoning arise among researchers themselves. Scientists are human too, and they work within a ruthlessly competitive "publish or perish" system that rewards the production of novel, surprising, or otherwise interesting findings, even if produced by flawed methods and reporting, such as p-hacking, fishing expeditions, and overgeneralization (Tullett 2015; Yarkoni 2022). In neuroscience, questionable research practices have led to overlooked errors in statistical analyses, underpowered studies with too few participants, immodest claims about what we can infer from brain images, and more. Other times, of course, researchers who are careful and properly motivated produce genuine findings that advance our knowledge of the mind and its ailments.

Motivated reasoning provides a useful framework for understanding these patterns. The desire for knowledge and truth does shape much reasoning in scientific practice, but so do non-epistemic values. Of these, there are at least five intrinsic motivations that shape reasoning in science: profit, credit, ideology, altruism, and spite (a list that builds on May 2021). For example, researchers can be motivated—often

unconsciously—to produce findings that will earn them a Nobel prize (credit and profit), promote their pet theory (ideology), help their graduate students get jobs (altruism), or ruin the career of their archenemy (spite). Acknowledging the existence of such motivations is compatible with science being one of our best sources of knowledge. Partly because it is adversarial and collaborative, the process can yield knowledge through a marketplace of ideas. Still, we can do better as producers and consumers of science; we can be less credulous and more vigilant.

With humility in mind, we turn in **Chapter 9** (Brain Reading) to whether neuroscience can be trusted to read the minds of criminals and consumers. The science is already being used to both convict and exonerate, through neurological exams and brain imaging. Some companies claim to be able to detect when someone is lying or recognizes a key piece of information about a crime. Yet there are many limitations to drawing such inferences from brain data (see e.g. Sinnott-Armstrong et al. 2008; Satel & Lilienfeld 2013). The base rate fallacy, for example, looms large over discussions of brain abnormalities in either defendants or plaintiffs. The massive variation in brain activity across individuals and times makes it difficult to even posit "normal" activity to which one might deviate. Reverse inference plagues conclusions about one's mental state made on the basis of functional magnetic resonance imaging. And judges and juries are apparently liable to overinflate the value of evidence simply because it's neurobiological (although much of that research has been overblown).

These become ethical issues if we rely uncritically on neurobiological evidence in the law to either convict the innocent or exonerate the guilty. Despite these serious limitations, however, I argue that we should generally embrace neuroscience in the courtroom (contra e.g. Sinnott-Armstrong et al. 2008). The primary reason is simply that there is already such a paucity of good evidence in the law that we should welcome more, even if it is limited. Solid research has revealed the unreliability of human vision and memory, yet we allow eyewitness testimony. By parity of reasoning, we should allow neurobiological evidence, provided we aren't overzealous or overly credulous.

A similarly anti-alarmist story can be told about the use of neuroscience in the marketplace. Big tech companies and startups aim to detect customer preferences through brain reading and to produce brain-computer interfaces that will not only rehabilitate patients but empower ordinary consumers. Companies like Neuralink aim to eventually provide brain implants for mass consumption that would be akin to smartwatches and other wearable devices that provide biofeedback to monitor one's heart rate and improve sleep. Ethicists and the public are understandably concerned about privacy and worsening mental health as technology creeps further into our lives and dominates our attention (e.g. Murphy, Illes, & Reiner 2008; Kreitmair 2019; Farahany 2023).

The serious limitations of neuroscience muffle these alarm bells. Researchers have recently been able to decode speech from a woman who can no longer talk due to ALS (Metzger et al. 2023), and the first human patient with a Neuralink implant is able to play chess just by thinking. Nevertheless, brain reading has proven much less efficient and cost-effective than the many ways our digital footprints are already monitored and monetized. We should worry about further losing privacy and unintended consequences of technology, but our fears and objections might be better trained on existing products—such as our smartphones, credit cards, and security cameras—and the big data they produce. These are arguably poised to constrain freedom and violate rights more than any reasonably accurate and affordable neurotechnologies.

Nuance and Neurodiversity

Neuroethics aims to provide a balanced view of neuroscience and the moral questions it raises and helps to answer. Free will is arguably compatible with our decisions being largely influenced by unconscious brain processes. Brain interventions don't threaten patient autonomy and identity any more than transformative experiences typical of ordinary life. Individuals grappling with mental disorders, including substance use disorder, often have the agency and autonomy to be held accountable and shape their own lives. Commonsense moral intuitions can generally be trusted even amid political disagreement. Although imperfect, we can ethically enhance our moral traits by altering our brain chemistry. The human brain is fundamentally a reasoning machine, but one that is inevitably biased by its own goals and values, which explains why even scientists themselves can engage in questionable research practices. Despite all these limitations, neuroscience can be useful in the law and the marketplace to promote justice and human flourishing.

The book covers only a fraction of the field. Various topics are entirely omitted or only briefly addressed, such as the criteria for death, minimally conscious states, brain organoids, a right to cognitive liberty, the use of animals in neuroscience research, and purported gender differences in brain structure and function. My hope is that the book provides a nuanced framework that can be useful for addressing all topics within neuroethics.

This framework, described in **Chapter 10** (Nuanced Neuroethics), includes five guiding principles. Philosophy and neuroscience should work in *harmony* as each provides checks and balances against one other. Ethics isn't all hairsplitting and neuroscience isn't all overhype. Yet *humility* does require us to avoid anxious alarmism and credulous enthusiasm, to more often revise rather than reject our conception of human agency. Scientific evidence should be *scrutinized* without credulity to identify the best empirical evidence on which to base theorizing. And our best science urges us to recognize the *complexity* of human agency as arising from conscious and unconscious processes that entangle reason and emotion. Finally, this complexity warrants an emphasis on *continuity* among human brains over categories that divide. Not all cases of addiction and autism are alike, and there is much more agency to be found than is often recognized. Similarly, the neurotypical are not a homogenous class of individuals who are largely rational and mentally well. It depends on the person and the time of day. Above all, we should recognize that diversity and variation in brain function are the norm. We all move about fluidly along the great Cognitive Continuum.

References

Buchanan, A. (2011). *Better than human: the promise and perils of enhancing ourselves*. Oxford University Press.

Chapman, R. (2019). Neurodiversity theory and its discontents: Autism, schizophrenia, and the social model of disability. In S. Tekin & R. Bluhm (eds.) *The Bloomsbury Companion to Philosophy of Psychiatry* 371-389.

Charland, L. C. (2002). Cynthia's dilemma: consenting to heroin prescription. *American Journal of Bioethics* 2(2):37– 47.

Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain*. Avon Books.

Dwyer, P. (2022). The neurodiversity approach(es): What are they and what do they mean for researchers? *Human Development* 66(2), 73-92.

Earp, B. D., Douglas, T., & Savulescu, J. (2018). Moral neuroenhancement. In Johnson & Rommelfanger (Eds.) *The Routledge handbook of neuroethics*. Routledge, pp.166– 184.

Farah, M. J. (Ed.). (2010). *Neuroethics: an introduction with readings*. MIT Press.

Farahany, N. A. (2023). *The battle for your brain: defending the right to think freely in the age of neurotechnology*. St. Martin's Press.

Forstmann, M., Yudkin, D. A., Prosser, A. M., Heller, S. M., & Crockett, M. J. (2020). Transformative experience and social connectedness mediate the mood- enhancing effects of psychedelic use in naturalistic settings. *Proceedings of the National Academy of Sciences* 117(5): 2338– 2346.

Gazzaniga, M. S. (2012). Free will is an illusion, but you're still responsible for your actions. *The Chronicle of Higher Education*. https://www.chronicle.com/article/michael-s-gazzaniga-free/131167

Glannon, W. (2013). *Brain, body, and mind: Neuroethics with a human face*. Oxford University Press.

Glenn, A. & Raine, A. (2014). *Psychopathy*. New York University Press.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics*, 124(4), 695-726.

Hall, W., Carter, A., & Forlini, C. (2015). The brain disease model of addiction: is it supported by the evidence and has it delivered on its promises? *The Lancet Psychiatry* 2(1): 105-110.

Hart, C. (2013). *High price: a neuroscientist's journey of self- discovery that challenges everything you know about drugs and society*. Harper.

Harris, S. (2012). *Free will*. Simon and Schuster.

Hirstein, W., Sifferd, K. L., & Fagan, T. K. (2018). *Responsible brains: neuroscience, law, and human culpability*. MIT Press.

Illes, J. (Ed.). (2005). *Neuroethics: Defining the issues in theory, practice, and policy*. Oxford University Press.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine* 2(8): e124.

Kennett, J. (2007). Mental disorder, moral agency, and the self. In Bonnie Steinbock (Ed.), *The Oxford handbook of bioethics*. Oxford University Press, pp. 90– 113.

Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? *Ethical Theory and Moral Practice* 12(1): 77– 96.

King, M., & May, J. (2018). Moral responsibility and mental illness: A call for nuance. *Neuroethics*, *11*, 11-22.

Koob, G. F., & Volkow, N. D. (2010). Neurocircuitry of addiction. *Neuropsychopharmacology* 35: 217–238.

Kreitmair, K. V. (2019). Dimensions of ethical direct-to-consumer neurotechnologies. *AJOB Neuroscience* 10(4): 152–166.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin* 108(3): 480-498.

Lewis, M. (2015). *The biology of desire: why addiction is not a disease*. PublicAffairs.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8: 529– 566.

Levy, N. (2007). *Neuroethics: Challenges for the 21st century*. Cambridge University Press.

Maibom, H. L. (2005). Moral unreason: The case of psychopathy. *Mind & Language* 20(2): 237-257.

May, J. (2018). *Regard for Reason in the Moral Mind*. Oxford University Press.

May, J. (2023a). *Neuroethics: Agency in the Age of Brain Science*. Oxford University Press.

May, J. (2023b). Moral Rationalism on the Brain. *Mind & Language* 38(1): 237-255.

May, J. (2021). Bias in science: natural and social. *Synthese* 199: 3345– 3366.

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., ... & Chang, E. F. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* 620(7976): 1037-1046.

Morse, S. J. (2011). Mental disorder and criminal law. *The Journal of Criminal Law and Criminology* 101(3): 885– 968.

Murphy, E. R., Illes, J., & Reiner, P. B. (2008). Neuroethics of neuromarketing. *Journal of Consumer Behaviour* 7(4‐5): 293– 302.

NBC News (2021). Experimental Brain Surgery Aims To Treat Severe Substance Abuse. https://youtu.be/FCP64qa2AOQ

Nichols, S. (2004). *Sentimental rules*. Oxford University Press.

NIDA (2023). Drug Overdose Death Rates. *National Institute on Drug Abuse*. https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

NPR (2021). More than a million Americans have died from overdoses during the opioid epidemic. *National Public Radio*. https://www.npr.org/2021/12/30/1069062738/more-than-a-million-americans-have-died-from-overdoses-during-the-opioid-epidemi

Paul, L. A. (2014). *Transformative experience*. Oxford University Press.

Pickard, H. (2017). Responsibility without blame for addiction. *Neuroethics* 10(1): 169– 180.

Pickard, H. (2022). Is addiction a brain disease? A plea for agnosticism and heterogeneity. *Psychopharmacology* 39: 993–1007.

Racine, E. (2010). *Pragmatic neuroethics: Improving treatment and understanding of the mind-brain*. MIT Press.

Railton, P. (2017). Moral learning: Conceptual foundations and normative relevance. *Cognition*, 167, 172-190.

Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy". *Philosophical Psychology* 16(1): 51-66.

Roskies, A. (2021). Neuroethics. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2021/entries/neuroethics/

Sandel, M. (2004). The case against perfection. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2004/04/the-case-against-perfection/302927/

Sapolsky, R. (2023). Determined: A Science of Life without Free Will. Penguin Press.

Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: the seductive appeal of mindless neuroscience*. Basic Books.

Shoemaker, D. (2015). *Responsibility from the margins*. Oxford University Press.

Singer, P. (2005). Ethics and intuitions. *The journal of ethics*, 9, 331-352.

Sinnott- Armstrong, W., Roskies, A. L., Brown, T., & Murphy, E. (2008). Brain images as legal evidence. *Episteme* 5(3): 359– 373.

Sparrow, R. (2014). Better living through chemistry? A reply to Savulescu and Persson on "moral enhancement." *Journal of Applied Philosophy* 31: 23– 32.

Sripada, C. S. (2022). Mental disorders involve limits on control, not extreme preferences. In M. King & J. May (eds.) *Agency in Mental Disorder*. Oxford University Press.

Stegenga, J. (2018). *Medical nihilism*. Oxford University Press.

Tullett, A. M. (2015). In search of true things worth knowing: Considerations for a new article prototype. *Social and Personality Psychology Compass* 9(4): 188-201.

Wang, J., Cherkassky, V. L., & Just, M. A. (2017). Predicting the brain activation pattern associated with the propositional content of a sentence: modeling neural representations of events and states. *Human Brain Mapping* 38(10): 4865– 4881.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences* 45:e1.