The Morality of Moral Neuroenhancement

Thomas Douglas
thomas.douglas@philosophy.ox.ac.uk

[This is a pre-publication version. The final version will be available here]

Abstract. This chapter reviews recent philosophical and neuroethical literature on the morality of moral neuroenhancements. It first briefly outlines the main moral arguments that have been made concerning moral *status* neuroenhancements. These are neurointerventions that would augment the moral status of human persons. It then surveys recent debate regarding moral *desirability* neuroenhancements: neurointerventions that augment that the moral desirability of human character traits, motives or conduct. This debate has contested, among other claims (i) Ingmar Persson and Julian Savulescu's contention that there is a moral imperative to pursue the development of moral desirability neuroenhancements, (ii) Thomas Douglas' claim that voluntarily undergoing moral desirability neuroenhancements would often be morally permissible, and (iii) David DeGrazia's claim that moral desirability neuroenhancements would often be morally desirable. The chapter discusses a number of concerns that have been raised regarding moral desirability neuroenhancements, including concerns that they would restrict freedom, would produce only a superficial kind of moral improvement, would rely on technologies that are liable to be misused, and would frequently misfire, resulting in moral deterioration rather than moral improvement.

In recent years, a number of philosophers and neuroethicists have become interested in moral enhancements. These are usually taken to be interventions that aim to and (expectably) succeed in augmenting the morality of an individual or one or more of her traits. This chapter follows recent literature in discussing only moral enhancements of *human persons* or their traits and the term 'moral enhancement' is henceforth used to refer only to enhancements of this sort.

Humans have traditionally sought to morally enhance themselves through means such as introspective reflection, engagement with literature or calm moral discussion with others, and to morally enhance others through rational persuasion, the application of incentives or disincentives or moral education. However, prompted by developments in the cognitive sciences, much recent discussion of moral enhancement has focused instead on moral neuroenhancement. Moral neuroenhancement can be distinguished from more traditional varieties of moral enhancement by the fact that it operates by altering brain states or processes directly, that is, not by (i) modifying the agent's environment (as in the application of incentives), or (ii) by engaging the agent's deliberative capacities (as in introspective reflection or rational persuasion). Paradigmatic examples of moral neuroenhancement would operate via the administration of drugs or application of brain modulation techniques such as electrical or magnetic brain stimulation.

Moral neuroenhancements all involve increasing the morality of an individual or trait. Thus, two natural ways of classifying moral enhancements are according to the *metric* by which morality is measured and the *target* of the enhancement—that is, the entity whose morality is augmented. Possible metrics of moral neuroenhancement would include be moral status, moral considerability, moral responsibility, moral understanding, moral virtue, moral desirability, moral rightness, or moral permissibility. Possible targets of moral neuroenhancement would include the human person or his character, motives or conduct. There are various ways in which one could combine different metrics and targets to describe different varieties of moral neuroenhancement. However, two kinds of moral neuroenhancement have dominated recent discussion of this topic, namely: (1) neuroenhancements of the moral status of humans (henceforth, moral status

neuroenhancements or MSNs) and (2) neuroenhancements of the moral desirability of human character traits, motives and conduct (henceforth, moral desirability neuroenhancements or MDNs).

Two main kinds of question have been asked about MSNs and MDNs: questions about their possibility, and questions about their morality. For example, there have been discussions of whether MSNs or MDNs are possible or likely to become technically feasible. And there have also been discussions of when, if ever, it would be morally permissible, desirable or obligatory to (a) pursue the development of technologies that would or might enable MSN or MDN, (b) undergo a MSN or MDN oneself, or (c) impose such an intervention on others.

This chapter outlines some of the main arguments that have been advanced concerning the morality of MSNs (section 1) and MDNs (sections 2-5). Discussions concerning the possibility of MSNs and MDNs are touched upon where relevant to arguments concerning their morality, but are not comprehensively surveyed.

1. Moral Status Neuroenhancements

Moral status neuroenhancements aim to and (expectably) succeed in bringing it about that a human enjoys greater moral status than she would otherwise have enjoyed. There is disagreement regarding how to understand moral status, but one way of understanding it is as a metric of the strength and breadth of moral protections (such as moral rights or claims) enjoyed by a being. Understood thus, moral status neuroenhancements could operate, for example, by conferring additional moral rights on humans.

Recent discussion of moral status enhancements has taken place primarily within debates on the morality of general cognitive or mental neuroenhancements (CROSSREF: section introduction 'Neuroenhancement' and chapters 'Reflections on Neuroenhancements', 'Neuroscience, drug policy, and drug regulation, including enhancement use of pharmaceuticals', 'The Ethics of Mood Enhancement'). It has in large part been stimulated by a speculative comment offered by Francis Fukuyama in a 2004 critique of transhumanism, the movement most committed to human neuroenhancement. Francis Fukuyama writes that

> Underlying [the] idea of the equality of rights is that we all possess a human essence that dwarfs manifest differences in skin color, beauty, and even intelligence. This essence, and the view that individuals therefore have inherent value, is at the heart of political liberalism. But modifying that essence is the core of the transhumanist project. If we start transforming ourselves into something superior, what rights will these enhanced creatures claim, and what rights will they possess when compared to those left behind? (Fukuyama 2004, p. 42)

One of Fukuyama's worries here is that enhanced beings might *claim* greater rights than are enjoyed by unenhanced humans. But another concern is that enhanced beings would *actually have* greater rights than the unenhanced. This latter concern can be interpreted as a concern that human neuroenhancement could result in the creation of beings with greater moral status than ordinary persons (call such beings status-enhanced beings or SEBs). This proposition has been the object of substantial subsequent analysis, most notably by Allen Buchanan (2009;2011;2012).

Why might neuroenhancement result in the creation of SEBs? The usual explanation draws on an analogy. It is often thought that ordinary adult humans enjoy higher moral status than at least some of our nonhuman mammalian relatives. In addition, it is often thought that our elevated moral status, relative to other mammals, can be attributed to our possessing greater mental capacity than other mammals. (Precisely which mental capacities confer higher moral status on us is open to dispute, but candidates include our capacities for self-awareness, rationality and moral agency.) But if we enjoy greater mental capacity than our mammalian relatives in virtue of our greater mental capacity, one might expect that beings whose mental capacity exceeded our own to a similar degree would enjoy greater moral status than us, and it

might seem that the neuroenhancement of humans could, at least in principle, produce such beings.

But the creation of SEBs would, it has been argued, be morally problematic since the existence of SEBs would be bad for ordinary persons (e.g. Agar 2012). Arguably, ordinary persons could be permissibly harmed for the sake of these SEBs in ways that they may not be permissibly harmed for the sake of one another. For example, perhaps persons could permissibly be used, without their consent, in medical experiments designed to aid SEBs. Or perhaps persons could be rightly excluded from the democratic institutions of the SEBs.

Three main responses have been made to this line of thought. First, some authors have questioned whether human neuroenhancements could produce beings of greater moral status than ordinary persons. For example, Allen Buchanan (2009; 2011) and James Wilson (2007; 2012) argue for a threshold account of moral status according to which there is some level of mental capacity above which all beings have the same moral status, and beyond which ordinary persons already lie (see also Savulescu 2009; Wikler 2009). Buchanan invokes Stephen Darwall's (2006) work to argue that the ability to engage in practices of mutual accountability is what takes one across the threshold, while Wilson argues that a being has crossed the threshold if it possesses Rawls' two moral powers—a sense of justice and a "capacity to have, revise, and rationally to pursue a conception of the good" (Rawls 2001, p. 19). In reply, a number of authors have questioned whether there is good reason to prefer this threshold account of moral status to (a) a scalar account, according to which moral status is continuously rising with some aspect of mental capacity, or (b) a multi-threshold account, according to which there are multiple thresholds of mental capacity making out differences in moral status, including at least one threshold that lies above the level of mental capacity possessed by ordinary human persons (McMahan 2009; Douglas 2011a; DeGrazia 2012; Agar 2012).[1]

A second response has been to argue that, even if human neuroenhancements could produce SEBs, this would not be bad for ordinary persons. For example, Buchanan (2009; 2011) argues that persons already possess enough moral status to be *inviolable* and that this protects them against being exploited in the ways that those concerned about the creation of SEBs fear. For example, because persons possess enough moral status to be inviolable, they could *not* permissibly be used, against their will, in medical experiments designed to aid SEBs. However, in reply, it has been argued that the inviolability enjoyed by ordinary persons is not absolute, so that SEBs could be more inviolable than us (Agar 2012; Douglas 2011a; McMahan 2009). Thus, for example, when faced with a choice between killing an ordinary person and killing an SEB in order to avert some catastrophe, it might be permissible to kill the person but not the SEB. It has also been argued that the existence of SEBs might be bad for mere persons in ways that are unrelated to inviolability. For example, SEBs might have stronger claims to scarce resources, such as healthcare, than ordinary persons (Douglas 2011a).

Finally, a third response to concerns regarding the creation of SEBs has been to argue that, even if human neuroenhancements could create SEBs *and* this would be bad for ordinary persons, it may still be morally permissible, or even morally desirable, to create such beings. For example, Douglas (2011a; 2012), Persson (2012) and Wasserman (2012) argue that costs

---

[1] Each of these authors relies to some extent on an inductive inference, most fully spelled out by Agar (2012): observed differences in mental capacity generate differences in moral status, and unobserved differences in mental capacity, between us and mentally enhanced beings, could be at least as great, thus we should expect that these too will give rise to differences in moral status. David Wasserman (2012) has recently offered an inductive argument militating in the opposite direction. He notes that the development of civilization, for example from feudalism to contemporary social arrangements, has tended to elide distinctions in moral status so that we now accept fewer tiers of moral status than previously. If we assume that this is due to moral-epistemic progress, then we might infer that we are in the process of converging on a correct view of moral status that accepts very few tiers of moral status (perhaps simply a binary view that entities either have moral status or not). Such a view might render implausible the claim that mentally enhanced beings could have greater moral status than us.

to ordinary persons associated with the existence of SEBs might be compensated by other advantages, or might be outweighed by goods enjoyed by the SEBs themselves. By contrast, Agar (2012) has argued that the costs of MSNs for the unenhanced are unlikely to be compensated by other benefits.

2. Moral Desirability Neuroenhancements

Moral desirability enhancements (MDNs) aim to and (expectably) succeed in augmenting the moral desirability of a person's character traits, motives or conduct (henceforth, collectively, 'psychological features'). The moral desirability of a psychological feature—often referred to simply as its moral goodness—corresponds to, or is at least correlated with, the degree to which there are typically agent-neutral moral reasons to promote that feature. It is to be distinguished, from the moral worth or moral praiseworthiness of a psychological feature (also sometimes referred to using the term 'moral goodness') which is instead standardly understood as the degree to which an agent merits praise for possessing that trait. Morally desirable traits are often also morally worthy, but they need not be.

Some psychological features may be *noninstrumentally* morally desirable. For example, Kant (1964) can be interpreted as arguing that a good will is noninstrumentally morally desirable; many have held that those character traits which qualify as virtues are noninstrumentally morally desirable; and many would also hold that respectful actions are, at least in one respect, noninstrumentally morally desirable. But character traits, motives and actions can also be *instrumentally* morally desirable, because they produce outcomes that are themselves morally desirable. For example, breaking a promise, though perhaps in one way morally undesirable, might also be in one way morally desirable when it results in lives being saved, pain being averted, or knowledge being gained. In what follows, moral desirability will be taken to include both instrumental and noninstrumental moral desirability. MDNs can thus be understood as interventions aim to and (expectably) succeed in increasing the total instrumental and noninstrumental value of a person's character traits, motives or conduct.

Note that, on some views, the same interventions will qualify as both moral status neuroenhancements and moral desirability neuroenhancements. For example, on some Kantian views, possession of a good will is both a morally desirable character trait and what confers moral status on a being. On this view, it seems possible that improving a person's will via neuroenhancement—supposing that this were possible—could constitute both a MSN and a MDN.

Types of Moral Desirability Neuroenhancement

MDNs are unified by the kind of moral value that they augment: moral desirability. However, as already noted, they may differ in their target, with some targeting a person's conduct, others her motives, and others still her character traits. MDNs that target a person's character or character traits could be aptly described as enhancements of moral *virtue* since, on some understandings of virtue, the moral virtues are often understood to be morally desirable character traits. However, moral virtue is also sometimes understood in other ways—for example, as referring to morally *praiseworthy* character traits.

MDNs can also be distinguished by the means they employ. As instances of neuroenhancement, all MDNs operate by directly modulating the brain states of an agent. However, we can distinguish them according to the way in which brain modulation is supposed to enhance the moral desirability of the agent's character, motives or conduct. Some MDNs might operate by directly altering an agent's mental *capacities or abilities* but without directly influencing any mental state. For example, Fröding (2011) suggests that certain cognitive neuroenhancements, for example, those which mitigate cognitive biases, would in some cases facilitate the development of moral virtues. Similarly, one can imagine that an intervention which enhanced an agent's ability to vividly imagine the consequences of her actions might contribute to morally more desirable conduct.

These interventions would alter mental states only indirectly, by influencing the agent's mental capacities. Other MDNs might operate by directly influencing an agent's mental states, for example, by directly altering the agent's beliefs, desires, intentions or emotions. Perhaps an intervention could enhance the moral desirability of an agent's conduct by augmenting her feelings of sympathy, her desire to help others or her belief in the importance of moral requirements. Within the category of MDNs that directly influence an agent's mental states we can distinguish between those that directly influence cognitive states (such as beliefs), conative states (such as desires and intentions) and affective states (such as emotions). As discussed below, some concerns that have been raised about MDNs are specific to MDNs that operate by directly altering certain kinds of mental state.

Scientific Prospects for Moral Desirability Neuroenhancement

Spence (2008) argues that within psychiatry, some medications, such as those used to treat antisocial personality disorder, substance abuse and psychosis, are already being used as MDNs in some circumstances. It might also plausibly be argued that anti-testosterone agents and other medications used to prevent sexual re-offending in some European and North American jurisdictions are intended to improve the moral desirability of the sex offender's conduct.

Most discussions of MDNs have, however, treated such enhancements as a prospect for the future rather than a contemporary phenomenon. To establish the plausibility of the claim that MDNs might become feasible or commonplace in the future, some have pointed to recent findings from moral psychology and neuroscience which are beginning to uncover the neural bases of morally significant character traits, motives and behavior, and in some cases, to suggest possible interventions capable of modifying these (e.g., Savulescu and Persson 2012). For example, the neurotransmitter oxytocin appears to have significant affects on trust and co-operativeness between people (e.g. Zak 2004; Kosfeld 2005; De Dreu 2010, 2011), and a number of widely used drugs, including the beta-blocker propranolol and a selective serotonin re-uptake inhibitors, a class of anti-depressants, appear to have significant effects on human moral psychology (e.g. Crockett 2008, 2010; Terbeck 2012). Shook (2012) and Levy et al. (forthcoming) review some recent empirical findings relevant to future prospects for MDN.

Motives for Discussing the Morality of Moral Desirability Neuroenhancements

There have been two main motives for assessing the morality of MDNs. Some have assessed hypothetical MDNs as a way to test arguments that have been made regarding the morality of neuroenhancement or bioenhancement more generally (CROSSREF: section introduction 'Neuroenhancement' and chapters 'Reflections on Neuroenhancements', 'Neuroscience, drug policy, and drug regulation, including enhancement use of pharmaceuticals', 'The Ethics of Mood Enhancement'). For example, Douglas (2008) presents certain kinds of MDN as counterexamples to the claim that it is always morally impermissible to biologically augment one's capacities from an already-normal level.

Other authors have discussed MDNs for more practical reasons: because they believe that there is a realistic prospect of MDNs being attempted in the future and wish to encourage, forestall or morally appraise such attempts or the technological developments that would enable them. For example, Ingmar Persson and Julian Savulescu have argued that MDNs might help humanity to escape its own destruction and should therefore be pursued, while John Harris, Robert Sparrow, Nick Agar and others have raised serious moral concerns regarding MDNs.

The following two sections outline some of the main moral positions that have been taken on MDNs and some of the dominant arguments that have been made for and against those positions. As in the discussion of MSNs, arguments concerning the *possibility* of MDNs are discussed only insofar as they have figured in arguments concerning their morality.

5

3. Persson and Savulescu's Defense of an Imperative to Pursue the Development of MDNs

The most forthright and provocative defence of MDNs has been offered by Ingmar Persson and Julian Savulescu, who argue that there is "an urgent imperative to enhance the moral character of humanity" and to pursue research into moral neuroenhancements as a possible means to this end (2008, p. 162; see also 2010, 2011a, 2011b, 2012). Persson and Savulescu argue that <u>scientific progress</u>, aided by <u>cognitive neuroenhancement</u>, is likely to bring it about that even small numbers of malevolent agents could cause great harm. For example, they suggest that scientific advances may allow even a single scientifically-trained individual to produce a designer pathogen capable of wiping out humanity, making it almost certain that humanity will indeed be destroyed. On the other hand, they argue, scientific progress will not have any benefits sufficiently great to make it rational to accept this risk of great harm, in part because it is generally easier to cause a harm than to cause a comparably large benefit, and in part because it is rational to be somewhat more averse to harms than attracted to benefits. They maintain that:

> It will be bad for us that scientific knowledge continues to grow by traditional means, and even worse if this growth is further accelerated by biomedical or genetic enhancement of our cognitive capacities. For if an increasing percentage of us acquires the power to destroy a large number of us, it is enough if very few of us are malevolent or vicious enough to use this power for all of us to run an unacceptable increase of the risk of death and disaster. To eliminate this risk, cognitive enhancement would have to be accompanied by a moral enhancement which extends to all of us, since such moral enhancement could reduce malevolence (2008, p. 166).

This moral enhancement, as they understand it, will consist in the acquisition of morally better (in the sense of more morally desirable) motivational dispositions. Persson and Savulescu argue that neither traditional forms of moral enhancement, such as moral education, nor cognitive enhancements (traditional or otherwise) are likely to produce the necessary moral enhancement. They also suggest that it is unlikely that the risks of great harm can be mitigated by holding back cognitive enhancement and scientific progress (2008, p. 173). Thus, they argue that we should pursue the possibility of moral neuroenhancements by, among other things, conducting scientific research that might facilitate their development (on this claim, see also Hughes 2006, 2011). This argument has attracted four main types of criticism.

Criticism 1: Misuse of Moral Enhancement Technologies

Persson and Savulescu are concerned that the products of technological development might be misused to devastating effect. Some authors have responded by noting that the moral enhancement technologies whose development they call for, or the scientific information that would need to be acquired in order to develop them, could themselves be misused by malevolent agents (Ehni & Aurenque p. 231-2; Fenton 2010; Harris 2011, p. 108; Sparrow, forthcoming, under review). Presumably, if we were technologically capable of bringing it about that people have *more* morally desirable character traits, motives or conduct, we could also bring it about that people have *less* desirable traits, motives or conduct, and some agents might well be more interested in decreasing moral desirability than increasing it. For example, we can imagine that a ruthless businessman or soldier might want to suppress pangs of conscience and that this might bring about morally undesirable motives and conduct.

Persson and Savulescu's response to these concerns has been to concede that technologies which enable MDNs could be misused, but maintain that pursuit of MDNs is nevertheless our only hope of avoiding ultimate harm (Persson and Savulescu 2011a, p. 443).

Criticism 2: Misconstrual of the Risks and Benefits of Scientific Progress

A second critical response has been to argue that Persson and Savulescu have overstated the risks or understated the benefits associated with further scientific progress. For example, Elizabeth Fenton maintains that, in characterising the likely benefits of scientific progress as consisting in "only a *small* increase in an *already high* quality of life", Persson and Savulescu both underestimate the size of the likely benefits of scientific progress and misconstrue the likely starting point (2010, p. 149, her italics). She argues that the benefits of scientific progress would in many cases be enjoyed by individuals whose quality of life is poor, for example, due to disease, famine or pollution, and for whom the benefits of such progress would be substantial. She also argues, appealing to Buchanan (2008), that further scientific progress might in fact be necessary to prevent significant falls in quality of life (p. 150).

John Harris (2011) also offers a response of this second variety. He disputes Persson and Savulescu's claim that it is generally easier to harm than to benefit others by appealing to cases in which it seems easy to cause large benefits by making donations, initiating public health programmes, providing vaccines or preventing others from inflicting great harm (pp 106-7). Suppose you know that someone is planning a mass shooting and have access to the gun they will use; you might then be able to save many lives by removing the bullets from this gun. Harris suggests that in this sort of case, it seems at least as easy for you to save the lives at stake as it is for the prospective shooter to end them. Harris has also argued that cognitive neuroenhancement is itself a variety of moral neuroenhancement (2011, e.g. pp. 106, 110). The implication is that, if future technological progress is driven in part by cognitive neuroenhancement, we can expect that future people will also have more morally desirable character traits, motives and conduct than us. This might somewhat mitigate the risk that new technologies will be used to harmful effect.

Persson and Savulescu offer a number of replies to claims that they have misconstrued the risk-benefit balance associated with further technological development, but perhaps the most important of these involves an appeal to the concept of *ultimate harm*: this is the harm of bringing it about that worthwhile life will never again exist. Persson and Savulescu argue that scientific progress will increase the risk of ultimate harm, and they take this to be an indefinitely bad outcome, since we have no way of knowing how much net value, in the form of continued worthwhile life, the ultimate harm prevents (e.g. 2011a, p. 442). On the other hand, they doubt that technological progress will similarly increase the likelihood of indefinitely good benefits.[2] Thus, they claim, we are in the position of weighing indefinitely bad harms against substantial but definite benefits, and faced with this choice, it is rational to prefer that there is not further technological progress.

Criticism 3: Implausible Implications

Persson and Savulescu have also been criticised for being committed to intuitively implausible views. For example, Fenton (2010) argues that they are committed to the view that "*all* forms of scientific progress are instrumentally bad for humans overall" (p. 149, her italics). Their response is to argue that there has been a turning point. Before that point, scientific progress could be expected to produce net benefits; afterwards, it could be expected to produce net harms. They suggest that this turning point occurred when "science and technology put in the hands of humans the means of destroying or seriously damaging forever the conditions of sentient life on this planet" and speculate that this point may have come in the mid-twentieth century (Persson and Savulescu 2011a, p. 441). Meanwhile, Harris (2011) argues that Persson and Savulescu are committed to the view that it would be in one way

---

[2] Persson and Savulescu (2011b, p. 4) do allow that one might contribute to an indefinitely large benefit by preventing someone else from causing ultimate harm, something that technological progress might allow. However, they argue that, in that sort of case, one would not be *guaranteeing* an indefinitely large benefit, because someone else (or a natural disaster) might cause ultimate harm. By contrast, when one causes ultimate harm, one does guarantee an indefinitely bad harm.

desirable to retard the cognitive powers of current and future humans, thus slowing scientific progress.

Criticism 4: Moral Neuroenhancement Won't Help

A fourth criticism of Persson and Savulescu's argument maintains that pursuit of moral enhancement would do little to diminish the risk of ultimate harm (Harris, 2011, pp. 109-10). Since the risks that worry Persson and Savulescu are risks that could be created by lone malevolent agents, and since those agents are also likely to be able to avoid undergoing moral enhancement, it seems doubtful whether pursuit of moral enhancement would do much to reduce these risks.

In response, Persson and Savulescu (2012) argue that technological progress can exacerbate risks of ultimate harm not only by allowing lone agents to wreak great havoc, but also by exacerbating the negative effects of widespread but mundane moral failings. For example, they argue that climate change may cause ultimate harm and that it can be attributed to a combination of technological development and human moral failings. Moral neuroenhancement might be capable of substantially reducing climate change even if not universally undergone. Thus, they suggest, MDNs might reduce some risks of ultimate harm even if they would fail to reduce others.

4. Other Defences of MDNs

Persson and Savulescu are interested in whether humanity falls under an imperative to pursue or promote the development of technologies that would enable MDNs. However, there are many other moral questions that we might ask regarding the morality of MDNs. For example, we might ask whether there is some weaker moral conclusion that could be made in favour of pursuing the development of MDNs: even if humanity falls under no imperative to promote MDNs, it might nevertheless be morally permissible, or even morally desirable, for it to do so. We could also move away from questions about what humanity as a whole ought to do to questions about what individual people, or groups of people, ought to do. Again, even if humanity falls under no imperative to promote MDNs, it might nevertheless be permissible, desirable or obligatory for individual people or groups of people to promote MDNs, to undergo MDNs themselves, or to encourage or require others to do so.

There is thus scope for weaker defences of MDN, and a number of such defences have been offered. For example, Thomas Douglas (2008) has argued that it would often be morally permissible for individuals to voluntarily engage in MDNs that expectably bring it about that they have more morally desirable motives than they would otherwise have had. More recently, David DeGrazia (forthcoming) has argued that moral desirability enhancements of motives or behavior would themselves be morally desirable, under certain idealizing assumptions. There have also been weaker defenses of specific means to MDN. For example, Douglas (2011b) argues that it could be morally permissible to enhance the moral desirability of one's motives or conduct *via directly modulating one's emotions*—that is, via means which, once set in train, modulate one's emotions without requiring the engagement of one's deliberative capacities.

It would be relatively uncontroversial that individuals have moral reasons to enhance the moral desirability of their character, motives and conduct, and that doing so is in one way morally desirable. These may even be analytic truths. Moreover, MDNs appear to be immune to many of the general concerns that have been raised about neuroenhancements (or bioenhancements more generally). These concerns have often focused on ways in which neuroenhancements undergone by some individuals might harm or wrong others, for example, by placing them at an unfair competitive disadvantage or undermining commitments to solidarity or equality. MDNs are unusual among the main types of neuroenhancements that have been discussed heavily in recent literature in that they might plausibly be expected to advantage, rather than disadvantage, others.

Nevertheless, some significant concerns have been raised regarding the permissibility and desirability of undergoing MDNs, or certain kinds of MDNs. Some of these are general concerns about enhancing the moral desirability of our characters, motives and conduct, whether through *neuro*enhancement or more traditional means. In this category are general skepticism about the possibility of moral improvement and concerns about whether we have adequate means for resolving disagreement and uncertainty about what character traits, motives and conduct are morally desirable and why. Debate regarding MDNs has elicited significant moral discussion on these points (Jotterand 2011; Schaefer 2011; Shook 2012; Wasserman 2011), but these concerns apply to traditional means of augmenting moral desirability as well as MDNs. Other concerns are general concerns about neuroenhancement that would apply to non-moral neuroenhancements as much as to MDNs. In this category are concerns regarding the unnatural means or hubristic motives that biomedical enhancement is said to involve (Kass 2003, Sandel 2007). Other concerns are, however, specific to MDNs—they would not apply equally to traditional means of enhancing moral desirability or to other kinds of neuroenhancement. The remainder of this section outlines three dominant concerns in this category as well as some responses that have been offered to them by those who wish to defend at least some varieties of MDN.

Concern 1: Restriction of Freedom

One concern that has been raised regarding MDNs, or certain MDNs, is that such enhancements might restrict <u>freedom</u> or <u>autonomy</u>. Thus, for example, John Harris (2011) argues that we might have reason to abstain from MDNs because they would restrict our freedom to perform morally undesirable actions or to have morally undesirable motives (see also Ehni and Aurenque 2012, p. 233 and, for a more general discussion of the effects of neuroenhancement on autonomy, Bublitz and Merkel 2009).

A similar thought underpins the well-known <u>free will defense of theism</u>. The free will defense maintains that the existence of an omnipotent, omniscient and benevolent God is consistent with the presence of evil because evil is a consequence of our possessing the freedom to do evil, which is, all things considered, good. Though the freedom to do evil possesses the great instrumental disvalue of allowing evildoing, it also possesses some other, greater value. It is no surprise, then, that an omnipotent, omniscient and benevolent God would have allowed evil to exist.

If the freedom to do evil is all-things-considered valuable despite its very great instrumental disvalue, one might expect that the more general freedom to have morally undesirable character traits or motives or to act in morally undesirable ways (henceforth, the freedom to be immoral) is also good despite its instrumental disvalue. Perhaps, then, MDNs will be all-things-considered disvaluable whenever they restrict this freedom.

Two main types of response have been made to this suggestion. The first has been to argue that, even where MDNs do restrict freedom, it might nevertheless be morally permissible or, all things considered, morally desirable to undergo such MDNs. The second has been to deny that all MDNs would restrict freedom, thus limiting the concern about freedom to a subset of MDNs.

Responses of the first type parallel a standard response to the free will defense of theism. That response holds that, in many cases, it seems preferable to sacrifice some freedom to do evil in order to prevent evil than to tolerate both the freedom to do evil and the associated evil. If I witness one person about to murder another, it seems that I should intervene to prevent the murder even though this involves restricting the prospective murder's freedom to do evil. Similarly, it seems that a would-be murderer should restrict his own freedom to do evil, thus preventing murder, if he is in a position to do so. The obvious way of explaining this conclusion is to suppose that, in at least some cases, any disvalue associated with restricting one's freedom to do evil is outweighed by the value of doing so. Similarly, it has been argued that there will be cases of MDN where the disvalue of any loss in freedom to be immoral is outweighed by the value of increasing the moral desirability of one's motives

or conduct (DeGrazia, forthcoming; Douglas 2008, 2011b; Savulescu et al., forthcoming; Savulescu and Persson 2012).

Responses in the second category sometimes begin by noting that concerns about the freedom-reducing effect of MDNs presuppose compatibilism—namely, the view that freedom is consistent with one's motives and conduct being causally determined. For suppose compatibilism were false, and we could be free only if we were causally undetermined. In that case, we would either already be completely unfree, because we are causally determined, in which case MDNs could not *reduce* our freedom, or we are free only because we are, at least some of the time, immune from causal forces, in which case MDNs, which operate through causal channels, could not affect us in circumstances where we are currently free (Blackford 2010; DeGrazia, forthcoming; Savulescu et al., forthcoming; Savulescu and Persson 2012).

Thus, it is argued, concerns about the freedom-reducing effect of MDNs will have to be grounded on compatibilism. But on compatibilist views of freedom, whether MDNs diminish freedom is likely to depend on precisely how they operate.

We can think of compatibilists as distinguishing between two aspects of our psychological life: a true or authentic self (which may be identified with higher order desires, or higher order desires that are not the result of inauthentic influences) and a brute self. Motives or actions are then thought of as free if and only if they are part of or flow from the true self.

On this picture, MDNs could reduce our freedom to be immoral if they operate by strengthening the influence of the brute self vis-à-vis the true self. Consider an individual who, whenever approached by poverty-relief charities, feels a conflict between, on the one hand, a firm commitment, formed through introspective reflection, to the view that she owes nothing to the poor, and on the other hand, feelings of sympathy towards the poor which are the result of extraneous social pressures and which she experiences as unwanted. Suppose further that this individual, again, under social pressure, undergoes an intervention that directly augments her feelings of sympathy. Such an intervention might perhaps qualify as a MDN, and it will also, on many compatibilist views, restrict her freedom. A proponent of these views would deem this agent's feelings of sympathy, both before and after the intervention, to be aspects of her brute self, while her moral commitments regarding charity are an aspect of her true self. The intervention might thus plausibly be construed as one in which the influence of the agent's brute self is expanded at the expense of the agent's true self. One might expect that, following the intervention, a greater proportion of this agent's motives and conduct would be unfree, where that implies that they are part of or flow from the brute, rather than the true, self.

However, defenders of MDNs have argued that, just as we can paint a compatibilist picture of a MDN that restricts the freedom to be immoral, so too we can paint a picture of one that operates by increasing the freedom to be moral. They argue that, on any plausible compatibilist account of freedom, there will be some ways in which the conduct of some people is already unfree—because it flows from aspects of our brute selves. MDNs could operate precisely by attenuating the influence of these aspects of our brute selves (Douglas 2008; 2011b; Savulescu et al., forthcoming). Thus, suppose that, when an agent is approached by worthy poverty-relief charities, she feels a pull between a reflectively formed commitment to the view that poverty-relief is morally required and a hoarding instinct not to part with any of one's property that is unwanted and the result of external factors. In this case, many compatibilist views will deem the moral commitment to be part of the true self, and the hoarding instinct to be part of the brute self. Thus, it will be plausible that an MDN that attenuated the hoarding instinct would increase the agent's freedom.

Concern 2: Superficiality

A second concern that has been raised regarding MDNs is targeted specifically at Douglas' defense of MDNs that increase the moral desirability of one's motives or conduct by directly modulating emotions (call such interventions 'Emotional MDNs'). John Harris worries that an

intervention which operates in this way "cannot be moral enhancement properly so called at all". It "is hardly an enhancement, and certainly not one that has much to do with morality" (2012a, p. 4). Indeed, he maintains that "the notion of moral behaviour has been attenuated to a vanishing point" once one claims that such behaviour could be produced by directly altering emotions (2012a, p. 6); "tinkering with the emotions is not a form of moral enhancement at all. It is more like the threat of punishment: it may make immoral behaviour less likely, but it does not enhance morality" (2012b, pp. 3-4; see also Harris and Chan 2010; Sparrow, forthcoming; Sparrow, under review).

One way of reading these passages would see them as asserting that Emotional MDNs are metaphysically, or at least physically, impossible. They violate some metaphysical or physical law. (Doubts regarding the possibility of MDNs more generally have also been raised by Ehni and Aurenque 2012, p. 232; Jotterand 2011; and Shook 2012.) However, it is difficult to see how this charge could be sustained. Emotions are mental states, mental states are normally taken to be either constitutively or causally dependent on brain states, and brain states are susceptible to direct modulation. It seems both metaphysically and physically possible to alter the emotions by directly influencing brain states. It also seems metaphysically and physically possible for alterations in one's emotions to alter the moral desirability of one's motives or conduct. For example, gratuitous killing of innocents in war is morally undesirable, and it is plausible that differences in the extent to which a solider experiences emotions such as a uncontrollable rage might influence whether or not that soldier engages in such morally undesirable conduct.

Perhaps a more plausible way of interpreting Harris' passages would see them as conceding that direct modulation of emotions could increase the moral desirability of one's motives or conduct, but maintaining that it could not produce some deeper kind of moral improvement for which the term 'moral enhancement' should be reserved. Fabrice Jotterand (2011) makes a similar point, arguing that "moral neuroenhancement is unlikely to morally enhance people in the true meaning of the word" (p. 8), as does Robert Sparrow (under review), who suggests that "while there is indeed evidence that certain pharmaceutical and neuro-scientific interventions can alter dispositions and behaviour in ways that we may be inclined to morally evaluate positively, this falls well short of constituting 'moral bioenhancement' in any interesting sense". These authors might be read as suggesting that the direct modulation of emotions could not produce *deep* moral improvement, where that might consist, for example, in an increase in the moral *worth* or *virtue* of one's motives or conduct.

It would not, of course, follow from this that Emotional MDNs are absolutely morally undesirable in any way. To claim that Emotional MDNs fail to produce deep moral improvement is to claim that they lack some desirable feature, not that they possess some absolutely undesirable one. However, if Emotional MDNs fail to produce deep moral improvement, it might follow that they are *less* morally desirable than certain other ways of enhancing the moral desirability of our conduct that do produce such improvement. Harris (2011, 2012a, 2012b) can be read as maintaining that, while Emotional MDNs fail to produce deep moral improvement, more traditional, deliberative means of enhancing moral desirability, such as introspective reflection and engagement with literature, do produce deep moral improvement and should thus generally be adopted in preference to MDNs.

Harris buttresses his claim that Emotional MDNs could not produce deep moral improvement by drawing a parallel between 'being moral', in the deeper sense in which he is interested, and 'being scientific':

> [o]ne can accidentally discover something of scientific importance, but one cannot be scientific, one cannot do science, accidentally. Doing science is a deliberative and disciplined process. It involves, for example, doing things like formulating and testing a hypothesis and looking for disconfirmatory evidence as well as for confirmatory evidence . . . Being moral is like being scientific (Harris 2012a, p. 6).

He suggests that, unlike traditional means of enhancing moral desirability, Emotional MDNs could bring about morally desirable motives and conduct only by accident. One might argue

for this view by noting that traditional means of enhancing moral desirability frequently operate by enhancing an agent's moral understanding: her understanding of what morality requires and why, both in general, and on particular occasions on which one must make a morally significant decision. This moral understanding is an all-purpose tool that helps the agent to be motivated and act in morally desirable ways in many circumstances. If an agent understands what morality requires and why, she will be well-placed to be motivated and act in morally desirable ways in almost any circumstance.

On the other hand, Harris (2012a, 2012b) suggests, Emotional MDNs would not operate by enhancing the agent's moral understanding. Others have made similar claims. For example, Fabrice Jotterand (2011) argues that "[w]hile the manipulation of moral emotions might change the behavior of an individual, it does not provide any content, for example, norms or values to guide one's behavioral response" (p. 6, see also p. 8). Similarly, Robert Sparrow (forthcoming) suggests that "It is hard to see how any drug could alter our beliefs in such a way as to track the reasons we have to act morally" and that "[s]omeone who reads Tolstoy arguably learns to be less judgemental and in doing so develops greater understanding: someone who takes a pill has merely caused their sentiments to alter". If these authors are correct, Emotional MDNs (and perhaps MDNs more generally) do not augment moral understanding. Instead, it might be argued, Emotional MDNs would typically work by removing some relatively straightforward emotional barrier to morally desirable motivation or conduct. The most obvious examples of such obstacles might include aggressive impulses, strongly xenophobic sentiments or a callousness or ruthlessness towards the suffering of others. But note that these are not *universal* barriers to moral desirable motivation and conduct. For example, one can imagine circumstances in which aggressive impulses or callousness would produce morally desirable motives and conduct; aggressiveness might do so when one is fighting a just war, or perhaps when one is confronted with one person assaulting another on the street, and ruthlessness might do so when one is a politician engaged in diplomatic negotiations (Douglas 2008; Savulescu et al., forthcoming; Wasserman 2011, forthcoming). Thus it might be thought that Emotional MDNs would produce only accidental or unreliable morally desirable motives and conduct.

There is little literature responding to Harris' concern about the superficiality of Emotional MDNs, nor to the suggestion that Emotional MDNs could produce no more than accidental morally desirable motives and conduct. However, some defenders of MDN have suggested that such neuroenhancements could operate by attenuating emotional barriers to sound moral deliberation (Douglas 2008) or by bringing it about that we are more like those existing people who we regard as especially moral (Savulescu and Persson 2012; Savulescu et al., forthcoming), and one might expect that where they operate in this way they will, like more traditional forms of moral enhancement, increase moral understanding. In addition, Wasserman (2011) has argued that, even if an Emotional MDN initially had no positive effect on moral understanding, we might expect the frequent presence of pro-moral emotions and the morally desirable motives or conduct to lead to a development in moral understanding over time. This parallels the Aristotelian point that one comes to know the good by being good (Burnyeat 1980).

Concern 3: Misfiring

A third concern that can be raised regarding MDNs maintains that attempts at MDNs are likely to misfire, bringing about moral deteriorations rather than improvements. This is not a concern about successful MDNs, but is rather the concern that actual attempts at MDN are likely to be unsuccessful.

Harris (2011) advances this concern by noting that "the sorts of traits or dispositions that seem to lead to wickedness or immorality are also the very same ones required not only for virtue but for any sort of moral life at all" (p. 104). He infers from this that the sorts of psychological alterations required for MDN would involve not the wholesale elimination or dramatic amplification of particular dispositions, but rather a kind of fine-tuning of our

dispositions (see also Jotterand 2011, p. 7; Wasserman 2011). However, he argues that the disposition-modifying neurotechnologies that we are actually likely to have available to us will be rather blunt: it will be difficult to determine or even predict their effects precisely or to target the specific dispositions that we wish to shape. Thus, he suggests, it is unlikely that any attempt to use these technologies will succeed in bringing about an improvement in the moral desirability of our motives or conduct.

Nicholas Agar (2010; forthcoming; under review) sets forward a more limited variant of the concern. He argues that attempted MDNs may have good chances of success when they aim only to correct subnormal levels of moral desirability, bringing an individual within the normal range, but that they are likely to misfire when they aim to produce levels of moral desirability above the normal range (he does not comment on MDNs that operate wholly within the normal range). Subnormal moral desirability is often the result of relatively isolated and easily identified defects such as, for example, the deficient empathy that characterizes psychopathy. Agar speculates that these defects could relatively safely be corrected. However, he argues that, to attain supra-normal levels of moral desirability, we would need to simultaneously augment or attenuate several different dispositions in a balanced way. This, he claims, will be very difficult. One might suppose that this difficulty derives from three sources. First, there may be moral uncertainty—uncertainty regarding what character traits, motives or conduct qualify as supra-normally morally desirable. There would be significant disagreement on this matter, and that this disagreement may be both evidence for and a source of uncertainty (see e.g. Schaefer 2011). Second, there may be empirical uncertainty about what changes would need to be wrought in a given individual to realize the putatively desirable traits. Third, there might, for reasons suggested by Harris (2011), be practical difficulties in realizing these transformations using technologies that we have available to us. Taken together, these difficulties might create a serious risk that attempts to bring about supra-normal MDNs will fail.

The main response made to these concerns by defenders of MDN has been concessive. Defenders of MDNs have acknowledged both that (1) in many cases, complex and subtle interventions would be needed in order to enhance moral desirability and that (2) this creates a risk that attempted MDNs will fail, perhaps instead resulting in moral deterioration (Douglas 2011b; Savulescu et al., forthcoming). However, some doubt has been cast on the thought that this concern would always count decisively against attempting MDNs. For example, Douglas (2011b) notes that there are other areas—such as clinical psychiatry— where we often also use rather blunt biological interventions as part of efforts to achieve subtle and multidimensional psychological changes. Yet in that area we normally think that attempting some interventions can be permissible and desirable if undergone cautiously, keeping open the option of reversing the intervention if it misfires. He suggests a similar approach might be justified in relation to MDNs.

5. Further Questions

Ethical debate on MDNs has focused primarily on (1) the ethical status of promoting the development of technologies that would enable MDNs, as in the case of discussion surrounding Persson and Savulescu's proposals, and on (2) the ethical status of voluntarily undergoing MDNs oneself, as in the case of discussion surrounding Douglas' and DeGrazia's proposals. Further questions might be asked regarding the ethical status of requiring or encouraging others to undergo such neuroenhancements. For example, one might wonder whether it would ever be permissible for a state to impose MDNs on its citizens, for courts to impose MDNs on convicted offenders, for parents to impose MDNs on their children, or for employers to impose MDNs on their employees. Such uses of MDNs would clearly raise, in addition to the concerns discussed above, concerns regarding, *inter alia*, coercion, manipulation and domination. For example, Sparrow suggests that imposing moral enhancements on others might, even if benevolently motivated, constitute an objectionable

form of domination (forthcoming; under review; see also Bublitz and Merkel 2009). Meanwhile Walker (2009) and Blackford (2010) have responded to similar concerns as they might be raised in relation to the imposition of MDNs on one's children or children-to-be.

Perhaps the context in which coercive use of MDNs is most likely to gain acceptance is that of <u>criminal justice</u>. Arguably, institutions of criminal justice, are, at least insofar as they offer criminal rehabilitation programs, already engaged in a kind of moral enhancement. In addition, concerns about moral disagreement, might be regarded as least serious in this context; criminal conduct is conduct that, if not universally taken to be morally undesirable, is at least widely accepted as conduct which the state and citizenry may legitimately seek to prevent. Concerns regarding coercion might also be regarded as weaker in the context of criminal justice, since a degree of coercion is also already accepted in that area: the paradigmatic criminal sanction, incarceration, is itself highly coercive. Nevertheless, it might be argued that forced MDNs in criminal justice would be more problematic than either incarceration or psychological rehabilitation programs, perhaps because more intrusive or manipulative. There is a burgeoning literature on this topic (Rosati 1994; Bomann-Larsen 2011; Ryberg and Petersen 2011; Vincent 2012; Shaw 2012; Bublitz and Merkel 2012).

Questions might also be asked regarding the morality of selecting for morally desirable character traits in one's children, for example, on the basis of genetic information acquired through <u>pre-implantation genetic diagnosis</u>. Moral enhancement and moral neuroenhancement have been understood in this chapter to involve augmenting the morality of some particular individual or her traits. On this understanding, selecting for moral traits in one's children would not qualify as moral neuroenhancement, or indeed moral enhancement, since it instead involves bringing one (expectably) more moral individual into existence in preference to another (expectably) less moral individual. However, some of the same issues arise in cases of selection as in cases of neuroenhancement discussed here, and there is some literature exploring these. For example, Faust (2008) argues that if parents were able to select for morally desirable traits in their future children it would be desirable, and probably obligatory, for them to do so, and Walker (2009) argues that principled reason not to select for moral virtues in one's children, unless there is a principled reason not to engage in selection at all.[3] For criticism of these claims, see Agar (2010), Andreadis (2010), Arnhart (2010) and Bronstein (2010).

References

Agar, N. (2010). Enhancing Genetic Virtue? *Politics and the Life Sciences* 29(1), 73–75. doi:10.2990/29_1_73.

Agar, N. (2012). Why Is It Possible to Enhance Moral Status and Why Doing so Is Wrong? *Journal of Medical Ethics*, forthcoming in print. doi:10.1136/medethics-2012-100597. http://jme.bmj.com/content/early/2012/08/23/medethics-2012-100597.short.

Agar, N. (Forthcoming). A Question About Defining Moral Bioenhancement. *Journal of Medical Ethics.*

Agar, N. (Under review). Against Moral Bioenhancement.

Andreadis, A. (2010). The Tempting Illusion of Genetic Virtue. *Politics and the Life Sciences* 29(1), 76–78. doi:10.2990/29_1_76.

Arnhart, L. (2010). Can Virtue Be Genetically Engineered? *Politics and the Life Sciences* 29(1), 79–81. doi:10.2990/29_1_79.

---

[3] Elster (2011) and Douglas and Devolder (forthcoming) have also defended ethical principles which imply that parents would typically have significant moral reasons to engage in such selection.

Blackford, R. (2010). Genetically Engineered People. *Politics and the Life Sciences* 29(1), 82–84. doi:10.2990/29_1_82.

Bomann-Larsen, L. (2011). Voluntary Rehabilitation? On Neurotechnological Behavioural Treatment, Valid Consent and (In)appropriate Offers. *Neuroethics*, forthcoming in print. doi:10.1007/s12152-011-9105-9. http://www.springerlink.com/content/314k2722666347j5/.

Bronstein, J. (2010). Objecting to the Genetic Virtue Program. *Politics and the Life Sciences* 29(1), 85–87. doi:10.2990/29_1_85.

Bublitz, J. C., & Merkel, R. (2009). Autonomy and Authenticity of Enhanced Personality Traits. *Bioethics* 23(6), 360–374. doi:10.1111/j.1467-8519.2009.01725.x.

Bublitz, J. C., & Merkel, R. (2012). Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. *Criminal Law and Philosophy*, 1-27, forthcoming in print. doi:10.1007/s11572-012-9172-y.

Buchanan, A. (2008). Enhancement and the Ethics of Development. *Kennedy Institute of Ethics Journal* 18(1), 1–34.

Buchanan, A. (2009). Moral Status and Human Enhancement. *Philosophy & Public Affairs* 37(4), 346–381.

Buchanan, A. (2011). *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford: Oxford University Press.

Buchanan, A. (2012). Still Unconvinced, but Still Tentative: A Reply to DeGrazia. *Journal of Medical Ethics* 38(3): 140–141.

Burnyeat, M. F. (1980). Aristotle on Learning to Be Good. In *Essays on Aristotle's Ethics*, ed. A. O. Rorty, 69–92. Berkeley, CA: University of California Press.

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., and Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science* 320, 1739.

Crockett, M. J., Clark, L., Hauser, M., Robbins, T. W. 2010. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc Natl Acad Sci* 107, 17433–8.

Darwall, S. L. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., Ten Velden, F. S., Van Dijk, E., & Feith, S. W. W. (2010). The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans. *Science* 328(5984), 1408 –1411. doi:10.1126/science.1189047.

De Dreu, C. K. W., Greer, L. L., Van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. J. (2011). Oxytocin Promotes Human Ethnocentrism. *Proceedings of the National Academy of Sciences* 108(4), 1262 –1266. doi:10.1073/pnas.1015316108.

DeGrazia, D. (2012). Genetic Enhancement, Post-Persons and Moral Status: A Reply to

Buchanan. *Journal of Medical Ethics* 38(3), 135–139. doi:10.1136/medethics-2011-100126.

DeGrazia, D. (Forthcoming). Moral Enhancement, Freedom, and What We (Should) Value in Moral Behavior. *Journal of Medical Ethics*.

Douglas, T. (2008). Moral Enhancement. *Journal of Applied Philosophy* 25(3), 228–245.

Douglas, T. (2011a). Human Enhancement and Supra-personal Moral Status. *Philosophical Studies*, forthcoming in print. doi:10.1007/s11098-011-9778-2. http://www.springerlink.com/index/10.1007/s11098-011-9778-2.

Douglas, T. (2011b). Moral Enhancement via Direct Emotion Modulation: A Reply to John Harris. *Bioethics*, forthcoming in print. doi:10.1111/j.1467-8519.2011.01919.x.

Douglas, T. (2012). The Harms of Status Enhancement Could Be Compensated or Outweighed: a Response to Agar. *Journal of Medical Ethics*, forthcoming in print. doi:10.1136/medethics-2012-100835. http://jme.bmj.com/content/early/2012/10/08/medethics-2012-100835.

Douglas, T., & Devolder, K. (Forthcoming). Procreative Altruism: Beyond Individualism in Reproductive Selection. *Journal of Medicine and Philosophy*.

Ehni, H-J, & Aurenque, D. (2012). On Moral Enhancement from a Habermasian Perspective. *Cambridge Quarterly of Healthcare Ethics* 21(02), 223–234. doi:10.1017/S0963180111000727.

Elster, J. (2011). Procreative Beneficence—Cui Bono? *Bioethics* 25(9), 482–488. doi:10.1111/j.1467-8519.2009.01794.x.

Faust, H. S. (2008). Should We Select for Genetic Moral Enhancement? A Thought Experiment Using the MoralKinder (MK+) Haplotype. *Theoretical Medicine and Bioethics* 29(6), 397–416.

Fenton, E. (2010). The Perils of Failing to Enhance: A Response to Persson and Savulescu. *Journal of Medical Ethics* 36:3, 148–151. doi:10.1136/jme.2009.033597.

Fröding, B. E. E. (2011). Cognitive Enhancement, Virtue Ethics and the Good Life. *Neuroethics* 4(3), 223-234. doi:10.1007/s12152-010-9092-2.

Fukuyama, F. (2004). Transhumanism. *Foreign Policy* 144 (September), 42–43.

Harris, J. (2011). Moral Enhancement and Freedom. *Bioethics* 25(2), 102–111. doi:10.1111/j.1467-8519.2010.01854.x.

Harris, J. (2012a). What it's like to be good. *Cambridge Quarterly of Healthcare Ethics*, forthcoming in print. doi: 10.1017/S0963180111000867.

Harris, J. (2012b). 'Ethics is for bad guys!' Putting the 'moral' into moral enhancement. *Bioethics*, forthcoming in print. doi: 10.1111/j.1467-8519.2011.01946.x.

Harris, J., & Chan, S. (2010). Moral behavior is not what it seems. *Proceedings of the National Academy of Sciences* 107.50, E183.

Hughes, J. (2006). Virtue Engineering: Applications of Neurotechnology to Improve Moral Behavior, presented at *Transvision*, Helsinki. http://ieet.org/index.php/IEET/more/hughes20071120.

Hughes, J. (2011). After Happiness, Cyborg Virtue. *Free Inquiry* 32(1).

Jotterand, F. (2011). 'Virtue Engineering' and Moral Agency: Will Post-Humans Still Need the Virtues? *AJOB Neuroscience* 2(4), 3–9. doi:10.1080/21507740.2011.611124.

Kant, I. (1964). *Groundwork of the Metaphysic of Morals*. Trans. H. J. Paton. 1st Harper torchbook ed. New York: Harper & Row.

Kass, L. R. (2003). Ageless Bodies, Happy Souls. *The New Atlantis* 1, 9–28.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin Increases Trust in Humans. *Nature* 435(2), 673–676.

Levy, N., Douglas, T., Kahane, G., Terbeck, S., Cowen, P. J., Hewstone, M., & Savulescu, J. (Forthcoming). Are You Morally Modified? The Moral Effects of Widely Used Pharmaceuticals. *Philosophy, Psychiatry, & Psychology.*

McMahan, J. (2009). Cognitive Disability and Cognitive Enhancement. *Metaphilosophy* 40(3-4), 582–605. doi:10.1111/j.1467-9973.2009.01612.x.

Persson, I. (2012). Is Agar Biased Against 'post-persons'? *Journal of Medical Ethics*. doi:10.1136/medethics-2012-100836. http://jme.bmj.com/content/early/2012/10/08/medethics-2012-100836.

Persson, I., & Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy* 25(3), 162–177.

Persson, I., & Savulescu, J. (2010). Moral Transhumanism. *Journal of Medicine and Philosophy* 35(6), 656 –669. doi:10.1093/jmp/jhq052.

Persson, I., & Savulescu, J. (2011a). The Turn for Ultimate Harm: a Reply to Fenton. *Journal of Medical Ethics* 37(7), 441-444. doi:10.1136/jme.2010.036962.

Persson, I., & Savulescu, J. (2011b). Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement. *Bioethics*, forthcoming in print. doi:10.1111/j.1467-8519.2011.01907.x. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8519.2011.01907.x/abstract.

Persson, I., & Savulescu, J. (2012). *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.

Rawls, J. (2001). *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

Rosati, C. S. (1994). A Study of Internal Punishment. *Wisconsin Law Review* 1994, 123–1579.

Ryberg, J., & Petersen, T. S. (2011). Neurotechnological Behavioural Treatment of Criminal Offenders—A Comment on Bomann-Larsen. *Neuroethics* 1–5. doi:10.1007/s12152-

011-9146-0.

Sandel, M. (2007). *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge, MA: Harvard University Press.

Savulescu, J. (2009). The Human Prejudice and the Moral Status of Enhanced Beings: What Do We Owe the Gods. In *Human Enhancement*, ed. J. Savulescu and N. Bostrom. Oxford: Oxford University Press.

Savulescu, J., & Persson, I. (2012). Moral Enhancement, Freedom, and the God Machine. *Monist* 95(3), 399–421.

Savulescu, J., Douglas, T., & Persson, I. (Forthcoming). Autonomy and the Ethics of Biological Behaviour Modification. In *Towards Bioethics in 2050: International Dialogues*.

Schaefer, G. O. (2011). What Is the Goal of Moral Engineering? *AJOB Neuroscience* 2(4), 10–11. doi:10.1080/21507740.2011.620593.

Shaw, E. (2012). Direct Brain Interventions and Responsibility Enhancement. *Criminal Law and Philosophy* 1–20, forthcoming in print. doi:10.1007/s11572-012-9152-2.

Shook, J. R. (2012). Neuroethics and the Possible Types of Moral Enhancement. *AJOB Neuroscience* 3(4), 3–14. doi:10.1080/21507740.2012.712602.

Spence, S. A. (2008). Can Pharmacology Help Enhance Human Morality? *The British Journal of Psychiatry: The Journal of Mental Science* 193, 179–180. doi:193/3/179.

Sparrow, R. (Forthcoming). (Im)Moral Technology? Thought Experiments and the Future of 'Mind Control'. In *Towards Bioethics in 2050: International Dialogues*, ed. Akira Akabayashi.

Sparrow, R. (Under review). Better Living Through Chemistry? A Reply to Savulescu and Persson on 'Moral Enhancement'.

Terbeck, S., Kahane, G. McTavish, S., Savulescu, J., Cowen, P., Hewstone, M. (2012). Beta-Adrenergic Blockade Reduces Implicit Negative Racial Bias. *Psychopharmacology*, 222(3), 419-424.

Vincent, N. (2012). Restoring Responsibility: Promoting Justice, Therapy and Reform Through Direct Brain Interventions. *Criminal Law and Philosophy* 1–22, forthcoming in print. doi:10.1007/s11572-012-9156-y.

Walker, M. (2009). Enhancing Genetic Virtue. *Politics and the Life Sciences* 28(2), 27–47. doi:10.2990/28_2_27.

Wasserman, D. (2011). Moral Betterness and Moral Enhancement, presented at the Uehiro-Carnegie Conference 2011, New York.

Wasserman, D. (2012). Devoured by Our Own Children: The Possibility and Peril of Moral Status Enhancement. *Journal of Medical Ethics*, forthcoming in print. doi:10.1136/medethics-2012-100843.
http://jme.bmj.com/content/early/2012/09/27/medethics-2012-100843.

Wasserman, D. (Forthcoming). When Bad People Do Good Things: Will Moral Enhancement Make the World a Better Place? *Journal of Medical Ethics*.

Wikler, D. I. (2009). Paternalism in the Age of Cognitive Enhancement: Do Civil Liberties Presuppose Roughly Equal Mental Ability. In *Human Enhancement*, ed. Julian Savulescu and Nick Bostrom, 341–55. Oxford: Oxford University Press.

Wilson, J. (2007). Transhumanism and Moral Equality. *Bioethics* 21(8), 419–425.

Wilson, J. (2012). Persons, Post-Persons and Thresholds. *Journal of Medical Ethics* 38(3), 143–144. doi:10.1136/medethics-2011-100243.

Zak, P.J., Kurzban, R., & Matzner, W. T. (2004). The Neurobiology of Trust. *Annals of the New York Academy of Sciences* 1032, 224–227.

Cross-references

Neuroenhancement

Reflections on Neuroenhancements

Neuroscience, drug policy, and drug regulation, including enhancement use of pharmaceuticals

The Ethics of Mood Enhancement

Keywords

Moral enhancement

Moral neuroenhancement

Cognitive enhancement

Cogntive neuroenhancement

Moral status

Moral desirability

Moral worth

Moral virtue

Genetic selection

Freedom