Moral neuroenhancement

Brian D. Earp, Thomas Douglas, and Julian Savulescu

University of Oxford

Authors' personal copy. Forthcoming book chapter (accepted manuscript). Please cite as:

Earp, B. D., Douglas, T., & Savulescu, J. (forthcoming). Moral neuroenhancement. In S. Johnson & K. Rommelfanger (eds.), *Routledge Handbook of Neuroethics*. New York: Routledge. Accepted manuscript online ahead of print at: https://www.academia.edu/27484573/Moral_neuroenhancement.

* Note: this is the accepted version of the chapter; further minor revisions may still be made prior to final publication. Feedback at this stage is welcome.

Abstract

In this chapter, we introduce the notion of "moral neuroenhancement," offering a novel definition as well as spelling out three conditions under which we expect that such neuroenhancement would be most likely to be permissible (or even desirable). Furthermore, we draw a distinction between first-order moral capacities, which we suggest are less promising targets for neurointervention, and second-order moral capacities, which we suggest are more promising. We conclude by discussing concerns that moral neuroenhancement might restrict freedom or otherwise "misfire," and argue that these concerns are not as damning as they may seem at first.

Key words: moral enhancement, neuroenhancement, biotechnology, moral capacities, ethics, autonomy, freedom

Introduction

In recent years, a number of philosophers, neuroethicists, and others have become preoccupied with "moral enhancement." Very roughly, this refers to the deliberate moral improvement of an individual's character, motives, or behavior. In one sense, such enhancement could be seen as "nothing new at all" (Wiseman 2016, p. 4) or as something philosophically mundane: as G. Owen Schaefer (2015) has stated, "Moral enhancement is an ostensibly laudable project … Who wouldn't want people to become more moral?" (p. 261). To be sure, humans have long sought to morally enhance themselves (and their children) through such largely uncontroversial means as moral education, meditation and other similar practices, engagement with moral ideas in literature, philosophy, and religion, and discussion of moral controversies with others. What is different about the recent debate is that it focuses on a new set of potential tools for fostering such enhancement, which might broadly be described as "neurotechnologies." These technologies, assuming that they worked, would work by altering certain brain states or neural functions directly, in such a way as to bring about the desired moral improvement.

What exactly this would look like, and the mechanisms involved, are unclear. As John Shook (2012, p. 6) notes: "there is no unified cognitive system responsible for the formation and enaction of moral judgments, because separable factors are more heavily utilized for some kinds of moral judgments rather than others." Moreover, "the roles of emotions in moral appreciation and judgment, alongside (and intertwining with) social cognition and deliberate reasoning, are so complex that research is only starting to trace how they influence kinds of intuitive judgment and moral conduct."

Nevertheless, suggestions in the literature for possible means of pursuing moral enhancement by way of direct modulation of brain-level targets—at least in certain individuals, under certain circumstances or conditions—range from the exogenous administration of neurohormones such as oxytocin (in combination with appropriate psychological therapy or social modification) to potentially increase "pro-social attitudes, like trust, sympathy and generosity" (Savulescu & Persson 2012, p. 402; see also Donaldson & Young 2008; but see Bartz et al. 2011; Lane et al. 2015, 2016; Nave, Camerer, & McCullough 2015; Wudarczyk et al. 2013), to the alteration of serotonin or testosterone levels to mitigate undue aggression while at the same time ostensibly enhancing fair-mindedness, willingness to cooperate, and aversion to harming others (e.g., Crockett 2014;

Montoya et al. 2012; Savulescu & Persson 2012; but see Wiseman 2014 re: serotonin), to the application of newly developed brain modulation techniques, such as non-invasive (but see Davis & Koningsbruggen 2013) transcranial electric or magnetic stimulation, or even deep brain stimulation via implanted electrodes (for scientific overviews, see, e.g., Fregni & Pascual-Leone 2007; Perlmutter & Mink 2005; for ethical overviews, see, e.g., Clausen 2010; Hamilton, Messing, & Chatterjee 2011; Maslen et al. 2014; Rabin et al. 2009; Synofzik & Schlaepfer 2008).

Potential uses of brain stimulation devices for moral enhancement include attempts to reduce impulsive tendencies in psychopaths (Glenn & Raine 2008; but see Maibom 2014), as well as efforts to treat addiction and improve self-control, thereby making associated "immoral behavior" less likely (Savulescu & Persson 2012, p. 402). In addition, some research has shown that disruptive stimulation of the right prefrontal cortex or the temporoparietal junction can affect moral judgments directly—for example, judgments relating to fairness and harm (Knoch et al. 2015; Young et al. 2010); however, the circumstances of these and other similar investigations have been thus far largely contrived, such that the real-world implications of the findings are not yet apparent (Wiseman 2016). More ecologically valid results pertain to the administration of drugs such as methylphenidate or lithium to violent criminals with ADHD or to children with conduct disorder to reduce aggressive behavioral tendencies (see, e.g., Ginsberg et al. 2013, 2015; Ipster & Stein 2007; Margari et al. 2014; Turgay 2009), as well as antilibidinal agents to reduce sexual desire in convicted sex offenders (Douglas et al. 2013; Lösel & Schumucker 2005; Thibaut et al. 2010). Such measures remain controversial, however, both ethically (Craig 2016; Earp, Sandberg, & Savulescu 2014; Gupta 2012; Singh 2008) and conceptually, i.e., in terms of their status as moral enhancers as opposed to mere forms of "behavioral control" (see Focquaert & Schermer 2015; see also McMillan 2014).

To date, the majority of the philosophical literature on moral enhancement has been oriented around two major strands of thought: (1) Ingmar Persson and Julian Savulescu's argument 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, 2011, 2012, 2013, 2014), and (2) Thomas Douglas's and David DeGrazia's arguments that it would sometimes be morally permissible (in Douglas's case) or morally desirable (in DeGrazia's case) for individuals to voluntarily pursue moral neuroenhancements of certain kinds (e.g., Douglas 2008; DeGrazia 2014).

Both strands of thought have been subjected to vigorous criticism (for an overview, see Douglas 2015; see also Parens 2013). For their part, Persson and Savulescu have primarily been interested in whether humanity falls under an imperative to pursue or promote the development of technologies that would enable moral neuroenhancement on some description. However, even if there is such an imperative, it might turn out that it would be morally impermissible to deploy any of the technologies that would be developed. On the other hand, even if there is no imperative to pursue such technologies, it might be morally permissible, or even morally desirable (or obligatory), for people to use some moral neuroenhancers that are nevertheless developed. Thus, there is a further question regarding the moral status of engaging in (as opposed to developing the technologies for) moral neuroenhancement, and it is this question to which we will confine ourselves in the latter part of this chapter. First, however, it is important to clarify what we mean by the term "moral neuroenhancement," and to show that such a thing could ever be possible. We will start by laying out some definitions.

What is moral (neuro)enhancement?

In her wide-ranging essay, "Moral Enhancement: What Is It and Do We Want It?," Anna Pacholczyk (2011) outlines three major ways of understanding the term "moral enhancement," two of which we will consider here. The first is: a moral enhancement is a *change in some aspect of a person's morality that results in a morally better person* (p. 251, paraphrased). This is broadly the sense we have in mind for this chapter, but it is not quite precise, nor is it sufficiently focused, for our purposes, on enhancements that work "directly" on the brain—i.e., moral *neuro*enhancements in particular. We therefore propose an alternative definition:

> **Moral neuroenhancement:** Any change in a moral agent—effected or facilitated in some significant way by the application of a neurotechnology—that results, or is reasonably expected to result, in the agent being a morally better (i.e., more moral) agent.

Let us call this the *agential* conception of moral neuroenhancement. Note that the moral "betterness" of an agent could be understood in various ways. For example, it could be taken to be the increased moral worth or praiseworthiness of the agent, the increased moral

excellence of the agent, or the increased moral desirability of the agent's character traits, taken together (see Douglas 2015 for further discussion). But however it is plausibly understood, as Pacholczyk notes, being moral (let alone more moral) is "a complex ability and there is a wide range of potentially enhancing interventions. Making morally better people could include making people more likely to act on their moral beliefs, improving their reflective and reasoning abilities as applied to moral issues, increasing their ability to be compassionate, and so on" (2011, p. 253). Of course, there are likely to be serious and substantive disagreements about what should or should not be included on this list, as well as what should or should not be counted as "morally better" in the first place. This is an important issue to which we will return throughout this chapter.

The second major sense of "moral enhancement" discussed by Pacholczyk is this: a moral enhancement is *a beneficial change in moral functioning* (p. 251, paraphrased). Here the idea is, first, to identify an underlying psychological or neurological function that is involved in moral reasoning, decision-making, acting, etc. (that is what makes the function "moral," a descriptive claim), and then to intervene in it "beneficially" (a normative claim). But "beneficially" could mean different things, depending on one's normative perspective, and also on what is to be benefitted or improved. Is it the agent? Her moral character? Her well-being? The function itself? The world? Pacholczyk explores several possibilities but does not settle on a single answer.

We will focus on "the function itself." In so doing, we will draw on what two of us have dubbed *the functional-augmentative approach to enhancement*, often encountered in the wider bioenhancement literature (Earp, Sandberg, Kahane, & Savulescu 2014; see also Savulescu, Sandberg, & Kahane 2011). According to this more general approach, "Interventions are considered enhancements … insofar as they [augment] some capacity or function (such as cognition, vision, hearing, alertness) *by increasing the ability of the function to do what it normally does*" (Earp, Sandberg, Kahane, & Savulescu, 2014, p. 2, emphasis added).

This way of understanding "enhancement" will serve as the foil to our preferred approach (the agential approach), so we will spell it out a bit further. Take the case of vision. A functional-augmentative enhancement to this capacity would be one that allowed a person to see more clearly, identify objects at a greater distance, switch focus more quickly and with less effort, and so on, than she could do before the intervention (on some accounts, regardless of whether she had been dealing with a so-called "medical" problem along any of the relevant dimensions; see Zohny in press for an in-depth discussion). For hearing, it would be one that

allowed a person to perceive a wider range of decibels, say, or to discriminate between auditory signals more easily and with greater accuracy. Or take the case of memory: on a functional-augmentative approach, a person's memory would be "enhanced" if—in virtue of some intervention—she could now recall more events (or facts), more vividly, or for a longer duration than before.

Importantly, none of this is to say that these functional augmentations would be *desirable*. Clearly, that would depend upon a number of factors, including the person's values, needs, and wishes (as well as those of relevant others), her physical and social environment, and her past experiences, to name but a few. To continue with the example of memory, one need only to think of soldiers who have experienced the traumas of war, or of survivors of rape or other forms of sexual assault, to realize that memory, and especially augmented memory, has the potential to be "a devastating shackle" (Earp, Sandberg, Kahane, & Savulescu 2014, p. 4; see also Earp 2015).

Or let us return to the case of hearing. Depending on how this capacity is described[1] and on the circumstances in which one found oneself, augmented hearing might turn out to be extremely undesirable: just imagine being trapped in a perpetually noisy environment. A similar analysis, we believe, applies to many other functions or capacities that are commonly discussed in the neuroenhancement literature. Simply put: "more is not always better, and sometimes less is more" (Earp, Sandberg, Kahane, & Savulescu 2014, p. 1). Indeed, in some cases, the *diminishment* of a specific capacity or function, under the right set of circumstances, could be required to achieve the best outcome.

And so it is for *moral* capacities. Whether having "more" of a morally-relevant capacity or emotion such as empathy, righteous anger, or a sense of fairness is desirable (morally or otherwise) depends upon numerous factors: the circumstances, one's baseline moral motivations and capacities, the social role one is fulfilling, and so on (see Douglas 2008; Douglas 2013). It seems plausible that a morally good agent would be able to respond flexibly to different situations, and to employ or tap into different cognitive and emotional

[1] This caveat points to an ambiguity in our functional-augmentative account of enhancement. As we wrote, such enhancement involves "increasing the ability of the function to do what it normally does" (Earp et al. 2014, p. 2). But what is the "ability of [a] function," exactly, and what does it mean to "increase" it? Plainly, it depends upon the function in question, which in turn depends upon, among other things, the level of description one uses to cordon off that function from alternative targets of intervention. In this case, if by "augmented" hearing, one meant simply *more sensitive* hearing, as implied by our illustration, then a noisy environment might indeed make this "enhancement" undesirable. If instead one meant the augmentation of a higher-order hearing capacity—one that allowed a person to pick up on subtle sounds in a quiet environment, but also to "tune out" loud and uncomfortable sounds in a noisy environment, then the augmentation of this more flexible, higher-order capacity would be much more likely to be regarded as desirable across a range of possible circumstances. This is very similar to what we have in mind when we talk about the neuroenhancement of higher-order *moral* capacities, as will become clear over the course of what follows.

resources as necessary to arrive at the motives, decisions, and behaviors that are morally desirable given the context. As we will argue, it is this higher-order capacity to respond flexibly and appropriately to a range of scenarios that should be augmented, if possible, to achieve reliable moral enhancement.

Consider the ability to empathize. This is, on any reasonable account, a capacity that is "implicated in moral reasoning, decision-making, acting and so forth" (Pacholczyk 2011, p. 253), and it is one whose potential modification has become a staple of the moral enhancement literature (see, e.g., Persson & Savulescu 2013). To see how this capacity might be biomedically "enhanced" in the functional-augmentative sense, imagine that someone took a drug similar to MDMA (see, e.g., Sessa 2007) which, at least temporarily, made it so that the person became able to experience *more* empathy, or to experience empathy *more readily,* in response to relevant stimuli. Would this be morally desirable? Would the person behave "more morally" while under the influence of the drug? Obviously, it depends. As we will see in the following section, the relationship between increasing or strengthening a morally-relevant capacity such as empathy ("enhancing" it, in the functional-augmentative sense), morally improving one's motives and behavior, and becoming a morally better agent, is complex and context-specific. It also depends on which moral theory is correct or most justified, which is open to dispute: obviously, people will disagree about what constitutes, e.g., "morally desirable behavior," and they may also disagree about how, if at all, the moral goodness of an agent depends upon the moral desirability of her behavior (or motivations, etc.).

In short, if the goal is to produce morally better agents, on whatever (plausible) conception of "morally better"—as we have suggested should be the case, and as we have highlighted with our agential definition of moral neuroenhancement—then a narrow focus on "boosting" specific moral capacities, we believe, is likely to be at best a small part of the story.

The limits of empathy

To see why this is the case, let us pursue the example of empathy in greater detail.[2] As the neuroscientist Simon Baron-Cohen (2011) has argued, even such "obviously" morally desirable capacities as the ability to empathize may have morally undesirable consequences in certain cases. Mark Stebnicki (2007), for example, has discussed the phenomenon of

[2] This paragraph is adapted from Earp, Sandberg, Kahane, and Savulescu (2014).

"empathy fatigue," which refers to the physical and emotional exhaustion that grief and trauma counselors sometimes come to face: their inability to *distance* themselves emotionally from the pain and suffering of their clients may ultimately interfere with optimal job performance (for related work, see, e.g., Melvin 2012 and Perry et al. 2011 on "compassion fatigue" among nurses). Likewise, Carol Williams (1989) has hypothesized that among helping professionals, high emotional empathizers may be disposed to earlier career burnout, thereby undermining their long-term effectiveness (see Zenasni et al. 2012 for a more recent discussion).

Empathy can also lead us astray when it comes to making moral judgments specifically. For example, there is the "identifiable victim" effect (but see Russell 2014), according to which people have a stronger emotional reaction to the suffering of a known individual (thereby motivating them to help that specific individual) than to the greater suffering of an "anonymous" individual (or group of individuals) that would benefit more from the same act or degree of help (see, e.g., Jenni & Loewenstein 1997; Small & Loewenstein 2003). As the economist Thomas Schelling (1984) once observed: "Let a six-year-old girl with brown hair need thousands of dollars for an operation that will prolong her life until Christmas, and the post office will be swamped with nickels and dimes to save her. But let it be reported that without a sales tax the hospital facilities of Massachusetts will deteriorate and cause a barely perceptible increase in preventable deaths—not many will drop a tear or reach for their checkbooks" (p. 115). Making the point more generally, Jesse Prinz (2011) has argued, "empathy is prone to biases that render moral judgment potentially harmful" (p. 214).

Similar statements have been made by Paul Bloom (2013, 2016), Peter Singer (2014), Ole Martin Moen (2014), and others. While this intellectual movement "against empathy" (Bloom 2016), and in favor of more "abstract" or "cold" cognition geared toward maximizing welfare on a utilitarian calculus has its detractors (e.g., Christian 2016; Cummins 2013; Srinivasan 2015; but see McMahan 2016), the broader point remains the same: moral agents require flexibility in how they "deploy" their lower-order moral capacities, so that they can respond appropriately to justified reasons for making certain kinds of decisions over others. By contrast, trying generally to "dial up" or "dial down" some discrete moral capacity or function (assuming that such a thing were even possible without incurring serious adverse side-effects) will be at best a highly unreliable means to becoming a morally better agent.

Thus, whether spraying a dose of oxytocin up someone's nose to increase empathy or trust, say, is likely to amount to an agential moral enhancement will depend not only upon the specific effects of the drug at various dosages, but upon the psychological and social context

in which this is done. For example, it will depend upon who is receiving the dose of oxytocin, what her values are, what her chronic and momentary mental states are, what situation(s) she is in both short and long-term, what particular decisions she faces, and is likely to face, and so on (see Wudarczyk et al. 2013 for a related discussion).

So it wouldn't be just "more empathy" (*tout court*) that would be expected to lead to the improvement of a moral agent, qua moral agent, but rather an increase in what might roughly be described as a kind of second-order empathic control – an ability to (1) know or to identify, whether consciously or unconsciously, when it is morally desirable to feel empathy and/or allow it to shape one's outward behavior (and in what way), as well as (2) to be able to feel such empathy, or if necessary, suppress such feelings (or their effects on behavior), in accordance with (1).

Similarly with a sense of fairness or justice, feelings of righteous anger or moral disgust, motivations associated with causing harm, etc.—the whole suite of underlying moral emotions, intuitions, and capacities (see generally, e.g., Haidt 2007; Haidt & Joseph 2004). If such capacities could be developed or augmented at their second-order level of description, this would be a more promising target, we believe, for interventions aimed at achieving (agential) moral enhancement, whether the intervention happened to be carried out with the assistance of a neurotechnology that acted directly on the brain, or whether it was of a more familiar kind (e.g., traditional moral instruction without the aid of, say, brain stimulation or pharmaceuticals). In other words, it is likely that augmenting higher-order capacities to *modulate* one's moral responses in a flexible, reason-sensitive, and context-dependent way would be a more reliable, and in most cases more advisable, means to agential moral enhancement.

Direct vs. indirect moral enhancement

We are not the first to distinguish between the direct modification of specific moral traits, functions, or emotions versus the modification of higher-order moral capacities. Instead, our discussion shares some features with, for example, Schaefer's recent examination of "direct vs. indirect" moral enhancement (Schaefer, 2015). Direct moral enhancements, according to Schaefer, "aim at bringing about *particular* ideas, motives or behaviors," which he sees as being problematic for a number of reasons, in much the same way that we see the functional augmentation of first-order moral capacities or emotions as being problematic. By contrast, what Schaefer calls indirect moral enhancements "aim at making people more reliably

produce the morally correct ideas, motives or behaviors without committing to the content of those ideas, motives and/or actions" (Schaefer, 2015, p. 261, emphasis added), an aim that is consistent with that of the second-order interventions we have just alluded to.

Briefly, Schaefer disfavors "direct" moral enhancement (especially if it were carried out programmatically, by, for example, a state, rather than undertaken voluntarily on a case by case basis) because he worries that such "enhancement" could suppress dissent: if everyone were forced to hold the exact same or even highly similar moral beliefs, dispositions, etc., then moral disagreement would likely fall by the wayside. But such disagreement is valuable, Schaefer argues, because without it, "conventional wisdom will go unchallenged and moral progress becomes essentially impossible" (Schaefer, 2015, p. 265). Schaefer also disfavors "direct" moral enhancement because, in his view, such enhancement might interfere with, bypass, or otherwise undermine conscious reasoning and rational deliberation. Instead of "coming to believe or act on a given moral proposition because it is the most reasonable," he fears, "we would come to believe or act on it because a particular external agent (the enhancer) said it is best" (p. 268) and perhaps even "implanted" it in our brains.

We are not confident that this fear is justified. At least, more work would need to be done to show how such enhancement would be significantly different from, or worse than, various current forms of moral education that aim at inculcating specific moral tendencies, values, and beliefs—sometimes, as in the case of children, without first explaining the reasons why (although such explanations may of course later be given or become apparent over time on their own). Insofar as this is a valid concern, however, it could plausibly be addressed by emphasizing the need for individual, voluntary enhancement, as opposed to top-down or coerced external enhancement, and indeed Schaefer seems open to this view. But whatever the solution to this problem, we agree that the ability to deliberate and to rationally evaluate different moral propositions is important, and that there would be strong reasons against pursuing any form of moral enhancement that had the effect of impairing such an ability.

In fact, this very same acknowledgement of the importance of rational deliberation (though note that we do not presume that genuine moral insights must always be strictly rationally derived) paves the way for one of Schaefer's main proposals for what he sees as an alternative to direct moral enhancement, namely "indirect" moral enhancement. "It is quite plausible to think," he writes, "that there is value in the process itself of deliberating over a moral proposition, both within one's own mind and in discussion with others" (2015, p. 268).

One form of indirect moral enhancement that would be at least *prima facie* permissible (and perhaps even desirable), then, in light of this consideration, would be to *improve the reasoning process itself*. The idea is that, all else being equal, better reasoning is likely to result in better moral beliefs and decisions, and consequently to better—that is, more moral—action.

What would this actually look like? Among other things, it might involve improving people's logical abilities (i.e., "people's ability to make proper logical inferences and deductions, spot contradictions in their own beliefs and those of others, as well as formulate arguments in a way that can highlight the true point of contention between interlocutors"); promoting conceptual understanding (since "vague and distorted ideas will lead to unreliable inferences, inducing behaviors that are not in line with someone's considered judgments"); and overcoming cognitive biases (Schaefer 2015, p. 276). Importantly, none of these enhancements would force a person to adopt any *particular* moral position, motivation, or behavior—thereby allowing for moral disagreement to persist, which is important, Schaefer claims, for moral progress—nor would they undermine rational deliberation, since, by definition, they would be expected to foster it. Certainly, allowing and/or helping people to reason better, with fewer biases, should be seen as uncontroversial (setting aside for now the crucial question of means); and this does seem to be plausible way of "mak[ing] people more reliably produce the morally correct ideas, motives, and/or actions without specifying the content of those ideas, motives, and/or actions" in advance (p. 262; see also Douglas 2008, p. 231; Douglas 2013, p.161).

Schaefer's other major proposal for "indirect" moral enhancement is something he calls "akrasia reduction," where akrasia is defined as acting against one's better judgment, typically due to weakness of will. As Schaefer writes:

> Weakness of will affects morality in a very straightforward way. Someone recognizes that some course of action is morally ideal or morally required, but nevertheless fails to carry out that action. For instance, someone might recognize the moral imperative to donate significant sums of money to charity because that money could save a number of lives, yet remain selfishly tight-fisted. This is a failure of someone's consciously-held moral judgments to be effective. (2015, p. 277)

Schaefer argues that individuals should be permitted to "work on" their weakness of will—in order to reduce associated akrasia—but that no one should be forced to undertake such (indirect) moral self-enhancement (with the possible exception of children being

brought up by their parents; for a related discussion see Maslen et al. 2014). Again, this seems uncontroversial: strengthening one's will to act in accordance with one's considered judgments, moral or otherwise, is usually[3] a virtue on any plausible account (see Persson & Savulescu 2016); the only significant debate in this area, as we have just suggested, has to do with the question of means (see Focquaert & Schermer 2015).

Traditional moral education, including the development and maintenance of good motivations and habits, is the most obvious—and least contentious—possibility. We take it that attempting to reduce one's weakness of will (and improve one's reasoning abilities) by such "traditional" methods as, e.g., meditation, Aristotelian habituation (see Steutel & Spiecker 2004), studying logic or moral philosophy, and engaging in moral dialogue with others, is clearly permissible—indeed laudable—and we expect that few would disagree. This is the "philosophically mundane" version of moral enhancement that we flagged in our introduction. It is rather moral enhancement[4] by means of, or at least involving, neurotechnological intervention, specifically, that we expect will be seen as more controversial, and we turn to this case in the following section.

The role of neurotechnology in moral enhancement

Is it permissible, or even desirable, to engage in "indirect" moral self-enhancement (on Schaefer's account), or agential moral self-enhancement via modulation of second-order moral capacities (on our account), *with the help of neurotechnologies*? Let us first re-emphasize that we are concerned only with *voluntary* moral (self) enhancement in this chapter, which we take to be the easiest case to justify, chiefly on liberal or libertarian grounds: we are setting aside the much more difficult question of whether wide-scale enhancement of, e.g., the moral character of all of humanity could be justified (if it were possible). Let us also state at the outset that *if* moral enhancement with the aid of neurotechnology is in fact permissible or even desirable, it is likely to be so only under certain conditions. For reasons we will soon discuss, the most promising scenario for permissible, much less desirable or optimal, agential moral (self) neuroenhancement seems to

[3] Obviously, there are exceptions. Consider Heinrich Himmler; he had firm (but false) moral beliefs, and given this, the weaker his will, the better (see, for discussion, Bennett 1974). Schaefer (2015) actually discusses this issue at length, arguing, essentially, that while there are always exceptions to the rule, in most cases and on balance, akrasia reduction will lead to moral improvement.

[4] Here we mean "indirect" moral enhancement (on Schaefer's account) or agential moral enhancement via modulation of second-order moral capacities (on our account).

us to be one in which at least the following conditions apply:

(1) the drug or technology in question is used as an *aid* or *adjunctive* intervention to well-established "traditional" forms of moral learning or education (rather than used, as it were, in a vacuum), such that

(2) the drug or technology allows for conscious reflection about and critical engagement with any moral insights that might be facilitated by the use of the drug (or by states-of-mind that are occasioned by the drug); and

(3) the drug or technology has been thoroughly researched, with a detailed benefit-to-risk profile, and is administered under conditions of valid consent

We are not prepared to argue that any currently available drug meets all three of these conditions. However, it does seems possible that some currently available putative cognitive enhancers, such as modafinil and methylphenidate (see, e.g., Greely et al. 2008; Turner et al. 2003; but see Lucke et al. 2011; Outram 2010), could, if used as an adjunct to moral education, potentially meet them in the future. So too might certain drugs or other neurointerventions that worked by attenuating emotional biases that would otherwise impede moral learning (although research in this area is currently nascent and scarce). Finally, although we will discuss the example of so-called psychedelic drugs in the following section, we must be clear that we do not advocate the use of these drugs by anyone, in any setting, but are rather flagging them as possible targets for future research (see Earp, Savulescu, and Sandberg 2012 for a related discussion).

With respect to conditions (1) and (2), it should be noted that "traditional" means of moral education frequently operate by enhancing an agent's moral understanding: her understanding of what morality requires and why. This requires some degree of rational engagement. Now, some critics of "direct" moral neuroenhancement, such as John Harris (2012, 2013) have suggested that interventions into what we are calling first-order moral emotions or capacities would *not* enhance the agent's moral understanding. Others have made similar claims. Fabrice Jotterand (2011), for instance, 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 (2014a) 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 "someone who reads Tolstoy arguably learns *reasons* to be less judgmental and in doing so develops greater understanding: someone who takes a pill has merely *caused* their sentiments to alter" (p. 2 and p. 3).[5]

But what about reading Tolstoy *while* taking a pill (i.e., a pill that enhances one's moral learning vis-à-vis the text)? The supposition here is that this hypothetical pill would occasion a state of mind that made the moral lessons of Tolstoy more apparent, or more compelling, to the reader.[6] Indeed, the importance of a robust educational or learning context cannot be overstated: what we envision is a *facilitating*, rather than *determining* role for any drug or neurotechnology (see Naar in press; see also Earp, Sandberg, & Savulescu in press), underscoring the need for critical engagement with some kind of actual moral "content" (e.g., "norms or values"). Arguably, we need not look to the distant future, or to hypothetical sci-fi scenarios, to imagine what such drug-assisted (as opposed to drug-caused, or drug-determined) agential moral enhancement might plausibly look like. Instead, we can look to the past and present.

Attempted moral neuroenhancements, past and present

In a recent book chapter, the theologian Ron Cole-Turner (2015) writes that technologies of moral enhancement "are not new. For millennia we have known that certain disciplines and techniques can enhance our spiritual awareness. We have also known that certain substances can alter our consciousness in interesting ways" (p. 369). Jonathan Haidt (2012) expands on this idea, noting that most traditional societies have a coming-of-age ritual designed to transform immature children into morally and socially competent adults, and that many of them use "hallucinogenic drugs to catalyze this transformation" (p. 266). The mental states induced by such drugs, according to anthropologists, are intended to "heighten" moral

[5] Please note that the page numbers for the Sparrow quotes come from the version of his essay available online at http://profiles.arts.monash.edu.au/wp-content/arts-files/rob-sparrow/ImmoralTechnologyForWeb.pdf.
[6] But note that there are other ways of responding to the above concerns as well. For example, some defenders of moral neuroenhancement have suggested that even "direct" interventions into first-order moral emotions or capacities could conceivably improve moral understanding, in certain cases, by attenuating emotional barriers to sound moral deliberation (Douglas 2008). And even if a first-order moral neuroenhancement intervention had no positive effect on moral understanding initially, Wasserman (2011) has argued that we might expect an agent's experience with morally desirable motives and conduct (as judged against a relatively stable background) 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).

learning "and to create a bonding among members of the cohort group" (quoted in Haidt 2012, p. 266).

Notice the words "enhance," "catalyze" and "heighten" in these quotations, which suggest a *facilitating*, rather than strictly *determining*, role for the hallucinogenic drugs in these societies, administered as part of a richly contextualized process of moral learning. This is worth highlighting, in our view, since moral lessons, abilities, dispositions, etc., that are achieved or developed with the *help* of a neurotechnology—as opposed to directly caused by it (thereby preserving space for conscious reflection, effort, and engagement)—could be seen as posing less of a threat to such important issues as authenticity, autonomy, and rational deliberation, as emphasized by (among others) Schaefer (2015).

Consider the use of ayahuasca, a plant-based brew containing MAO-inhibitors and N,N-dimethyltryptamine or DMT, which has been employed in traditional shamanic ceremonies across the Amazon basin and elsewhere for hundreds of years (McKenna, Towers, & Abbott 1984; Homan 2011). According to Michael J. Winkelman (2015, p. 96) the active ingredients in ayahuasca, in combination with a certain restrictive diet, may occasion an "altered state of consciousness" in the initiate in which her "artistic and intellectual skills" are seen as being enhanced, thereby allowing her to better appreciate the teachings of the shaman. Winkelman stresses, however, the *interactive* relationships among: healer and patient (initiate), various "ritual factors," and what he calls psycho- and socio-therapeutic activities, in shaping the learning experience. A similar emphasis is given by William A. Richards (2015, p. 140) in reference to the drug psilocybin:

> It is clear that psilocybin … never can be responsibly administered as a medication to be taken independent of preparation and careful attention to the powerful variables of [one's mindset] and [physical] setting. One cannot take psilocybin as a pill to cure one's alienation, neurosis, addiction, or fear of death in the same way one takes aspirin to banish a headache. What psilocybin *does* is provide an opportunity to explore a range of non-ordinary states. It unlocks a door; how far one ventures through the doorway and what awaits one … largely is dependent on non-drug variables.

We caution the reader that it is not currently legal in many jurisdictions to consume these substances (see Ellens & Roberts 2015 for further discussion), and we reemphasize that we are not advocating their use by any person, whether for attempted moral enhancement or anything else. Our point is merely that the intentions for which, and manner in which, some

hallucinogenic drugs have been used in certain settings resemble the approach to moral enhancement for which we argue in this chapter (i.e., a facilitating role for the neurotechnology, active engagement with moral content, a rich learning context, etc.), suggesting that this approach is not a radical departure from historical practices. That said, rigorously controlled, ethically conducted scientific research into the effects of such drugs on moral learning or other moral outcomes (in concert with appropriate psychosocial and environmental factors) may well be worth pursuing (Tennison 2012; see also Frecska, Bokor, & Winkelman 2016; Griffiths et al. 2006, Griffiths et al. 2008).

Objections and concerns

We see it as uncontroversial that individuals have moral reasons to increase the moral desirability of their character, motives, and conduct, and that actually doing so is morally desirable. Moral neuroenhancements in particular appear to be immune to many of the more common moral concerns that have been raised about neuroenhancements (or bioenhancements 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 by undermining commitments to solidarity or equality. Moral neurenhancements are unusual among the main types of neuroenhancements that have been discussed heavily in the recent literature in that they might plausibly be expected to advantage, rather than disadvantage, others (though see, for a criticism of this view, Archer 2016).

Nevertheless, some significant concerns have been raised regarding the permissibility and desirability of undergoing moral neuroenhancements, or certain kinds of moral neuroenhancements. Some of these are general concerns about enhancing the moral desirability of our characters, motives, and conduct, regardless of whether this is undertaken through moral neuroenhancement or through more familiar means such as traditional moral education. In this category are concerns that stem from a general skepticism about the *possibility* of moral improvement, as well as concerns about whether we have adequate means for resolving disagreement and uncertainty about what character traits, motives, and conduct are morally desirable and why. However, the first of these concerns strikes us as implausible: even if people disagree on certain moral issues, there are surely *some* moral behaviors and/or dispositions that everyone can agree are better than *some* alternative moral behaviors and/or dispositions—even if only at the far extremes—and if it is psychologically

possible to move even a little bit from the less desirable side of things toward the more desirable side, then (agential) moral improvement is also possible. As for the second concern about resolving disagreements, this does not seem to us to be damning even if it is true: of course there will be disagreements about what counts as "morally desirable"—in the realm of traditional moral enhancement as well as in the realm of moral neuroenhancement—but as Schaefer (2015) points out, such disagreement is in fact quite healthy in a deliberative society and is perhaps even necessary for moral progress.

Other points of contention have to do with general concerns about neuroenhancement that would also apply to non-moral neuroenhancements. In this category are concerns regarding the unnatural means or hubristic motives that biomedical enhancement is said to involve (Kass 2003, Sandel 2007). We will set those issues aside as being tangential to the focus of this chapter. There are, however, also more specific concerns about moral neuroenhancements—concerns that would not apply equally to traditional means of moral enhancement or to other kinds of neuroenhancement. The remainder of this section outlines two dominant concerns in this category.

Concern 1: Restriction of Freedom

One concern that has been raised regarding moral neuroenhancement, or at least certain variants of it, is that it might restrict freedom or autonomy. Harris (2011) argues that we might have reason to abstain from moral neuroenhancements because they would restrict our freedom to perform morally undesirable actions or to have morally undesirable motives (see also Ehni & Aurenque 2012, p. 233 and, for a more general discussion of the effects of neuroenhancement on autonomy, Bublitz & Merkel 2009).

Two main types of response have been made to this line of argument. The first is that, even where moral neuroenhancements do restrict freedom, it might nevertheless be morally permissible or, all things considered, morally desirable to undergo such enhancements (DeGrazia 2014; Douglas 2008, 2011; Savulescu et al. 2014; Savulescu & Persson 2012; Persson & Savulescu 2016). Suppose that you come across one person about to murder another. It seems that you should intervene to prevent the murder even though this involves restricting the prospective murderer's freedom to act in a morally undesirable way. Similarly, it seems that, if he could, the would-be murderer should have restricted his own freedom to commit the murder, for example, by having a friend lock him in his room on days when he knows he will be tempted to commit a murder. The obvious way of accounting for these

intuitions is to suppose that, in at least some cases, any disvalue associated with restricting one's freedom to act in morally undesirable ways is outweighed by the value of doing so.

The second response has been to deny that *all* moral neuroenhancements would in fact restrict freedom, thus limiting the concern about freedom to a subset of moral neuroenhancements. Responses in the second category sometimes begin by noting that worries about the freedom-reducing effect of moral neuroenhancements presuppose that freedom is consistent with one's motives and conduct being causally determined (Persson & Savulescu, 2016). If we could be free only if we were causally undetermined, we would already be completely unfree, because we are causally determined, in which case moral neuroenhancements could not *reduce* our freedom. Alternatively, we are free only because, at least *some* of the time, we act on the basis of reasons, in which case moral neuroenhancements which operate without affecting (or by actually enhancing) our capacity to act on the basis of reasons would not reduce our freedom (DeGrazia 2014; Savulescu et al. 2014; Savulescu and Persson 2012; Persson and Savulescu 2016).

Finally, although we cannot pursue this argument in detail, we have suggested that agential moral neuroenhancement could plausibly be achieved by targeting second-order moral capacities, thereby increasing a kind of "moral impulse control." On this account, we should be open to the idea that moral neuroenhancments could actually *increase* a person's freedom, that is, her ability to behave autonomously. Niklas Juth (2011, p. 36) asks: "Can enhancement technologies promote individuals' autonomy?" And answers: "*Yes*. In general plans require capacities in order for them to be put into effect and enhancement technologies can increase our capacities to do the things we need to do in order to effectuate our plans." Similarly, Douglas (2008) has argued that diminishing counter-moral emotions (things that tend to interfere with whatever counts as good moral motivation) is also a kind of second-order moral enhancement, and in many cases it will also increase freedom (since the counter-moral emotions are also constraints on freedom; see also Persson & Savulescu, 2016).

Concern 2: Misfiring

A second concern that can be raised regarding moral neuroenhancements maintains that attempts at moral neuroenhancement are likely to misfire, bringing about moral deteriorations rather than improvements. This is not a concern about successful moral neuroenhancements, but is rather a concern that actual attempts at moral neuroenhancement 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 that would be required for genuine moral neuroenhancement 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, so that attempts at such fine-tuning are likely to fail.

We might respond to this in two ways. First, we agree, as we argued earlier, that the elimination or amplification of particular moral dispositions or capacities is likely, on balance, to be an unreliable way of bringing about genuine (agential) moral enhancement. But we are less convinced that the technologies we are likely to have available to us could not bring such enhancement about. This is based on our exploration of the possibility of drug-assisted moral learning, where we drew on the examples from certain so-called traditional societies, where such moral learning is generally understood not only to be possible, but also (at least sometimes) actually occurrent. Whether such moral learning involves, or amounts to, a "fine-tuning of our [moral] dispositions," then, would be beside the point, because agential moral neuroenhancement would, by whatever mechanism, be taking place (thus showing that it is indeed possible).

Agar (2010, 2013) sets forward a more limited variant of the concern raised by Harris. He argues that attempted moral neuroenhancements may have good chances of success when they aim only to correct subnormal moral functioning (such as, e.g., might be exhibited by a hardened criminal), bringing an individual within the normal range, but that they are likely to misfire when they aim to produce levels of moral functioning above the normal range (he does not comment on moral neuroenhancements that operate wholly within the normal range). Subnormal moral functioning is often the result of relatively isolated and easily identified defects such as, for example, the deficient empathy that characterizes psychopathy (but see Bloom 2016 for further discussion). Agar speculates that these defects could relatively safely be corrected. However, he argues that, to attain super-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 and there is a serious risk that it would misfire.

Defenders of moral neuroenhancement have conceded to these concerns, acknowledging 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 moral neuroenhancements will fail, perhaps resulting in moral deterioration (Douglas 2013; Savulescu et al. 2014). However, it is not obvious that achieving super-normal moral functioning would, as Agar suggests, always require the alteration of multiple capacities. Imagine a person who would function at a super-normal level, except for the fact that she performs suboptimally on a single moral dimension. An intervention affecting that dimension alone might be sufficient to achieve super-moral functioning. Moreover, focusing on augmenting the powers of more traditional moral education, as we have proposed here, could be expected to produce moral improvements across a range of dimensions, and might in this way produce the breadth and balance of moral improvement that Agar takes to be necessary without requiring multiple distinct enhancement interventions.

Finally, some doubt has been cast on the notion that neurointerventions are invariably inapt when complex and subtle psychological alterations are sought. For example, Douglas (2011) 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 or modifying the intervention if it misfires. Douglas suggests that a similar approach might be justified in relation to moral neuroenhancers.

Conclusion

In this chapter, we have considered moral enhancement in terms of agential moral neuroenhancement. This means any improvement in a moral agent, qua moral agent, that is effected or facilitated in some significant way by the application of a neurotechnology. We have distinguished between first- and second-order moral capacities. First-order capacities include basic features of our psychology which are relevant to moral motivations and behavior, such as empathy and a sense of fairness. As we argued, there is no straightforward answer to whether augmenting these functions constitutes agential moral enhancement, just as one cannot say that having super-sensitive hearing is good for a person without knowing that person's situation (for a related discussion in the context of disability, see Kahane & Savulescu 2009 and Kahane & Savulescu 2016). What makes having a capacity valuable is

being able to employ it in the right circumstance and in the right way, which means having appropriate control over its regulation.

In addition, we have emphasized a *facilitating* role for neurotechnologies in bringing about moral enhancement, rather than a determining role, which leaves room for rational engagement, reflection, and deliberation: this allowed us to address concerns that such processes might be undermined. Another consequence of thinking in terms of facilitation is that moral neuroenhancers should not ideally be used "in a vacuum," but rather in a meaning-rich context, as we illustrated briefly with "traditional" examples (e.g., ayahuasca). Finally, we have responded to two main moral concerns that have been raised regarding the pursuit of moral neuroenhancement, namely that it would restrict freedom or else "misfire" in various ways. We argued that these concerns, while worth taking seriously, are not fatal to the view we have presented.

Acknowledgment

Minor portions of this chapter have been adapted from Douglas (2015). The first author (Earp) wishes to thank the editors of this volume for the opportunity to contribute a chapter, and Michael Hauskeller for inviting him to present some of these ideas at the 2016 Royal Institute of Philosophy Annual Conference held at the University of Exeter.

Key Readings

DeGrazia, D. (2014). Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of Medical Ethics* 40(6), 361–8.

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

Harris, J. (2016). *How to be Good*. Oxford: Oxford University Press.

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

Persson I., & Savulescu, J. (2016). Moral bioenhancement, freedom and reason, *Neuroethics*. Available online ahead of print at http://link.springer.com/article/10.1007/s12152-016-9268-5.

Sparrow, R. (2014). Better living through chemistry? A Reply to Savulescu and Persson on "moral enhancement." *Journal of Applied Philosophy* 31(1), 23–32.

Wiseman, H. (2016). *The Myth of the Moral Brain: The Limits of Moral Enhancement*. Cambridge, M.A.: MIT Press.

References

Agar, N. (2010). Enhancing genetic virtue? *Politics and the Life Sciences* 29(1), 73–75.

Agar, N. (2013). A question about defining moral bioenhancement. *Journal of Medical Ethics*, 40(6), 369–70.

Archer, A. (2016). Moral enhancement and those left behind. *Bioethics*. Available online ahead of print at http://onlinelibrary.wiley.com/doi/10.1111/bioe.12251/full.

Baron-Cohen, S. (2011). Autism, empathizing-systemizing (e-s) theory, and pathological altruism. In *Pathological Altruism*, eds. B. Oakley, A. Knafo, G. Madhaven, & D. S. Wilson, pp. 344-348. Oxford: Oxford University Press.

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences*, 15(7), 301-309.

Bennett, J. (1974). The conscience of Huckleberry Finn. *Philosophy*, 49(188), 123–134.

Bloom, P. (2013, May 20). The baby in the well. *The New Yorker*. Available at http://www.newyorker.com/magazine/2013/05/20/the-baby-in-the-well.

Bloom, P. (2016). *Against Empathy*. New York: HarperCollins.

Bublitz, J. C., & Merkel, R. (2009). Autonomy and authenticity of enhanced personality traits. *Bioethics*, 23(6), 360–374.

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

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.

Christian, R. (2016, March 23). Should you give money to beggars? Yes, you should. *Practical Ethics*. University of Oxford. Available at http://blog.practicalethics.ox.ac.uk/2016/03/guest-post-should-you-give-to-beggars-yes-you-should/.

Clausen, J. (2010). Ethical brain stimulation–neuroethics of deep brain stimulation in research and clinical practice. *European Journal of Neuroscience*, *32*(7), 1152-1162.

Cole-Turner, R. (2015). Spiritual enhancement. In *Religion and Transhumanism: The Unknown Future of Human Enhancement*, eds. C. Mercer & T. J. Trothen, pp. 369-383. Santa Barbara, C.A., and Denver, C.O.: Praeger.

Craig, J. N. (2016). Incarceration, direct brain intervention, and the right to mental integrity–a reply to Thomas Douglas. *Neuroethics*. Available online ahead of print at http://link.springer.com/article/10.1007/s12152-016-9255-x.

Crockett, M. J. (2014). Moral bioenhancement: a neuroscientific perspective. *Journal of Medical Ethics*, *40*(6), 370-371.

Cummins, D. (2013, October 20). Why Paul Bloom is wrong about empathy and morality. *Psychology Today*. Available at https://www.psychologytoday.com/blog/good-thinking/201310/why-paul-bloom-is-wrong-about-empathy-and-morality.

Davis, N. J., & Koningsbruggen, M. V. (2013). "Non-invasive" brain stimulation is not non-invasive. *Frontiers in Systems Neuroscience*, 7(76), 1-4.

DeGrazia, D. (2014). Moral enhancement, freedom, and what we (should) value in moral behavior. *Journal of Medical Ethics* 40(6), 361-368.

Donaldson, Z. R., & Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science*, 322(5903), 900-904.

Douglas, T. (2015). The morality of moral neuroenhancement. In *Handbook of Neuroethics,* eds. J. Clausen & N. Levy, pp. 1227-1249. Dordrecht: Springer.

Douglas, T. (2014). Enhancing moral conformity and enhancing moral worth. *Neuroethics*, 7(1), 75–91.

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

Douglas, T. (2013). Moral enhancement via direct emotion modulation: A reply to John Harris. *Bioethics*, 27(3), 160–168.

Douglas, T., Bonte, P., Focquaert, F., Devolder, K., & Sterckx, S. (2013). Coercion, incarceration, and chemical castration: An argument from autonomy. *Journal of Bioethical Inquiry*, 10(3), 393-405.

Earp, B. D. (2015). "Legitimate rape," moral coherence, and degrees of sexual harm. *Think*: *A Journal of the Royal Institute of Philosophy*, 14(41), 9-20.

Earp, B. D., Sandberg, A., Kahane, G., and Savulescu, J. (2014). When is diminishment a form of enhancement? Rethinking the enhancement debate in biomedical ethics. *Frontiers in Systems Neuroscience*, 8(12), 1-8.

Earp, B. D., Sandberg, A., & Savulescu, J. (2014). Brave new love: the threat of high-tech "conversion" therapy and the bio-oppression of sexual minorities. *AJOB Neuroscience*, 5(1), 4-12.

Earp, B. D., Sandberg, A., & Savulescu, J. (in press). The medicalization of love: response to critics. *Cambridge Quarterly of Healthcare Ethics*, in press. Available online ahead of print at https://www.researchgate.net/publication/301748753_The_medicalization_of_love_Response_to_critics?ev=prf_pub.

Earp, B. D., Savulescu, J., & Sandberg, A. (2012, June 14). Should you take ecstasy to improve your marriage? Not so fast. *Practical Ethics*. University of Oxford. Available at http://blog.practicalethics.ox.ac.uk/2012/06/should-you-take-ecstasy-to-improve-your-marriage-not-so-fast/.

Earp, B. D., Wudarczyk, O. A., Sandberg, A., & Savulescu, J. (2013). If I could just stop loving you: anti-love biotechnology and the ethics of a chemical breakup. *The American Journal of Bioethics*, *13*(11), 3-17.

Ehni, H.-J., & Aurenque, D. (2012). On moral enhancement from a Habermasian perspective. *Cambridge Quarterly of Healthcare Ethics*, 21(02), 223–234.

Ellens, J. H., & Roberts, B. (Eds). (2015). *The Psychedelic Policy Quagmire: Health, Law, Freedom, and Society*. Santa Barbara, C.A. and Denver, C.O.: Praeger.

Fenton, E. (2010). The perils of failing to enhance: A response to Persson and Savulescu. *Journal of Medical Ethics* 36(3), 148 –151.

Focquaert, F., & Schermer, M. (2015). Moral enhancement: do means matter morally? *Neuroethics*, *8*(2), 139-151.

Frecska, E., Bokor, P., & Winkelman, M. (2016). The therapeutic potentials of ayahuasca: possible effects against various diseases of civilization. *Frontiers in Pharmacology*, 7(35), 1-17.

Fregni, F., & Pascual-Leone, A. (2007). Technology insight: noninvasive brain stimulation in neurology—perspectives on the therapeutic potential of rTMS and tDCS. *Nature Clinical Practice Neurology*, *3*(7), 383-393.

Ginsberg, Y., Långström, N., Larsson, H., & Lichtenstein, P. (2013). ADHD and criminality: could treatment benefit prisoners with ADHD who are at higher risk of reoffending? *Expert Review of Neurotherapeutics*, 13(4), 345-348.

Ginsberg, Y., Långström, N., Larsson, H., & Lindefors, N. (2015). Long-term treatment outcome in adult male prisoners with attention-deficit/hyperactivity disorder: three-year naturalistic follow-up of a 52-week methylphenidate trial. *Journal of Clinical Psychopharmacology*, 35(5), 535-543.

Glenn, A. L., & Raine, A. (2008). The neurobiology of psychopathy. *Psychiatric Clinics of North America*, 31(3), 463-475.

Greely, H., Sahakian, B., Harris, J., Kessler, R. C., Gazzaniga, M., Campbell, P., & Farah, M. J. (2008). Towards responsible use of cognitive-enhancing drugs by the healthy. *Nature*, *456*(7223), 702-705.

Griffiths, R. R., Richards, W. A., McCann, U., & Jesse, R. (2006). Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology*, *187*(3), 268-283.

Griffiths, R. R., Richards, W. A., Johnson, M. W., McCann, U. D., & Jesse, R. (2008). Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later. *Journal of Psychopharmacology*, *22*(6), 621-632.

Gupta, K. (2012). Protecting sexual diversity: rethinking the use of neurotechnological interventions to alter sexuality. *AJOB Neuroscience*, 3(3), 24-28.

Haidt, J. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Vintage.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*(5827), 998-1002.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: how innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55-66.

Hamilton, R., Messing, S., & Chatterjee, A. (2011). Rethinking the thinking cap: ethics of neural enhancement using noninvasive brain stimulation. *Neurology*, 76(2), 187-193.

Harris, J. (2011). Moral enhancement and freedom. *Bioethics* 25(2), 102–111.

Harris, J. (2012). What it's like to be good. *Cambridge Quarterly of Healthcare Ethics*, 21(3), 293-305.

Harris, J. (2013). 'Ethics is for bad guys!' Putting the 'moral' into moral enhancement. *Bioethics*, *27*(3), 169-173.

Harris, J., & Chan, S. (2010). Moral behavior is not what it seems. *Proceedings of the National Academy of Sciences*, 107(50), E183.

Homan, J. (2011). *Charlatans, Seekers, and Shamans: The Ayahuasca Boom in Western Peruvian Amazonia*. Dissertation (University of Kansas). Available at https://kuscholarworks.ku.edu/handle/1808/8125.

Ipser, J., & Stein, D. J. (2007). Systematic review of pharmacotherapy of disruptive behavior disorders in children and adolescents. *Psychopharmacology*, 191(1), 127-140.

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, 14(3), 235-257.

Jotterand, F. (2011). 'Virtue engineering' and moral agency: will post-humans still need the virtues? *AJOB Neuroscience*, 2(4), 3–9.

Juth, N. (2011). Enhancement, autonomy, and authenticity. In *Enhancing Human Capacities*, eds. J. Savulescu, R. Muelen, and G. Kahane, pp. 34–48. Oxford: Blackwell.

Kahane, G., & Savulescu, J. (2009). The welfarist account of disability. In *Disability and Disadvantage*, eds. A. Cureton & K. Brownlee, pp. 14–53. Oxford: Oxford University Press.

Kahane, G., & Savulescu, J. (2016). Disability and mere difference. *Ethics*, 126(3), 774-788.

Kass, L. R. (2003). Ageless bodies, happy souls. *The New Atlantis*, 1, 9–28.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829-832.

Lane, A., Luminet, O., Nave, G., & Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. *Journal of Neuroendocrinology*, *28*(4), n. p.

Lane, A., Mikolajczak, M., Treinen, E., Samson, D., Corneille, O., de Timary, P., & Luminet, O. (2015). Failed replication of oxytocin effects on trust: the envelope task case. *PLOS ONE*, *10*(9), e0137000.

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1(1), 117-146.

Lucke, J. C., Bell, S., Partridge, B., & Hall, W. D. (2011). Deflating the neuroenhancement bubble. *AJOB Neuroscience*, 2(4), 38-43.

Maibom, H. L. (2014). To treat a psychopath. *Theoretical Medicine and Bioethics*, 35(1), 31-42.

Margari, F., Craig, F., Margari, L., Matera, E., Lamanna, A. L., Lecce, P. A., ... & Carabellese, F. (2014). Psychopathology, symptoms of attention-deficit/hyperactivity disorder, and risk factors in juvenile offenders. *Neuropsychiatric Disease and Treatment*, 11, 343-352.

Maslen, H., Earp, B. D., Kadosh, R. C., & Savulescu, J. (2014). Brain stimulation for treatment and enhancement in children: an ethical analysis. *Frontiers in Human Neuroscience*, 8(953), 1-5.

Melvin, C. S. (2012). Professional compassion fatigue: what is the true cost of nurses caring for the dying. *International Journal of Palliative Nursing*, 18(12), 606-611.

McMahan (2016). Philosophical critiques of effective altruism. *The Philosophers' Magazine*, 73, 92-99. Available at http://jeffersonmcmahan.com/wp-content/uploads/2012/11/Philosophical-Critiques-of-Effective-Altruism-refs-in-text.pdf.

McMillan, J. (2014). The kindest cut? Surgical castration, sex offenders and coercive offers. *Journal of Medical Ethics*, *40*(9), 583-590.

McKenna, D. J., Towers, G. N., & Abbott, F. (1984). Monoamine oxidase inhibitors in South American hallucinogenic plants: tryptamine and β-carboline constituents of ayahuasca. *Journal of Ethnopharmacology*, 10(2), 195-223.

Moen, O. M. (2014). Should we give money to beggars? *Think: A Journal of the Royal Institute of Philosophy*, 13(37), 73-76.

Montoya, E. R., Terburg, D., Bos, P. A., & Van Honk, J. (2012). Testosterone, cortisol, and serotonin as key regulators of social aggression: A review and theoretical perspective. *Motivation and Emotion*, *36*(1), 65-73.

Naar, H. (2015). Real-world love drugs: Reply to Nyholm. *Journal of Applied Philosophy*, 33(2), 197-201.

Nave, G., Camerer, C., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science*, 10(6), 772-789.

Outram, S. M. (2010). The use of methylphenidate among students: the future of enhancement? *Journal of Medical Ethics*, 36(4), 198-202.

Pacholczyk, A. (2011). Moral enhancement: what is it and do we want it? *Law, Innovation and Technology*, 3(2), 251-277.

Parens, E. (2013). The need for moral enhancement. *The Philosophers' Magazine*, 62, 114-117.

Perlmutter, J. S., & Mink, J. W. (2005). Deep brain stimulation. *Annual Review of Neuroscience*, *29*, 229-257.

Perry, B., Toffner, G., Merrick, T., & Dalton, J. (2011). An exploration of the experience of compassion fatigue in clinical oncology nurses. *Canadian Oncology Nursing Journal*, 21(2), 91-97.

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.

Persson, I., & Savulescu, J. (2011). The turn for ultimate harm: a reply to Fenton. *Journal of Medical Ethics* 37(7), 441-444.

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

Persson, I., & Savulescu, J. (2013). Getting moral enhancement right: the desirability of moral bioenhancement. *Bioethics*, *27*(3), 124-131.

Persson, I., & Savulescu, J. (2014). Should moral bioenhancement be compulsory? Reply to Vojin Rakic. *Journal of Medical Ethics*, 40(4), 251-252.

Persson, I., & Savulescu, J. (2016). Moral bioenhancement, freedom and reason. *Neuroethics*. Available online ahead of print at http://link.springer.com/article/10.1007/s12152-016-9268-5.

Prinz, J. (2011). Against empathy. *The Southern Journal of Philosophy*, 49(s1), 214-233.

Rabins, P., Appleby, B. S., Brandt, J., DeLong, M. R., Dunn, L. B., Gabriëls, L., ... & Mayberg, H. S. (2009). Scientific and ethical issues related to deep brain stimulation for disorders of mood, behavior, and thought. *Archives of General Psychiatry*, *66*(9), 931-937.

Raus, K., Focquaert, F., Schermer, M., Specker, J., & Sterckx, S. (2014). On defining moral

enhancement: a clarificatory taxonomy. *Neuroethics*, 7(3), 263-273.

Richards, W. A. (2015). Understanding the religious import of mystical states of consciousness facilitated by psilocybin. In *The Psychedelic Policy Quagmire: Health, Law, Freedom, and Society*, eds. J. H. Ellens & B. Roberts, pp. 139-144. Santa Barbara, C.A. and Denver, C.O.: Praeger.

Russell, L. B. (2014). Do we really value identified lives more highly than statistical lives? *Medical Decision Making*, 34(5), 556-559.

Sandel, M. (2007). *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge, MA: Harvard 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. (2014). Autonomy and the ethics of biological behaviour modification. In *The Future of Bioethics: International Dialogues*, ed. A. Akabayashi, pp. 91-112. Oxford: Oxford University Press.

Savulescu, J., Sandberg, A., & Kahane, G. (2011). Well-being and enhancement. In *Enhancing Human Capacities*, eds. J. Savulescu, R. ter Meulen, & G. Kahane, pp. 3–18. Oxford: Wiley-Blackwell.

Schaefer, G. O. (2015). Direct vs. indirect moral enhancement. *Kennedy Institute of Ethics Journal*, *25*(3), 261-289.

Schaefer, G. O. (2011). What is the goal of moral engineering? *AJOB Neuroscience* 2(4), 10–11.

Schelling, T. C. (1984). *Choice and Consequence*. Cambridge, M.A.: Harvard University Press.

Sessa, B. (2007). Is there a case for MDMA-assisted psychotherapy in the UK? *Journal of Psychopharmacology*, 21(2), 220-224.

Shook, J. R. (2012). Neuroethics and the possible types of moral enhancement. *AJOB Neuroscience* 3(4), 3–14.

Singer, P. (2014). Against empathy: commentary by Peter Singer. *Boston Review*. Available at http://bostonreview.net/forum/against-empathy/peter-singer-response-against-empathy-peter-singer.

Singh, I. (2008). Beyond polemics: science and ethics of ADHD. *Nature Reviews Neuroscience*, *9*(12), 957-964.

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, *26*(1), 5-16.

Soler, J., Elices, M., Franquesa, A., Barker, S., Friedlander, P., Feilding, A., ... & Riba, J. (2016). Exploring the therapeutic potential of ayahuasca: acute intake increases

mindfulness-related capacities. *Psychopharmacology*, *233*(5), 823-829.

Sparrow, R. (2014a). (Im)moral technology? Thought experiments and the future of 'mind control.' In *The Future of Bioethics: International Dialogues,* ed. A Akabayashi. pp. 113-119. Oxford: Oxford University Press. Cited page numbers from the online version available at http://profiles.arts.monash.edu.au/wp-content/arts-files/rob-sparrow/ImmoralTechnologyForWeb.pdf.

Sparrow, R. (2014b). Better living through chemistry? A Reply to Savulescu and Persson on 'moral enhancement.' *Journal of Applied Philosophy*, 31(1), 23-32.

Srinivasan, A. (2015). Stop the robot apocalypse: the new utilitarians. *London Review of Books*, *37*(18), 3-6.

Stebnicki, M. A. (2007). Empathy fatigue: healing the mind, body, and spirit of professional counselors. *American Journal of Psychiatric Rehabilitation*, 10(4), 317-338.

Steutel, J., & Spiecker, B. (2004). Cultivating sentimental dispositions through Aristotelian habituation. *Journal of Philosophy of Education*, *38*(4), 531-549.

Synofzik, M., & Schlaepfer, T. E. (2008). Stimulating personality: ethical criteria for deep brain stimulation in psychiatric patients and for enhancement purposes. *Biotechnology Journal*, *3*(12), 1511-1520.

Tennison, M. N. (2012). Moral transhumanism: the next step. *Journal of Medicine and Philosophy*, *37*(4), 405-416.

Thibaut, F., Barra, F. D. L., Gordon, H., Cosyns, P., & Bradford, J. M. (2010). The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of paraphilias. *The World Journal of Biological Psychiatry*, 11(4), 604-655.

Thomas, G., Lucas, P., Capler, N. R., Tupper, K. W., & Martin, G. (2013). Ayahuasca-assisted therapy for addiction: results from a preliminary observational study in Canada. *Current Drug Abuse Reviews*, 6(1), 30-42.

Turgay, A. (2009). Psychopharmacological treatment of oppositional defiant disorder. *CNS Drugs*, 23(1), 1-17.

Turner, D. C., Robbins, T. W., Clark, L., Aron, A. R., Dowson, J., & Sahakian, B. J. (2003). Cognitive enhancing effects of modafinil in healthy volunteers. *Psychopharmacology*, 165(3), 260-269.

Wasserman, D. (2011). Moral betterness and moral enhancement. Presented at the 2011 Uehiro-Carnegie Conference, New York, USA.

Williams, C. A. (1989). Empathy and burnout in male and female helping professionals. *Research in Nursing & Health*, 12(3), 169-178.

Winkelman, M. J. (2015). Psychedelic medicines. In *The Psychedelic Policy Quagmire: Health, Law, Freedom, and Society*, eds. J. H. Ellens & B. Roberts, pp. 93-117. Santa Barbara, C.A. and Denver, C.O.: Praeger.

Wiseman, H. (2016). *The Myth of the Moral Brain: The Limits of Moral Enhancement*. Cambridge, M.A.: MIT Press.

Wiseman, H. (2014). SSRIs as moral enhancement interventions: a practical dead end. *AJOB Neuroscience*, 5(3), 21-30.

Wudarczyk, O. A., Earp, B. D., Guastella, A., & Savulescu, J. (2013). Could intranasal oxytocin be used to enhance relationships? Research imperatives, clinical policy, and ethical considerations. *Current Opinion in Psychiatry*, 26(5), 474-484.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, 107(15), 6753-6758.

Zenasni, F., Boujut, E., Woerner, A., & Sultan, S. (2012). Burnout and empathy in primary care: three hypotheses. *British Journal of General Practice*, 62(600), 346-347.

Zohny, H. (in press). A defence of the welfarist account of enhancement. *Performance Enhancement & Health,* in press. Available online ahead of print at http://dx.doi.org/10.1016/j.peh.2015.09.002.