

Neuromodulation in the Service of Moral Enhancement

Bernard Baertschi

Received: 19 November 2012 / Accepted: 20 December 2012 / Published online: 29 December 2012
© Springer Science+Business Media New York 2012

Abstract Human enhancement is a much debated topic in the bioethical literature. Human beings have long tried to improve their capacities and their performances through training and with the aid of tools; but more recently new means have come to the fore, such as drugs and biotechnological devices, especially in the domain of bodily strength and cognitive powers. Moral enhancement has been more seldom discussed. However, this question has recently been hotly debated between two philosophers, Thomas Douglas and John Harris. Douglas claims that modulating certain ugly emotions *directly* would consist in moral progress—*directly*, that is, without using cognitive means like persuasion or deliberation. Harris makes three objections against this thesis: such a direct neuromodulation would be inefficacious, would put our liberty in jeopardy and would lead to a moral decline. In this paper, I examine the third argument: with direct modulation, we risk intervening too much or too little, inducing an inappropriate emotion or an inappropriate level of an otherwise appropriate emotion—two upshots that will put morality in jeopardy. I conclude that the validity of this objection depends on several meta-ethical positions: if you are a rationalist or think that intentionality and consciousness are at the core of morality, you will agree with Harris, but if you are a sentimentalist or someone for whom results count, you will disagree. Here as elsewhere, ethical questions cannot be divorced from meta-ethical ones.

Keywords Neuromodulation · Moral enhancement · Thomas Douglas · John Harris · Emotions · Cognition

Introduction: The Special Case of Moral Enhancement

Human enhancement is a much debated topic in the bioethical literature. Human beings have long tried to improve their capacities and performances through training and with the aid of tools; but more recently new means have come to the fore, such as drugs and biotechnological devices, especially in the domain of bodily strength and cognitive powers. Doping and cognitive enhancers have received a very divisive reception, from downright rejection to enthusiastic promotion. Moral enhancement has been more seldom mentioned, with the exception of mood enhancement, a state of mind that has some ethical impact. Is it for the reason that it is less controversial? Maybe, because, contrary to the case with much bodily and cognitive enhancement, the fact that somebody takes a moral enhancer does not put anybody else in a worse position: if I become morally better through swallowing a moral pill, the people who have relations with me may even benefit. Morality is usually not a competitive context with winners and losers. Nevertheless, the actual reason probably lies elsewhere: it exists rather in the fact that we do not have a clear idea of what exactly ‘morally better’ means and, therefore, what are the appropriate means to reach such a state. Contrary to the case with cognitive enhancement, where the goal is known and its content determinate—to have a better memory consists in being able to store more information and to retrieve it more easily—with moral enhancement, the goal has no precise content. It consists in becoming more moral, but what does it mean exactly?

B. Baertschi (✉)
Institute for Biomedical Ethics, University of Geneva, Geneva,
Switzerland
e-mail: Bernard.Baertschi@unige.ch

B. Baertschi
CMU, 1, Rue Michel-Servet, 1211 Geneva 4, Switzerland

However, this is still not the whole truth, because really we have some idea of the end and of the means: to be more generous is to be morally better, and education has been and still is the classical means of creating less egoistic attitudes among children. A lot of situations are indeed less clear. Would improving trust be a sign of moral progress or would it only create more opportunities for manipulating people through enhancing credulity? For this case, we already have a drug at our disposal: oxytocin (Churchland 2011, pp. 75–78). But the fact that there exist fuzzy situations is not a good reason to claim that we have no notion of what ‘morally better’ consists of. So, why has moral enhancement been little discussed to date, despite its importance? In my opinion, the main reason is that many authors deem that such an enhancement is not a possible option. No drug or biotechnological device can afford it. The argument is rather straight. Imagine that being more trusting would consist in moral progress and that oxytocin improves trust; nevertheless, improving trust through oxytocin would not be moral progress, because nothing moral can be achieved through a means that bypasses conscious efforts and cognition.

In this sense, morality is like climbing. Climbing a mountain is not the same as reaching its top. It depends on the means used. If a helicopter drops you off at the summit of the Matterhorn and, after returning, you tell your friends that you have climbed this beautiful mountain, they will laugh at you.¹ In the same manner, if you claim that you have become more moral by sniffing oxytocin, you will not be considered as having performed a moral feat. Conscious efforts and cognition are required as necessary ingredients in moral improvement.

Of course, this requirement leaves the door open to cognitive enhancers with a moral impact. For example, if a drug could lessen our irrationality concerning long-term decisions, it would aid morality, but enhancing morality *directly* is not possible. This view is very traditional and is seldom challenged. Nevertheless, some authors have recently claimed that direct interventions (i.e. interventions that by-pass cognition) are not necessarily precluded. Neurological research tends to show that morality has its base (or one base) in emotions, such as empathy, fear or anger. Psychopaths are unable to feel empathy and they act strikingly immorally; if it were possible to induce empathy in them, they would surely become less immoral persons: their morality would be enhanced. Morally speaking it would be a good thing, for them as well as for us. Emotions can be modified through cognitive means: if I realise that the dog that is barking at me is a Yorkshire terrier, my fear will disappear. But they can be modified through other means, like drugs, i.e. through neuromodulation,

by-passing cognition.² From a moral point of view, we can concede to the traditional view that it is better to alter our emotions through conscious and cognitive means, but what if they resist our efforts? Would neuromodulation be morally worthless? This question has recently been the subject of hot debate between two philosophers, Thomas Douglas and John Harris. In the next section, I will examine their disagreement and the underlying reasons. To my mind, each has a point, and I will develop arguments in the following sections to show that neuromodulation has a genuine moral worth in some situations, but that it could not be a universal means of moral progress if we stick to the traditional meaning of this expression.

The Debate Between Thomas Douglas and John Harris

In 2008, Thomas Douglas published a paper entitled ‘Moral Enhancement’, in which he claimed that modulating certain emotions *directly* would be an act of moral progress—*directly*, that is, without using cognitive means like persuasion or deliberation. To avoid the objection that there exists no agreement on what immorality consists in, he chose two ugly emotions or dispositions: racial aversion and aggressiveness. These emotions are motives for evil actions as they cause immoral behaviour. Behaviours prompted by these emotions are immoral because people are wronged through them, and it is a basic requirement of every moral doctrine that we should not wrong people; therefore, there is no difficulty in agreeing that such emotions are ugly and that it would be morally better if we could mitigate them.

In his 2008 paper, Douglas states his thesis and answers possible objections based on bioconservative arguments: we should accept what is given to us (our emotional equipment), we ought to improve ourselves through educative means and biotechnological means are unnatural. I think that his answers are persuasive, but I will not dwell on them, because I am interested in another group of objections—those formulated later by John Harris.

The situation is *prima facie* paradoxical, because Harris is well-known for his very positive evaluation of human enhancement (see for example Harris 2007); but, as we will see, that is precisely what makes his objections insightful and relevant. Harris has formulated them in a 2011 paper entitled ‘Moral Enhancement and Freedom’, and Douglas summarised them in his answer the same year (Douglas 2011). Three main objections are set out: direct

¹ I borrow this example from Schermer (2008, 360).

² Pharmacological means are not the only tools available for neuromodulation—think of Transcranial Magnetic Stimulation (TMS) or Deep Brain Stimulation (DBS). However, the problems they raise are essentially the same for the debate I will examine; consequently, pharmacological means will constitute my paradigm.

neuromodulation of ugly emotions would be inefficacious, would put our liberty in jeopardy and would lead to a moral decline. I will make some quick comments on the two first objections, and then examine the third and its impacts in greater detail.

The inefficiency objection could be dismissed in three words: wait and see! (or: try and see!). It is indeed only when we can point to an impossibility that we should dispense with useless attempts, and this is not the case with the modulation of emotions: it is possible, since it has already been realised with oxytocin and in the treatment of nervous breakdowns. Of course, the problem of knowing if it leads to genuine moral enhancement remains, but this is the gist of the third objection. Leaving this aside for the moment, it is nevertheless useful to elaborate a little on the reply to the inefficiency objection. Neuromodulation could fail, but traditional means already fail completely in some situations. Douglas observes that such is the case with some ugly emotions, for the reason that they have no cognitive basis. In these cases, there exists unfortunately no target for traditional means of moral reform. This happens sometimes with racism: ‘Psychologists and neuroscientists study certain deeply ingrained fear- or disgust-based responses towards people of different race that seem aptly characterized as variants of racial aversion (they are often described as *implicit racial bias*) but which entail the presence of no false, stereotypic beliefs.’ (2011, p. 5) The persons affected by those emotional responses do not entertain any racist belief: their emotions are without a cognitive basis, i.e. a belief that foreigners are dangerous or inferior. They have correct beliefs, but do not feel the corresponding emotions. Here, traditional means of reform are completely inefficacious; neuromodulation cannot be worse and could probably do better.

If neuromodulation is efficacious it will endanger our freedom, says Harris, because it could be used to change the emotional responses of someone without his consent or even against his will. However, if neuromodulation could be employed to diminish freedom, it could be used to enhance it, too. ‘Suppose that I undergo an intervention that mitigates some of the many emotional biases that afflict my prudential and moral reasoning, or that reduces my temptation to act against my sincere normative judgements. This intervention seems aptly characterized as *increasing* my freedom to be moral by removing a brute constraint on that freedom’ (2011, p. 7), says Douglas, and he is right. A person who suffers from racist emotions that run against his beliefs will be relieved by the silencing of these emotions; a burden will be removed from him and he will be more free in his relations with foreign people. As is clear, this

question regards freedom of mind and of action, and has nothing to do with the free will debate.

The moral decline objection is the more serious one for Douglas, and I agree with him. But what does it mean exactly? Remember that Douglas’ topic is moral emotions. As is well-known from Aristotle on—Douglas and Harris concur—in order to be morally appropriate, an emotion must be felt at the right level. It is a happy medium: neither too much, nor too little. If you feel too much fear, you will probably manifest cowardice in your behaviour, but if you feel too little you will evince foolhardiness. You will only be courageous—the morally correct attitude—if you feel fear at the right level, somewhere in the middle. In the case of the emotions studied by Douglas, racial aversion and aggressiveness, they must be at least lessened: ‘As Harris rightly notes, the emotion would have to be attenuated to *the right degree*. This is because, as Harris puts it, “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”’ (2011, p. 7), observes Douglas. He is not completely convinced by the argument because, he continues, ‘it is not clear that xenophobia is, in normal circumstances, at all conducive to morality’, but he concedes the main point. Before going on, I will add a comment in passing: xenophobia is not a counter-example, because it is not a generic emotion but a form of hate (and/or fear) already in excess: there exist good reasons to be wary of unknown people, but not to hate them. I return to Harris’ argument. For him, in order to reach this ‘right degree’, a degree that is a necessary condition to have a correct moral motivation or to improve it, ‘what is necessary [...] is the *fine tuning* of certain emotions in a person-specific way that is sensitive to prevailing circumstances’ (Douglas 2011, p. 7), and he doubts seriously that this could be possible through direct modulation. We risk intervening too much or too little, inducing an inappropriate emotion or an inappropriate level of an otherwise appropriate emotion.

Douglas concedes this point, but replies that it is sometimes worth running the risk in the face of the ugliness of these emotions. We must nevertheless be cautious, because our past history shows numerous examples of abuses backed by good intentions (think of lobotomy). Formulated in this manner, Douglas’ proposal is very modest and Harris’ objection rather easy to satisfy. Harris himself seems sometimes to avow that his refusal is only a matter of feasibility, such as when he says: ‘Of course if ever we do have the prospect of such precise and unequivocally good producing interventions, I will welcome them’ (2011, p. 105). Actual feasibility would cancel the risk of moral decline.

Sentimentalism and Rationalism

However, the disagreement between both authors is much deeper than is shown by the preceding discussion, and amounts to a meta-ethical one. Each entertains a different conception of ethics, a sentimentalist one for Douglas and a rationalist one for Harris. The difference explains why Harris insists that emotions should only be acted upon through cognitive means, i.e. indirectly, whereas it is not a requirement for Douglas. In the papers I have examined, the difference does not clearly surface, except perhaps in this observation from Harris: ‘The most important thing about the prejudices that most, perhaps all of us, have in one form or another, is to recognize them and learn to be ashamed of them and above all not to act on them’ (2011, 105). If I have racist beliefs, and if I feel racial aversion, the requirements of morality are above all to refrain from acting on their basis, and not to change my beliefs or to mitigate or eradicate my ugly emotions. Of course, it is better if I can, but not *morally* better, since morality is foremost a matter of action and behaviour. In other words, I behave morally if I act for the right reasons, and not if I follow the motives I feel or the false beliefs I have. Reason tells me not to discriminate against foreigners, not to wrong them, whatever emotions I feel and whatever beliefs I have. In this sense, emotions—the topic I am interested in—do not matter morally.

However, this is not the whole story. In order to be motivated well, it is of course better to feel correct emotions, and consequently it could be desirable to change our emotions so that we mitigate or raise them. In order to succeed at this task, we dispose of cognitive means. But what if it would be possible to reach this result also through direct neuromodulation? For Harris, the result will not count as a moral improvement at all: the only means that possess moral worth are cognitive, i.e. proceed through reason. Here, we get to the climbing-a-mountain argument I mentioned in my introduction: being dropped off by a helicopter at the top of a mountain does not count as climbing, and in the same manner mitigating an ugly emotion by neuromodulation does not count as a moral improvement.

It is easy to see that this claim buttresses the moral decline argument in a second sense—or constitutes a second argument. If traditional means of moral enhancement were replaced by non-cognitive ones, this would lead to moral decline. In the preceding section, we saw Harris’ objection that direct neuromodulation is less sensitive than cognition in order to modulate emotions in a fine-grained manner; now we see that it will restrict the field of morality and replace it by a kind of social engineering.

For Douglas, the situation is of course not quite the same: if he would probably agree that morality is fore and

foremost a question of action and behaviour, he would not accept that reason furnishes the only genuine moral motives. Ultimate moral motives are for him of an emotional nature, so that we cannot act morally if we do not feel the right emotions. Consequently, the use of any mean of modifying some ugly emotion that proves efficacious constitutes a genuine moral enhancement.

Douglas is never explicit concerning his meta-ethical position—in this sense, what I have said is a kind of reconstruction—but Harris is of the same mind as me when he observes that ‘Douglas’s problem seems to me to be that he equates moral enhancement with having the right feelings’ (2012a, p. 4). Harris himself is quite clear on his own position, especially in some papers he has written with Sarah Chan. In one of them, the authors acknowledge that reason and emotions play their role, but not on a par: ‘Although both emotion and reasoning thus affect moral decisions, of the two, it must be reasoning that pulls in the direction of morality’ (Chan and Harris 2011, p. 130). As we have just seen, to make a moral decision is to decide upon reason, not upon emotion, because we must weigh the various values at stake and take into account the many factors that shape the context of our action. In such a context, some emotions will be awakened and will motivate the action in a certain direction, but not always the right one. To know if it is the right one, we must reflect on the emotions to see if they are adequate in order to promote the values we want to. To illustrate this point, the authors mention a study conducted by Molly Crockett and colleagues where subjects have been given citalopram, a drug that modulates serotonin (Chan and Harris 2011, pp. 130–131). In the *Ultimatum Game*, the subjects were more prone to accept unfair offers, because citalopram heightens their aversion to cause harm. Now, to refuse an offer, even if it is unfair, is to harm the offerer. In such a case, observe Harris and Cham, modulation of emotion through citalopram tends to discount reasoning and rational attachment to moral values like justice. In one word, it is moral *de-enhancement*. Here, we meet the moral decline argument again.

It follows, as Harris states elsewhere, that a person who feels that he is not sufficiently altruistic ‘knows what’s right and needs no moral enhancement, for these purposes at least. He may need some help in feeling sympathy but that is a different problem and not, I would suggest, a moral one’ (2012a, p. 4). Modulating emotions is then devoid of any moral worth—a rather extreme rationalist claim—and Harris ends this paper recalling the motto of Socrates, that a meaningful life cannot be lived without examination of oneself, concluding that ‘only self-conscious reflection on conduct can deliver answers to the question as to whether what one feels is right is indeed right. It is only such an examination, and the resolve to put its conclusions into

effect, that constitutes a moral life and, a fortiori, a morally enhanced life' (2012a, p. 5). Emotions are in need of justification, they do not justify themselves, because the gut is not an organ of thought. This claim belongs to the last quotation I want to offer in order to underline the rationalist stance adopted by Harris in his ethics: 'To believe that emotions can deliver answers to moral dilemmas or generate moral judgements is like believing that the gut is an organ of thought, or one that can answer complex, combined theoretical and empirical, questions' (2012b, p. 4). If we replace 'deliver' by 'rightly motivate', we will give more punch to his thesis, because nobody has ever claimed that emotion could deliver answers, even an extreme sentimentalist.

Emotions and Reason in Moral Judgements

The upshot of this debate seems to be that if you are not a steady opponent to human enhancement, you will accept direct neuromodulation as a tool for moral improvement or not, depending on your meta-ethical position. If you are a rationalist like Harris, following authors like Charles Larmore when he states that 'morality is not possible except for beings that can respond to reasons' (2010, p. 2), you will not, but you will perhaps accept cognitive neuroenhancement; if you are a sentimentalist, like Douglas, you will accept both. If this diagnosis is correct, then the desirability of by-passing reason in some situations will depend on a meta-ethical thesis. Hence a new question: which of these theses is right?

Nowadays, sentimentalism is very widespread, especially in neuroscience and neuroethics. Jonathan Haidt and Joshua Greene, for instance, clearly defend such a position. In order to do that, Greene has devoted a lot of time and paper to the well-known *Trolley Problem*, where a trolley is running downhill and will kill five workers. In a first version, the *switch dilemma*, it is possible for a bystander to hit a switch that will divert the trolley onto a side-track where only one worker stands. On a second version, the *footbridge dilemma*, it is possible to save the five workers' life too, but only in pushing a fat man off a footbridge in order to stop the trolley. The upshot of the two versions is the same: we can save four lives (the five workers minus the person killed), but when we ask people if they think that it is permissible to switch the trolley and to throw the fat man, the answers are very different. In the first case, almost 90 % think it is permissible, whereas in the second only 10 % agree (Hauser et al. 2007).

This astonishing result has prompted a lot of research and reflection. There exist of course several ways to interpret the discrepancy between the reactions. For Greene, the best explanation is given by a dual-process

theory: 'Moral judgment is the product of both intuitive and rational psychological processes, and it is the product of what are conventionally thought of as "affective" and "cognitive" mechanisms' (Cushman et al. 2011, p. 48). In such a theory, there exist two types of processes that issue a moral judgement. One is emotional, that is intuitive, automatic and non-voluntary: it prohibits harming a person through forceful physical contact, and therefore motivates the judgement that it is not permissible to throw the fat man. The second is rational and conscious, motivating the judgement that it is better that only one person dies rather than five, and therefore it is permissible to switch the trolley. If this second process is not activated in the footbridge dilemma, it is because emotion pre-empts reasoning here (see also Greene et al. 2009).

People are generally not aware of the emotional process; when they are asked to explain the motivation behind their decision not to harm the fat man, they make up reasons, but of course post hoc reasons. As these reasons are disconnected from the real causes of their judgements, Greene speaks of rationalisation, and even of confabulations: 'We are all confabulators of a sort. We respond to the conscious deliverance of our unconscious perceptual, mnemonic, and emotional processes by fashioning them into a rationally sensible narrative, and without any awareness that we are doing so' (Greene 2008, p. 63). The real causes are emotions. They act independently of reason. This does not mean that they are devoid of any cognitive basis, because they are triggered by the perception of a potential harm, but they are not sensitive to rational considerations. However, Greene contends, this is not completely true; he thinks that we have the possibility to become more rational, and therefore to overcome our emotional reaction, even if he is not very forthcoming on how we are to do this (Greene et al. 2004 p. 398). Would Greene eventually become a rationalist, too? Not exactly, because he claims to be a follower of David Hume, a famous sentimentalist, in thinking that the rational process of weighing the goods and evils that is at the basis of rational morality is grounded in emotions too, but emotions of another kind. He is presently devising experiments to put this into a better light (Cushman et al. 2011, sect. 4).

As we can see, if the controversy between proponents of direct neuromodulation and their opponents can be framed as a disagreement between sentimentalists and rationalists, it can also be framed inside the sentimentalist approach, in distinguishing between several kinds of emotions, some in line with reasoning and the others being more or less impervious to it. For Greene, as for Harris, the imperviousness never seems to be complete; therefore, there seems always to be a way to modulate these emotions through cognition and reflection. Does it mean that Greene would agree with Harris' position, in that direct neuromodulation

could lead to moral decline? Not certainly, because as a strict utilitarian, he thinks that results count morally and could be tempted, like Peter Singer, to judge that VMPC patients, i.e. patients with a brain lesion that prevent emotions from impinging on rationality, could in certain situations make better moral judgements than ‘normal’ people usually do—‘better’ should obviously be understood in the consequentialist sense of the term, i.e. in focusing on outcomes rather than on motives or reasons to act. These patients are justly characterised by a dissociation between reason and emotions, due to a brain lesion, and they are less reluctant than us to push the fat man (Koenigs et al. 2007).³ If a brain lesion can issue in a moral improvement, it is of course without the mediation of reason. It is indeed rather counterintuitive and controversial to suppose that a brain lesion can issue in such an improvement, but all that I want to insist on is the bypassing of reason it implies. Consequently if direct neuromodulation can result in moral enhancement, it would certainly be welcomed by Greene.

The Moral Importance of Means versus Ends

The debate between Douglas and Harris was about motivation and action, whereas Greene’s reflections concerned judgements. This difference is nevertheless irrelevant, because judgements about what we ought to do, or what is permissible, are backed by motives. The discussion of Greene’s position has brought to the fore another distinction that is more important—the distinction between results or ends and means. If only results count morally, then neuromodulation will be welcome if it is efficacious, but if means also count, it will perhaps not be. The controversy between Douglas and Harris could be read along these lines, too, as I have already hinted at. For Harris, even if neuromodulation is efficacious, it will not count as a *moral* enhancement. Two readings can be given for this thesis:

1° Means count morally, and a good end should not be realised through evil means (this claim for instance is made in the well-known doctrine of double-effect).

2° Means determine ends. If certain means are chosen, they will change the meaning of the end. Think of the climbing-a-mountain argument I have spoken of.

³ As an anonymous reviewer rightly observes, if we interpret the actions of these patients from a moral viewpoint that is not strictly utilitarian, the situation becomes more complicated, because there is no evidence that these patients are able to empathize with the imaginary victim and to judge that each of the two outcomes is bad because at least one person will be killed or that one outcome is worse than the other in terms of how many people will be killed. In consequence, empathy and the capacity to recognize when a person can be harmed could play no role in their judgement.

With the first reading, the choice of an inappropriate means deprives the end of its moral worth, and the action becomes immoral; with the second, the action could become immoral, but not necessarily so; nevertheless, it changes its nature: it is not an action of the same type.

The second reading is the relevant one for the question addressed here. It has been adopted by bioconservatives. Leon Kass, for example, asks: ‘Where is biotechnology taking us? What does it mean for our humanity? What kind of people do we wish to be and what sort of a society do we wish to become?’ (2005, p. 225). Means (biotechnology) and ends (‘the kind of people we want to be’) are intermingled. Will Harris be of the same mind? Not exactly, because he would probably accept enhancement through neuromodulation for socially welcomed traits even if it is not a moral enhancement, because for him, as a moderate utilitarian, results also count. Following this thread, could we not say that we should not mind if an improvement is labelled moral or not as long as it is an improvement, something that is desirable from a human and social point of view? Even if pro-social attitudes are not necessarily moral, depending on how you understand this expression, they remain desirable.⁴ Besides, Harris himself seems to be of this mind and is willing to quit the strict moral point of view when he states that social and political means are to be used in order to fix some very important moral problems like global poverty: ‘If we couple universal education with the eradication of poverty and more, increasing affluence, we will I believe be doing the most promising thing as far as moral enhancement goes’ (2012b, p. 6). Such means bypass personal cognition and rationality, of course; they rather belong to social engineering.

Following this thread does not mean that any means to change our emotions are acceptable. Some training, in the guise of *A Clockwork Orange* or through physical violence, is clearly not, because it violates values we cherish, but I think there exist situations where neuromodulation is quite acceptable. First, for views that give a moral role to emotions in motivation, be it a dual-process theory or not, some emotions motivating immoral behaviour could well be immune from rationality, and we could be wise to alter them directly, if we can. Notice that in so doing we do not necessarily by-pass any cognitive means, as in taking a pill, since some kinds of neuromodulation are cognitive, but in an indirect way. Think of neurofeedback, where computer games are used: of course, we cannot play without using our cognitive powers.⁵ Second, as Douglas has underlined, some emotions, even if they are not problematic in

⁴ We could say that pro-social attitudes are moral from the point of view of society, but not necessarily from the point of view of the agent. For that, cognition would be necessary.

⁵ See the paper by Niels Birbaumer in this volume.

themselves, can be felt at inappropriate levels which resist reform through cognitive means.

Both situations include degrees. Rational influence is never excluded and we are able to modulate consciously in ourselves and in our children appropriate emotional responses. In this sense, as traditional morality has taught us, we can fight against *vice*. But sometimes the vice continues to escape our rational powers and our conscious will; in this sense, it becomes a *sickness*. Of course, it is not always the case. In *The Scarlet Letter*, the famous novel written by Nathaniel Hawthorne, Hester Prynne hates her husband, and she is aware of this ugly emotion; the author observes: ‘She upbraided herself for the sentiment, but could not overcome or lessen it’ (2008, p. 98). In this case, there is no need to take medicine, voluntary moral reform is all that is needed—and maybe acceptance of this emotion, because her husband is really hateful. Our attitude in front of vices and sicknesses is very different, as Tristram Engelhardt observes: ‘If the individual is placed in a full-fledged sick role, the sick person is usually held not to be responsible for being in that role, is excused from social duties that the illness impairs, and is enjoined to seek treatment from a set of individuals socially recognized as appropriate therapists’ (1997, p. 217). If we have other expectations and demands, it is, Engelhardt adds, for the reason that ‘medicine focuses primarily on “caused” phenomena, and ethics on “chosen” ones’. In the case of sickness, we perceive consequently no difficulty in using means that by-pass conscience; they do not meet with objections. If a vice is a condition we should reform, sickness is a condition we should heal, and neuromodulation seems to be well suited as a form of healing.⁶ Since the distinction between vice and sickness is only one of degree, as I have said, direct neuromodulation would also be well suited at least for some kinds of moral difficulties. However, the claim that there exists a continuum between vice and sickness raises many difficulties, in particular because, as Engelhardt notices, we react differently to them, socially and legally. Consequently, we tend to see them apart, as two separate conditions. In my opinion, the problem of direct neuromodulation in morality would be a good opportunity to revise some of our social and legal practices, but this is the topic of another paper.

We arrive at an analogous conclusion if we reflect on personal responsibility and integrity. What characterises human actions is intentionality, i.e. the property to be

directed at an aim or an end. Without intentionality, an action is not an action, but a physical movement: it only happens. Intentionality is linked with responsibility, because I am responsible for what I do when I aim consciously at a goal. However, there exist some conditions that prevent me from being responsible. If I am coerced or if my mind is badly confused, my responsibility can be mitigated, and even cancelled. In other words, I am not fully responsible if the motives I act upon are not mine. ‘We can be genuine authors of our actions if we are not impeded in consciously deliberating, choosing, and acting. If our motivational states are not compelled, coerced, or manipulated by artificial means, and if we have the capacity to reflect on, identify with, and execute these mental states in choices and actions, then we can act freely’ (2007, p. 55), says Walter Glannon. Emotions are, as we know, amongst the motives of action. The racist feels racial aversion and is prompted to act upon it. But, as Douglas has correctly observed, there exist persons who feel such aversion but reject it because it does not fit with their moral norms or values. In other words, these emotions are motives that these persons cannot acknowledge as theirs; they are like foreign elements put into their minds. Some cases are clearly pathological—think of obsessive-compulsive disorders (OCD)—but others are not; they just manifest a moral difficulty. Wouldn’t it be a good thing if, through direct neuromodulation, we could mitigate or silence these emotions? It would be a clear benefit for liberty, as Douglas has said, and for the integrity of the persons affected—two important values for morality. Of course, gains in liberty and in integrity do not guarantee moral progress, in the sense that if the liberty and integrity of wrong-doers are enhanced, they will be more noxious for their fellow men. But in another sense, to become more able to act intentionally is genuine moral progress; it enhances responsibility. This reminds us that ‘moral enhancement’ has two senses. First, it designates better motives and better outcomes: there is moral enhancement if there is more good and less evil in our actions and in the world (a world with less racism and racist motives is a better world). Second, it applies to a better capacity to act from a moral point of view. Ugly emotions are a hindrance for both, and therefore neuromodulation can be an aid for both, too.

Conclusion

I said in my introduction that neuromodulation could produce genuine moral worth in some situations, but that it could not be a universal mean of moral progress. Douglas is therefore right in his claim. But Harris has a point, too, as far as he insists correctly on the role of reason and

⁶ A difficult question regards *involuntary* modulation. If results count and if their pro-social character can justify direct modulation, then it can also justify involuntary modulation, as a part of education or of punishment. Some convicts are already offered the opportunity to follow a cure—sometimes a harsh one—and consequently to have a shorter sentence, but it is voluntary. Would it be acceptable to force them into direct neuromodulation?

deliberation: morality consist for an important part in their use. Through this statement he thinks he utters a critique of Douglas; I am not sure that he is right, because Douglas would concede much of what he says. However, this point is not essential for me. What is important in my mind is that morality is a matter of reason *and* emotions. Each has its role. Recent neuroscience has insisted upon the importance of emotions, showing in particular that empathy is a component—maybe a necessary one—of conscience or moral sense. Consequently, some authors have hypothesised that psychopaths, and even some autistic patients, could not be full participants in our moral community. Stephen Morse reports that ‘Paul Litton denies that psychopaths are rational at all because they lack any evaluative standards to assess and guide their conduct’ (2008, p. 209). It is true that psychopaths have a big problem with morality, but this fact should not hide another from us: empathy may be necessary for morality, but it cannot be sufficient. Reason is another necessary component of it, because morality consists also in reflection and deliberation. It still consists in evaluating emotions, as Harris reminds us: ‘We will, I believe, always need to use moral reasoning to act as a guide to our emotions and as a way of checking that we are having appropriate feelings in appropriate circumstances and for appropriate objects’ (2012a, p. 4). Moreover, emotions can incorporate some reason; i.e. they motivate not only as a (blind) cause.⁷

This duality of reason and emotion should not be equated with the dualism of dual-process theories which are now so often offered as explanations of our moral judgments and motives. I would say that it is not an ‘either-or’ question, but an ‘and’ one. I think that we have now a lot of data which point in this direction and have substantiated claims such as, ‘Determining whether an action is moral or not does not rely only on affective sensitivity. Determinations of moral relevance also require the capacity to integrate a representation of the mental states of others together with the consequences of their actions’ (Decety et al. 2012, p. 217). Even if emotions are one of the roots of morality, they are not alone and, moreover, the roots are far from being the whole plant. This was already Antonio Damasio’s conclusion in *Descartes’ Error* (1994, chap. XI).

Emotions and reason are therefore on a par in the domain of moral enhancement, and if we accept improvement of the second through direct cognitive enhancement (like drugs), we should not refuse emotional enhancement through direct neuromodulation when emotions do not respond to reason, even if conscious efforts and training should continue to play a prominent role in both in order to fight cognitive bias and bad habits. Of

course, both enhancements pose normative problems, moral and conceptual: it is notably difficult to determine the threshold separating normality from pathological condition, and this has a moral relevance, because it is always easier to justify an intervention when a condition is medically labelled. My conclusion is only that an absolute prohibition of direct neuromodulation is not justified, because of the role emotions play in our moral life, as the examples afforded by Douglas have shown, and even if it remains sometimes hard to reach an agreement concerning which emotions and dispositions are ugly and, more generally, what it means to become more moral.

References

- Chan S, Harris J (2011) Moral enhancement and pro-social behaviour. *J Med Ethics* 37:130–131
- Churchland P (2011) *Braintrust*. Princeton University Press, Princeton
- Cushman F, Young L, Greene J (2011) Our multi-system moral psychology: towards a consensus view. In: Doris J, Harman G, Nichols S, Prinz J, Sinnott-Armstrong W, Stich S (eds) *The Oxford handbook of moral psychology*. Oxford University Press, Oxford, pp 47–71
- Damasio A (1994) *Descartes’ Error: emotion, reason and the human brain*. Putnam’s Sons, London
- Decety J, Ickes W (2011) *The social neuroscience of empathy*. MIT Press, Cambridge
- Decety J, Michalska K, Kinzler K (2012) The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. *Cereb Cortex* 22:209–220
- Douglas T (2008) Moral enhancement. *J Appl Philos* 25:228–245
- Douglas T (2011) Moral enhancement via direct modulation: a reply to John Harris. *Bioethics*. doi:10.1111/j.1467-8519.2011.01919.x
- Engelhardt HT (1997) *The foundations of bioethics*. Oxford University Press, Oxford
- Glannon W (2007) *Bioethics and the brain*. Oxford University Press, Oxford
- Greene J (2008) The secret joke of Kant’s soul. In: Sinnott-Armstrong W (ed) *Moral psychology*, vol 3. MIT Press, Cambridge, pp 35–79
- Greene J, Nystrom L, Engell A, Cohen J (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44:389–400
- Greene J, Cushman F, Stewart L, Lowenberg K, Nystrom L, Cohen J (2009) Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition* 111:364–371
- Harris J (2007) *Enhancing evolution. The ethical case for making better people*. Princeton University Press, Princeton and Oxford
- Harris J (2011) Moral enhancement and freedom. *Bioethics* 25:102–111
- Harris J (2012a) ‘Ethics is for Bad Guys!’ Putting the ‘Moral’ into Moral Enhancement. *Bioethics*. doi:10.1111/j.1467-8519.2011.01946.x
- Harris J (2012b) Moral progress and moral enhancement. *Bioethics*. doi:10.1111/j.1467-8519.2012.01965.x
- Hauser M, Cushman F, Young L, Kang-Xing R, Mikhail J (2007) A dissociation between moral judgments and justification. *Mind Lang* 22:1–21
- Hawthorne N (2008) *The scarlet letter*. Tark Classic Fiction, London
- Kass L (2005) Reflections on public bioethics: a view from the trenches. *Kennedy Inst Ethics J* 15:221–250

⁷ In particular, it is true of empathy. See Decety and Ickes 2011.

- Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446:908–911
- Larmore C (2010) Reflection and morality. *Soc Philos Policy* 27:1–28
- Morse S (2008) Psychopathy and criminal responsibility. *Neuroethics* 1:205–212
- Schermer M (2008) Enhancements, easy shortcuts, and the richness of human activities. *Bioethics* 22:355–363