

6

ETHICAL REQUISITES FOR NEUROENHANCEMENT OF MORAL MOTIVATION

Francisco Lara

Abstract: No agreement exists among ethical theories on what can count as a right moral motivation. This hampers us from knowing whether an intervention in motivation biology can be considered positive for human morality. To overcome this difficulty, this paper identifies minimal requirements for moral enhancement that could be accepted by the major moral theories. Subsequently four possible scenarios are presented where the most promising neural interventions on moral motivation are implemented, by means of drugs, electromagnetic stimulation of brain, or biotechnological brain implants. The ultimate goal of this paper is to evaluate the results of each one of these interventions according to their capacity to fulfill the identified requirements.

Keywords: *human enhancement, moral motivation, neuroethics, enhancement, moral enhancement, neuroenhancement, bioenhancement.*

The latest advances in neurological sciences provide extensive information on the biological bases of morality, the field of human behavior connected with prescriptive judgments about what is good and right. These advances have revived interest in determining whether, with greater development in this field, it would be acceptable to use all this knowledge to enhance human capacities related to judgment and moral action. Such interest has turned into controversy characterized, like those related to the enhancement of other capacities (e.g. physical or cognitive), by confrontations between extreme and speculative positions. On the one hand, positions

often oscillate between either the appeal to an untouchable human essence¹ or an unlimited rational capacity to manipulate what is natural.² On the other, participants in these debates tend to base their positions on highly theoretical arguments that are far beyond the real possibilities and reasonable risks of the interventions.³ These aspects of the controversy have served only to polarize it between the optimism of the bioenhancers, who pay little attention to the limitations of interventions of complex and still quite unknown biological mechanisms,⁴ and the pessimism of bioconservatives, who distrust any biological manipulation, no matter how minor, how controlled, or how beneficial it might be.⁵ Furthermore, the particular controversy on moral enhancement suffers from an excessively simplistic perspective, in two ways. First, with respect to the aims. The participants in the controversy tend to position themselves with regard to the question of whether or not we accept to perfect humans morally – of whether to convert them suddenly into individuals who are fully virtuous or who will always do what is correct, something patently unattainable in the short or even long term.⁶ However, the consideration of the means is also simplistic. Bioenhancement is often discussed as though it were homogeneous, without differentiating between different techniques, the peculiarities of which can considerably alter the debate.⁷

To avoid these errors, here I undertake a separate initial ethical analysis, supported by the latest scientific discoveries of each of the currently most viable neural treatments meant to enhance only one aspect of moral agency, specifically that of moral motivation, that is, the influence

¹ Fukuyama (2003); Kass (2002); Habermas (2003); and Sandel (2004).

² Savulescu (2007); Kamm (2005); Agar (2003); and Caplan (2009).

³ Walker (2002).

⁴ Bostrom (2003), (2004), (2005a) and (2005b).

⁵ Elliot (2003); Ida (2009).

⁶ Jotterand (2011) questions whether neuroenhancement can make us virtuous, since we progress morally only through an interaction of emotion and reflexive exercise, something that can never be achieved by mere neural manipulation. Along this line, Fröding (2011) and Schaefer (2011) have argued that thanks to the enhancement in cognitive capacities and the elimination of certain unsuitable tendencies, neural interventions, although they are not themselves intended to make people virtuous, can be of great help to reach this goal. On how some proponents of moral neuroenhancement suffer from overvalued expectations, see Pacholczyk (2011).

⁷ Thus, for example, moral enhancement through genetic manipulation bears aspects of irreversibility and, in some cases, of offspring modification, which is not involved in drug treatment and which requires a separate evaluation for each technique.

that our formed judgments about what is *right* or *wrong*, *good* or *bad*, have on us.⁸ These interventions involve techniques that manipulate the nervous system by means of drugs, electromagnetic stimulation of the brain, or biotechnological brain implants. I begin by proposing a list of requisites that, I believe, could be accepted by the main ethical theories, such as utilitarianism, deontology and virtue ethics, and finally I evaluate the aforementioned treatments according to their capacity to fulfill these requirements.

Of course, because many of these treatments are currently in an early stage of research, the possible secondary effects are unknown, and therefore those that would supposedly enhance subjects could result in a clear detriment and would therefore be inadvisable to administer. Consequently, the analysis will in all cases be based on the condition of adequate certainty that the treatments analyzed pose no serious risk for human health.

ETHICAL REQUISITES FOR ENHANCEMENT

The core problem in determining whether neural treatments in moral motivation truly involve an enhancement is that there is no consensus on what is right and thus on when a person is morally well disposed. For some ethical theorists, an individual behaving in a certain way does what is right only by virtue of certain rational characteristics of his or her behaviour. For others, motivation is a mixture of reason and emotions.

This controversy could nevertheless be avoided if we could determine certain achievements that the treatment in question would provide, and that, being minimal, would be acceptable according to the main ethical theories as manifestations of an enhancement in moral motivation. These achievements could be measured by virtue of their capacity to fulfill certain requirements that, when satisfied, convert a neural treatment into

⁸ I consider this type of moral neuroenhancement to pose fewer difficulties than others do. Thus, for example, an interface that, connected to the brain, prevents the individual from engaging in certain types of actions would only succeed in making someone behave according to moral judgments at the price of worsening their (deliberative) condition as a moral agent. And the attempt to enhance only the cognitive abilities necessary for proper moral deliberation would have limited efficiency, given the prominence that recent studies confer to the motives in moral decisions. (Haidt 2012; Crockett 2012).

something acceptable for everyone. The requisites that I propose would be the following three:

- *Requisite of minimal impartial perspective (RI)*: “Does the treatment enable the agent or those of his/her group, when free from the profound harm that others suffer, to be capable to make decisions after imagining him/herself actually in the position of those who were harmed?”⁹

A treatment that fosters this faculty in a subject could be accepted within the different ethical theories. It would be clearly acceptable for utilitarianism, even with the recognition that the treatment would be insufficient. This theory demands that with our actions we seek the maximum well-being possible of all those involved, giving all equal consideration. This would also be valid, and even more justified, when others are seriously harmed but the subject and those of his/her group are not. This type of treatment would also be positively valued by virtue ethics. On succeeding in implanting in the character of the subject a concern for situations with a markedly unequal distribution of suffering, we would be, this theory might state, clearing the path for any agent wishing to attain moral virtue. To a certain extent, we would be faced with something similar to what is purported by deontological ethics, since a treatment that fulfills this requirement would promote the correct behaviour of the agent undergoing the treatment, predisposing him/her to impartiality, whether by universalizing our rules of conduct or by treating all humans with respect. A Kantian deontologist would of course need the agent to understand that he/she should want to behave impartially because this is what is rational; but, even so, this deontologist would invariably have to recognize that this type of intervention would at least facilitate authentically moral behaviour.

⁹ We might ask whether this greater capacity to put oneself in the place of others really implies a moral improvement. As Julian Savulescu observed after reading this paper, could a sadist not continue to be so after having imagined the victim’s pain? Would something more than the imagination not be needed to improve morally? Something of empathy, for example? Surely, but I think that often this empathy fails to arise because of the lack of a capacity of putting oneself vividly in the place of the sufferer. Sometimes one can gain an idea of the way others suffer but this is far different from imagining how *oneself* would suffer in the position of the sufferer. This interiorized imagination is not a guarantee that one would have pity for a sufferer but it would be, even in the worst of cases, a great step forward.

- *Requisite of reduction of extremely negative consequences (RC)*: “Does the treatment of the agent’s motivation result in fewer situations of profound harm for many humans?”¹⁰

Clearly, an evaluation of the treatment according to the resulting overall degree of well-being can hardly fail to be accepted by utilitarianists, since for them the consequential criterion is the only one that finally determines moral rightness. In turn, virtue ethics should not reject it either, since the good life, which is the aim of all virtuous beings, does not depend exclusively on their behavioral and deliberative excellence but also on the fact that external factors should neither frustrate their expectations nor seriously harm them. And what about the deontologists? For them, the consideration of the consequences is morally irrelevant only when ineluctable obligations are at stake. But even in these situations most deontologists accept a kind of catastrophic clause that would explain that, when much is at stake, what is correct is defined not by virtue of these obligations or the nature of the action, but by the consequences of what a person does or does not do.¹¹

- *Requisite of self-determination (RA)*: “Does the treatment respect the autonomy and the identity of the agent?”

¹⁰ Both this requisite as well as the previous one have been defined purposefully using an imprecise term such as “profound” to facilitate the objective basis of this work in order to reach an initial agreement among the different ethical theories on what could be considered an enhancement of moral motivation. Subsequent research in the field of effects of the particular treatments should enable the specification of the threshold of fulfillment of each of the requisites.

¹¹ Ch. Fried, a recognized deontologist, holds that “we can imagine extreme cases where killing an innocent person may save a whole nation. In such cases it seems fanatical to maintain the absoluteness of the judgment, to do right even if the heavens will in fact fall. And so the catastrophic may cause the absoluteness of right and wrong to yield, but even then it would *non sequitur* to argue (as consequentialists are fond of doing) that this proves that judgments of right and wrong are always a matter of degree, depending on the relative goods to be attained and harms to be avoided. I believe, on the contrary, that the concept of the catastrophic is a distinct concept just because it identifies the extreme situations in which the usual categories of judgment (including the category of right and wrong) no longer apply» (1978, 10). Something similar is expressed by Donagan (1977, 206-7).

All moral theories, specially virtue ethics and deontological ethics, start from the stance that actions take on moral relevance only if they come from an agent that could have done otherwise. Also, they coincide in that this freedom of action has to be limited by behaviour according to the reasoning or principles agents identify with. It is true that utilitarianism has often been criticized for not safeguarding the personal autonomy and integrity of agents who, according to this theory, should always subordinate their personal preferences and projects to the general well-being. Nevertheless, there are less demanding utilitarian versions that, inspired in one way or another by the ideas of J.S. Mill, reconcile the principle of utility and the reasons centered on the agent.¹²

EVALUATION OF INTERVENTIONS

Having explained my proposal of ethical requisites for enhancement, I apply below a tentative test to the currently most promising neural interventions in the sphere of moral motivation, in order to ascertain whether or not they fulfill these requisites. I frame these in the four following scenarios:

SCENARIO 1: STRENGTHENING OF WILL

Ann is a professor of moral philosophy who, after much study, has reached certain conclusions on what is right, which she wishes to apply to her personal life. She has no doubts in this regard. Her fundamental values are upheld now in these beliefs and she is firmly decided that they also govern her conduct. However, usually, before doing what she believes is correct, always based on the impersonal consideration of all those affected by the action, another notion dominated by self-interest suddenly appears and moves her to carry out a different act. In fact, she has the capacity for self-control, but she does not always exercise it. This weakness of will power makes her feel badly and she strongly wishes she could be more resolute in her moral decisions.

¹² For an introduction to proposals of this type, such as rule, motive, or two-level utilitarianism, see Shaw (1999). I appreciate the suggestion of Julian Savulescu to discuss this nuance.

Anne tells her problem to a colleague who specializes in neuroethics. He helps her see that the cause of her problem resides in her difficulty to delay the gratification of her behaviour. In her moments of crisis, he explains, she places the satisfaction of her immediate desires before the more gratifying state that she would enjoy over the long term for acting correctly. He also says that this impulsiveness is associated with greater activity in the parietal and frontal cortex. Despite its partial environmental conditioning, parietal and frontal cortex activity is strongly determined by certain genotypes.¹³

Thus, on the suggestion of her colleague, Ann decides to take a psychotropic drug to control this impulsiveness that prevents her from behaving as rightly as she wishes. This substance has balanced effects on neurotransmitters, such as dopamine, adrenaline, noradrenaline, and serotonin. Thanks to this treatment, when Ann takes the drug, she displays strong will power to act according to the intentions and reasoning that are consistent with her values.¹⁴

A problem with this treatment, given that the drug is administered regularly, could be its negative side-effects on the nervous system. Ann could quit the treatment and try another type of intervention that, despite being more costly, is not invasive, lacks the undesired effects of regular drug use, and is far more effective. This consists in stimulating the prefrontal ventromedial cortex by the activation of an electric current through electrodes placed on the scalp. This zone of the brain has an essential function in regulating reasoning and decision making, and therefore this technique, compared to drugs, has an even more pronounced positive effect on decision making by fortifying her resolution to act according to her principles and to disregard immediate impulses.

It seems evident that both treatments fulfill the requisites proposed above. On the one hand, on helping people to match their conduct to their moral principles, which for their own nature they should do from impartial consideration of the situations, the treatments would fulfill the RI.¹⁵ However, fulfillment of the RC depends on whether this impartial

¹³ Boettiger *et al.* (2007).

¹⁴ See Stahl (2006).

¹⁵ From this, it can be deduced that these techniques can enhance moral motivation only in the case in which, like Ann, the subject has knowledge of what is right and the will, though weak, to follow it. If not, then more than an enhancement, it becomes a serious setback. With these treatments, people could also be expected to become more

consideration of the situations is accompanied by an interest in determining what is right, each of person paying attention to the consequences of his/her actions towards others and not only to the nature of these deeds. Even so, we can accept that it is very probable that this will occur – that the subject will do whatever is possible to reduce serious injury to others, given that what is achieved with these treatments on fulfilling the RI is that the subject is predisposed to put him/herself in the place of others when they are severely harmed, and that this predisposition is usually expressed in acts of solidarity. Furthermore, these treatments also fulfill the RA, since both before as well as after taking the drug or receiving the electric stimulation, Ann would recognize the reasons for and against these interventions and she would autonomously choose to receive them and continue with them. Therefore, at no time would Anne stop being the force behind her own decisions and actions.

SCENARIO 2: REDUCTION OF AGGRESSIVENESS

Paul realizes that he is becoming progressively more aggressive and violent. Problems at work and at home impel him, at critical moments of intense stress, to start a fight compulsively with anyone around him. This behavior has caused him court judgments and has aggravated his problems at work and in his family. As a result of it, he decides to act.

He follows psychological treatment that partially palliates his stress situations, but this proves insufficient. An anatomical study reveals certain peculiarities of his brain that would account for his tendency to be aggressive. The orbitofrontal cortex of his brain, the area where emotions are regulated, due to perturbations in the system of the serotonergic transmitter, receives an inadequate dose of serotonin and therefore he fails to control his violent impulses or regulate emotional reactions to provocation. Therefore, Paul decides to take selective serotonin reuptake inhibitors (SSRIs), a substance widely used to combat depression, anxiety, and compulsive obsessive disorder by slowing down the absorption of serotonin, making it more available to stimulate the receptors and thus helping to reduce aggressiveness.¹⁶

immoral or criminal, facilitating their control, for example, of the impulse to help people that suffer.

¹⁶ Almeida *et al.* (2005); Miczek *et al.* (2007); Coccaro (2012); Krakowski (2003).

Like Ann, after some time under treatment Paul begins to worry about the possible side effects of prolonged drug use. He decides to try brain stimulation, specifically of the posteromedial hypothalamus, which pacifies his aggressive impulses.¹⁷ He tries this while development proceeds on the experimental technique of a neural implant with sensors that measure the brain or blood levels of the substances responsible for aggressive impulses and that block imminent violent behaviour, whether by releasing substances that counteract the others, or by electric stimulation.¹⁸

Without a doubt, the fact that Paul receives some of these treatments contributes to the fulfillment of RI. By avoiding his aggressiveness, Paul frees himself from strong impediments so that, at least in extreme situations, he can consider the suffering of others to be as relevant as his own. On the other hand, given his prior belief that it is not good for him to be so aggressive, the intervention can only be seen as a means of increasing in his freedom and bringing him closer to his true identity, thereby fulfilling RA.¹⁹

Also, at first sight, it appears that there would be no problems, either, in satisfying the RC. With less aggressiveness, the world would be better off, as we free ourselves from the suffering provoked by aggressive individuals. It can be argued that the loss of aggressiveness might also entail negative consequences. In certain contexts, the disposition of being aggressive could promote moral conduct, for example by preventing someone from abusing other people.²⁰ However, the objection does not make much sense if the treatments to which we are referring in this second scenario are reserved for exceptional cases such as Paul's. In other cases, normal ones, perhaps the proper solution would be not to undergo any neural treatment and permit, as is usual, a certain degree

¹⁷ Experiments performed with psychiatric patients with a strong tendency towards aggressiveness have demonstrated the effectiveness of this technique. See Franzini *et al.* (2005).

¹⁸ Van Gompel *et al.* (2010)

¹⁹ It could be conjectured that RA might not be fulfilled if the enhanced person, as opposed to our example, had always been violent and aggressive. However, as we shall see further on, personal identity need not always be taken as something previously established, but also as something that could consist of a way of being that is created or striven towards over a lifetime, thereby enabling oneself at times to be truly authentic if the change is radical. I appreciate the suggestion of Julian Savulescu to discuss this nuance.

²⁰ This is what is adduced in Chan and Harris (2011).

of violent aggressiveness in extreme situations. Even if a generalization of these types of treatments were defended, the objection would continue to lack sense. Given the rarity of these extreme situations, overall we gain more when people are, in general, less aggressive. The strength of this reply resides in the evidence that if being prepared for these sorts of extreme situations were positive, we should consider trying to make people more aggressive than they are by nature, and this is not the case. This seems unadvisable. We believe that the proper way to prevent and resolve situations that require aggressiveness is to rely on institutions created for this purpose and that should have the exclusive right to use force and violence. If we had to justify it, we might say that violence could be a lesser evil only to avoid more violence, and therefore should be used only in an instance that, being institutional, has better means to apply it appropriately, is more impartial and reduces the resentment and the urge for vengeance that would come from an act of violence committed by an individual.

SCENARIO 3: INCREASE IN EMPATHY

Peter feels privileged to have been brought up in an atmosphere of keen social awareness. His family as well as his friends at school are exceptional for their high degree of commitment to altruistic causes. However, Peter feels anxiety because when faced with dedicating part of his resources or time to charitable organizations with which those around him collaborate, he invariably finds some excuse not to participate. He truly feels badly because he thinks that he should act like those close to him, with whom he fully identifies.

His anxiety takes him to an expert, who informs him that his lack of authentic feeling of solidarity results from certain biological determinants. Peter is told that just his being male, as opposed to most of those close to him, who are females, makes him more predisposed to lack empathy for others.²¹ Moreover, explains the expert, all the pro-social surroundings in which he was raised may have been insufficient to motivate him towards

²¹ Baron-Cohen (2003). More recently, Hastings *et al.* (2014, 420) have corroborated this thesis, demonstrating that certain levels of testosterone, a male hormone, could impede affectionate responses towards others.

altruism because, probably due to his genetic makeup,²² those areas of the brain related to empathy (amygdala, insula, orbitofrontal cortex, dorso-lateral prefrontal cortex, and the mirror neuron system) function differently in him.²³ Therefore, the expert tells Peter, his disposition to help others is due to the intervention of certain substances, in particular a hormone called oxytocin. This, together with the arginine vasopressin, forms an essential part of a system that predisposes monogamous mammals to form a pair and care for offspring (among other things).²⁴ This hormone can be measured in the blood, in urine, and in the cerebrospinal fluid. It has been synthesized and, when administered in humans, nasally or intravenously, key capacities for social conduct are heightened, such as recognizing faces,²⁵ inferring the emotions and intentions of others from facial expressions,²⁶ subjectively experiencing empathy or generosity,²⁷ feeling greater trust in others²⁸ or becoming more disposed to sacrificing for others.²⁹

Would this type of treatment fulfill the requisites proposed here? Not always. Thus, although at first sight its generalization would appear to satisfy the RC, given that it could help make the world a better place, where people would be more concerned for others and more willing to work together, thereby helping those that suffer, there is also the risk

²² Studies in behavioral genetics with twins have demonstrated the existence of a genetic predisposition to have concern for others, and subsequent research has located the cause of this predisposition in genes DRD4 and COMT. See Hastings *et al.* (2014).

²³ See Ramachandran, V.S. and Oberman (2006); Hastings *et al.* (2014, 416); Crockett *et al.* (2010).

²⁴ On how empathy arises in the evolution of basic emotions related to the care of offspring, see Hasting *et al.* (2014).

²⁵ Ferguson *et al.* (2000); Savaskan *et al.* (2008).

²⁶ Domes *et al.* (2007).

²⁷ Insel, T.R., and Fernald (2004); Zak *et al.* (2007), Barraza, J., and Zak (2009); Hurlmann *et al.* (2010); Rilling *et al.* (2011).

²⁸ The relation between oxytocin and trust has been demonstrated with experiments in which a member of a couple is asked to choose a quantity of money to give to the other, knowing that the money received will be tripled. Then, the second member will decide the quantity of money to give back to the first one. In the experiments, the subjects that had been administered with oxytocin demonstrated great trust in the other, giving a great quantity of money even without knowing how much would be returned. See Kosfeld *et al.* (2005); Zak *et al.* (2004).

²⁹ Morhenn *et al.* (2008).

that individuals enhanced with oxytocin would be more trusting in contexts in which they should be wary.³⁰ Thus, for example, thanks to a greater benevolence, they could make it easy for free riders (those who seek only their own well-being by taking advantage of others) to extend their selfish behaviour without restraints and consequently provoke more misery and insecurity; in this regard, RC would not be fulfilled.

Furthermore, this treatment does not satisfy RI in all cases. It is true that oxytocin plays a central role in a subject's understanding of morality. Individuals with very low levels of empathy have exhibited difficulties in distinguishing moral rules from merely conventional ones such as etiquette, and to arrange them correctly according to their relevance³¹. This appears to demonstrate that we perceive the relevance of moral norms only when we are capable of really putting ourselves in the place of others and thereby understanding how important, also for others, basic interests are, such as not suffering or not dying. And this capacity to empathize is bolstered by oxytocin, thus facilitating, in principle, the fulfillment of RI, a demand that the person receiving the intervention should be more willing to put him/herself in the place of others in extreme situations. Nevertheless, this type of intervention might not always satisfy RI because, as demonstrated in some studies, the increase in oxytocin also accentuates favoritism. It has been demonstrated that this substance fosters a protectionist attitude only for members of our own group at the same time as a strongly defensive attitude against strangers,³² which is expressed at times as patent racial favouritism,³³ as an excessive conformity to the group itself,³⁴ and even a dishonest component, without hesitation, to benefit the group.³⁵

And what about RA? Given the relevant similarities between this treatment and the one of the next scenario, we will leave the consideration of whether or not both fulfill this requisite for later.

³⁰ Baumgartner *et al.* (2008).

³¹ Blair *et al.* (2005).

³² Declerck *et al.* (2010); Sheng *et al.* (2013).

³³ De Dreu *et al.* (2010; 2011), Sheng *et al.* (2013)

³⁴ Stallen *et al.* (2012).

³⁵ Shalvi and De Dreu (2014). Because oxytocin could foster discriminatory attitudes such as racism or xenophobia, Persson, I., and Savulescu (2012b, 118-120) believe that the use of this substance should always be accompanied by an enhancement in the capacity of achieving anti-racist reasoning. Douglas (2008), however, is supported by Terbeck *et al.* (2012) to defend a modification of the racist attitudes by using betablocking propranolol.

SCENARIO 4: INCREASED SENSE OF JUSTICE

Andrea looks after her own interests. It is not that she does not trust others. She simply believes that each person should pursue his/her own interests and, to protect her own, she sees nothing wrong in taking advantage of the naïveté of others and in avoiding responsibilities whenever she can. With expressions such as “that way they will learn they cannot go through life like that”, she justifies to herself each act of taking advantage of others, accepting their help without any intention of reciprocating. Since she is not willing to cooperate, she often has no choice but to deceive others.

Given that people around her increasingly know about her real intentions of not reciprocating, she feels socially punished and isolated for her attitude. She has also had problems with the law, as she has repeatedly failed to fulfill her part in contracts with others. She realizes she needs to change and asks for help from a moral educator. Andrea believes that her problems come from an upbringing without values. After an exhaustive examination, the educator explains that her problem is not her upbringing but that rather that her behaviour has a biological explanation.³⁶

She consults a neuroenhancement specialist, who recommends her to take selective serotonin reuptake inhibitors (SSRIs). She is informed about experiments in which subjects who had taken this drug to boost their serotonin levels showed greater willingness to share money fairly,³⁷ to reject unjust offers of money,³⁸ and to cooperate with others.³⁹ SSRIs

³⁶ For this conclusion, the research by Wallace *et al.* (2007) is highly relevant, demonstrating that twins share the same idea of justice, something that does not necessarily occur among siblings that do not share the same genetic makeup.

³⁷ The results of these experiments, based on this type of context of decision known as the Dictator Game, where the subject has to divide money as he/she likes with another participant, are found in Tse, W.S., and Bond (2002).

³⁸ For this, the team of M.J. Crockett placed experimental subjects in an Ultimatum Game. The proposer offered the division of the money between the two. The responder could accept this division or reject it, in which case nobody received anything. It was common for subjects to reject offers that they considered very unjust, despite that they themselves would end up with nothing. Nevertheless, the individuals with a higher level of serotonin reached the highest levels of rejection of unjust offers (Crockett *et al.*, 2008). For Crockett *et al.* (2010), however, this increase in the rejection of unjust offers should be interpreted more as an aversion to help others than as a stronger sense of justice.

³⁹ In this case, the group of R.M. Wood designed the experiments according to the

could therefore be the solution for her. As these substances would reinforce her sense of fairness, Andrea would develop a predisposition to collaborate.

From the above, it might be deduced that the sense of fairness that these drugs enhance is based on reciprocity – on the propensity to return a favor with gratitude, and an offence with proportionate anger. Also they would induce Andrea to prefer, under conditions of equality, a fair deal over an unfair one. Plus, it is assumed that if individuals strengthen their commitment to reciprocity and equality, they will be more willing to put themselves in the place of others to make moral decisions, especially in situations in which some are badly disadvantaged with respect to others. It might be assumed, therefore, that Andrea's use of these drugs satisfies the RI.

It fulfills the RC as well, since the keener sense of fairness means, as we have seen, a greater disposition to collaborate, and this alone would make it easier for people to commit to agreements of mutual help if one of those involved undergoes a hardship. Furthermore, this desire to cooperate triggers deeper concern for others who are worse off, since, as is well known, the greater the willingness to collaborate and to abandon a selfish attitude, the freer the person is from overriding worries about satisfying basic needs and avoiding serious losses.

However, doubts persist as to whether this treatment, like the one that increases empathy, satisfies the RA. It can be objected that the enhanced individual becomes less autonomous and forfeits identity. With respect to autonomy, the objection would hold that Andrea and Peter, resorting to biological interventions to change their motivation, would be basing their decisions and their behaviour on an external element, rather than what is most intrinsic to them, which is their inner capacity for deliberation. Thus, ultimately, they would quit being true moral agents and transform themselves into some sort of puppets moved by the enhancers and treatments that they would have been submitted to.⁴⁰ To this

game of the Prisoner's Dilemma. This game represents the usual manner of thinking of two persons who are incommunicado and that could get some benefit from harming the other. In the version used by Wood, the game is the same but played more than once while permitting the participants to remember the actions of their opponents during previous encounters – that is, they remember a past grievance from the opponent (Wood *et al.*, 2006).

⁴⁰ Harris (2011) poses this objection in other terms. The enhanced agent would lose autonomy, he holds, because the agent could not quit doing what is correct. Therefore, this type of moral enhancement would be worse than the cognitive type

objection, it can be responded that, in terms of autonomy, there would be no significant differences between individuals enhanced by biological methods and by traditional ones. People who, by education or any other form of socialization, have acquired a proper moral conduct have also changed thanks to others. In fact, moral education is a process of indoctrination that is not always characterized by a discursive method (using persuasion, oversimplification, deceit, or even punishment), even when the objective is finally for the person educated to be capable of thinking independently. Education and biological intervention can be seen as two forms of externally facilitating the subjects' autonomy, helping the person in different ways to overcome the motivational (as well as cognitive) limitations that prevent him/her from matching his/her conduct to values or preferences to be reached through independent and thoughtful deliberation. In this sense, it could be added that the subject would achieve still more autonomy with biological intervention because, as opposed to the methods of moral education, if it is undertaken in the adult age, it derives from the subject's prior freely made decision and not from the unilateral initiative of the educator.⁴¹

Similar arguments can be used to address a second claim that the RA is not fulfilled, in this case adducing that these treatments are an attack against the subject's identity. This objection to bio-enhancement makes sense only from an essentialist conception of the self. Such an understanding of the self, however, hardly fits reality. We do not believe that humans are born with a way of being that can or must be maintained for life. We believe that either humans proceed to develop themselves based on experiences throughout life, or else that these experiences help each person to discover their authentic self, which differs from the one by which that

to avoid, for example, racist attitudes, since the latter type will always permit the agent to decide not to listen to his/her enhanced judgment on the irrationality of racial discrimination. But this result is strange. If the argument presented by Harris were true, we would have to accept also that a person who, whether by nature or by indoctrination, has a predisposition to do what is correct, he/she is therefore less free. This response to Harris has been suggested by Persson, I., and Savulescu (2012a, 409); DeGrazia (2014, 5-7). The latter holds, furthermore, that, assuming that freedom is lost with enhancement, this should be offset by the possible overall good consequences of this situation.

⁴¹ Dees (2007; 2011). Bublitz *et al.* (2009) go even further and hold that the agent would really not lose autonomy even without having given consent prior to the rehabilitation, in that the agent ends up identifying with the newly acquired trait.

person usually characterized. With both readings, i.e. self-realization or self-discovery, the interventions, especially if they are reversible –as is the case of those studied here– can nevertheless be a useful tool more than an impediment against reaching that sought-after identity. Peter and Andrea were unsatisfied with their way of acting and therefore turned to neuroenhancements to turn into the way they wanted to be or as they thought they should be. Moreover, in the case of heightened empathy, the usefulness of constructing or discovering a more authentic self becomes especially evident. Individuals with a deficit of empathy normally have a limited understanding of themselves because, being incapable of putting themselves in someone else's shoes, they cannot see themselves as others see them, nor can they formulate their own proposals to change their personality from the empathetic knowledge of others.⁴²

CONCLUSIONS AND A NOTE OF CAUTION

The debate on whether it is ethical to enhance ourselves morally by modifying our biology can, I believe, advance only if analyses are made separately of each type of intervention and starting in all cases from the empirical data provided by recent scientific discoveries.

In addition, these analyses should consist in corroborating not whether the intervention in question converts us into perfect moral agents overnight, but to what extent with such treatment we manage to exert a positive influence on some morally relevant capacities. In this work, I have looked specifically at the neural interventions that could enhance moral motivation. Explaining what counts as an enhancement for moral motivation, I have proposed four requisites. Three ethical requisites would consider the intervention in a subject's motivation to be an enhancement if 1) they permit that subject to have a predisposition to reason morally from an impartial perspective in extremely disadvantageous situations for outsiders; 2) the intervention contributes to or does not diminish the

⁴² A different matter is how the subjects might feel after suddenly changing important features of their motivation and, therefore, their personality. There is probably a period of feelings of estrangement and insecurity. But psychologists already know the facility of human beings to fully appropriate outside actions and thoughts. It also occurs in the ambit of moral indoctrination. Therefore, we might infer that the same internalization would eventually occur in the case of the individual enhanced through neurological techniques.

subject's ability to avoid undesired situations for everyone, and 3) the subject should not lose self-determination. I have argued that these requisites would be admitted by the main ethical theories as allowing progress, although minimal, in moral motivation. The fourth, non-ethical requisite, consists in the demand that the intervention should not involve secondary effects that harm the subject's health.

From these methodological assumptions, I have examined only four neural treatments that, given our current scientific knowledge, can conceivably imply a positive change in moral motivation. The results of this examination indicate that, to a great extent, these treatments could be viewed as moral enhancements.

Nevertheless, I believe that such treatments should be delayed as long as it is not possible to modulate them by establishing certain thresholds, above which the effects would clearly no longer be valuable. The clearest example in this respect is the situation in which greater empathy after oxytocin administration can either reduce the impartial perspective of the treated individual or else deepen trust in others to the extent that the individual fails to protect him/herself against parasitic behaviour of others.

The demand of modulation goes much farther. The application of these types of intervention should be decided only with a considerable degree of certainty that their effects will be positive given the particularity of the individual to be enhanced. This constitutes a difficult challenge, since, despite the great development of neurological sciences, the complex participation of neural elements in different systems can not be known in detail,⁴³ and therefore the treatment can always carry unexpected results related to the biology of each individual and, even having the same biology, to the different environment of each individual.⁴⁴

This being said, it is worth noting that to a certain extent this also often occurs in the context of medical therapy, which in many cases is effective only if very precise and personalized treatments are provided. Just as this difficulty is not an impediment to continuing research in biomedicine and applying its achievements only when the obstacle of par-

⁴³ On functional complexity, for example, of serotonin, see Zarpentine (2013, 147-8).

⁴⁴ Thus, for example, it has been demonstrated that the effects of oxytocin over empathetic interest can be moderated by certain individual characteristics, such as social abilities (Hastings *et al.* 2014, 420.) or a high degree of stress (Zak 2011, 56-7).

ticularization is overcome, it should not be an obstacle to consider the possibilities of using the discoveries of the neurological sciences with caution and rigor in order to enhance the moral predisposition of humans.

ACKNOWLEDGMENTS

I thank the Department of Economy and Competitiveness of the Spanish Government (grant number FFI2012-23565) for their funding. I also thank Julian Savulescu and David Rodriguez-Arias for their helpful comments on a draft of this paper.

REFERENCES

- Agar, N. 2003. *Liberal Eugenics*, Blackwell.
- Almeida, R.M., Ferrari, P.F., Parmigiani, S., and K.A. Miczek. 2005. "Escalated Aggressive Behavior: Dopamine, Serotonin and GABA." *European Journal of Pharmacology* 526.
- Baron-Cohen, S. 2003. *The Essential Difference: Men, Women and the Extreme Male Brain*. Penguin/Basic Books.
- Barraza, J., and P.J. Zak. 2009. "Empathy Toward Strangers Triggers Oxytocin Release and Subsequent Generosity." *Annals of the New York Academy of Sciences* 1167.
- Baumgartner, T.M., Heinrichs, M., Vonlanthen, A., Fischbacher, U. and E. Fehr. 2008. "Oxytocin Shapes the Neural Circuitry of Trust and Trust Adaptation in Humans." *Neuron* 58.
- Blair, J.M.D., and K. Blair. 2005. *The Psychopath: Emotion and the Brain*. Blackwell.
- Boettiger, C.A., Mitchell, J.M., Tavares, V.C., Robertson, M., Joslyn, G., D'Esposito M., and H.L. Fields. 2007. "Immediate Reward Bias in Humans: Fronto-Parietal Networks and a Role for the Catechol-O-Methyltransferase 158^{Val/Val} Genotype." *The Journal of Neuroscience* 27(52): 14383-14391.
- Bostrom, N. 2003. "Human Genetic Enhancements: A Transhumanist Perspective." *Journal of Value Inquiry* 37 (4):493-506.
- Bostrom, N. 2004. "Transhumanism. The World's Most Dangerous Idea?" *Betterhumans* 10/19/2004.
- Bostrom, N. 2005a. "Future of Human Evolution." In *Death and*

- Anti-Death*, edited by Ch. Tandy. Ria University Press.
- Bostrom, N. 2005b. "In Defence of Posthuman Dignity." *Bioethics* 19 (3): 202-2014.
- Bublitz, J.C., and R. Merkel. 2009. "Autonomy and Authenticity of Enhanced Personality Traits." *Bioethics* 23 (6).
- Caplan, A.L. 2009. "Good, Better, or Best?" In *Human Enhancement*, edited by J. Savulescu and N. Bostrom. Oxford University Press.
- Chan, S., and J. Harris. 2011. "Moral Enhancement and Pro-Social Behaviour." *Journal of Medical Ethics* 37 (3).
- Coccaro, E.F. 2012. "What Is the Nature of Serotonergic Abnormalities in Human Aggression?" *Biological Psychiatry* 72.
- Crockett, M.J. 2012. "Moral Bioenhancement: A Neuroscientific Perspective." *Journal of Medical Ethics*. Published online first: 28/09/2014. Doi: 10.1136/medethics-2012-101096.
- Crockett, M.J., Clark, L., Hauser, M.D., and T.W. Robbins. 2010. "Serotonin Selectively Influences Moral Judgment and Behavior Through Effects on Harm Aversion." *Proceeding of the National Academy of Sciences* 107 (4).
- Crockett, M.J., Clark, L., Tabibnia, G., Lieberman, M.D., and T. W. Robbins. 2008. "Serotonin Modulates Behavioral Reactions to Unfairness." *Science* 320.
- De Dreu, C.K.W., Greer, L.L., Handgraaf, M.J., Shalvi, S., Van Kleef, G.A., Baas, M., Ten Velden, F.S., Van Dijk, E., and S.W. Feith. 2010. "Neuropeptide Oxitocin Regulates Parochial Altruism in Intergroup Conflicts among Humans." *Science* 328.
- De Dreu, C.K.W., Greer, L.L., Van Kleef, G.A., Shalvi, S., and M. J. Handgraaf. 2011, "Oxitocin Promotes Human Ethnocentrism." *Proceedings of the National Academy of Sciences* 108 (4).
- Declerck, C., Boone C, and T. Kiyonari. 2010. "Oxytocin and cooperation under conditions of uncertainty: the modulating role of incentives and social information." *Hormones and Behavior* 57: 368-374.
- Dees, R.H. 2007. "Better Brains, Better Selves? The Ethics of Neuroenhancements." *Kennedy Institute of Ethics Journal* 17.
- Dees, R.H. 2011. "Moral Philosophy and Moral Enhancements." *AJOB Neuroscience* 2 (4).
- DeGrazia, D. 2014. "Moral Enhancement, Freedom, and What We (Should) Value in Moral Behaviour." *Journal of Medical Ethics* 40 (6) (Published online first: 25/01/2013. Doi:10.1136/medethics-2012-101157)

- Domes, G., Heinrichs, M., Michel, A., Berger, C., and S.C. Herpertz. 2007. "Oxotocin Improves 'Mind-Regarding' in Humans." *Biological Psychiatry* 61.
- Donagan, A. 1977. *The Theory of Morality*. Chicago University Press.
- Douglas, T. 2008. "Moral Enhancement." *Journal of Applied Philosophy* 25.
- Elliot, C. 2003. *Better Than Well: American Medicine Meets the American Dream*. W.W. Norton.
- Ferguson, J.N., Young, L.J., F. Hearn, E.F., Matzuk, M.M., Insel, T.R. and J.T. Winslow. 2000. "Social Amnesia in Mice Lacking the Oxytocin Gene." *Nature Genetics*, 25.
- Franzini, A., Ferroli, C.M.P., Bugiani, O. and G. Broggi. 2005. "Stimulation of the Posterior Hypothalamus for Medically Intractable Impulsive and Violence Behavior." *Stereotactic and Functional Neurosurgery* 83.
- Fried, Ch. 1978. *Right and Wrong*. Harvard University Press.
- Fröding, B.E.E. 2011. "Cognitive Enhancement, Virtue Ethics and the Good Life." *Neuroethics* 4 (3).
- Fukuyama, F. 2003. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. Profile.
- Habermas, J. 2003. *The Future of Human Nature*. Polity Press.
- Haidt, J. 2012. *The Righteous Mind*. Random House.
- Harris, J. 2011. "Moral Enhancement and Freedom." *Bioethics* 25 (2).
- Hastings, P.D., Miller, J.G., Kahle, S., and C. Zahn-Waxler. 2014. "The Neurobiological Bases of Empathic Concern for Others." In *Handbook of Moral Development*, edited by K. Killen, and J. Smetana. Psychology Press.
- Hurlemann, R., Putin, A., Onur, O.A., Cohen, M.X., Baumgartner, T., Metzler, S., Kendrick, K.M. 2010. "Oxytocin Enhances Amygdala-dependent, Socially Reinforced Learning and Emotional Empathy in Humans." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30.
- Ida, R. 2009. "Should We Improve Human Nature? An Interrogation from an Asian Perspective." In *Human Enhancement*, edited by J. Savulescu, J and N. Bostrom. Oxford University Press.
- Insel, T.R., and R. D. Fernald. 2004. "How the Brain Processes Social Information: Searching for the Social Brain." *Annual Review of Neuroscience* 27.

- Jotterand, F. 2011. "Virtue Engineering' and Moral Agency: Will Post-Humans Still Need the Virtues?" *AJOB Neuroscience* 2 (4).
- Kamm F. 2005. "Is There a Problem with Enhancement?" *American Journal of Bioethics*, 5 (3): 5-14.
- Kass, L.R. 2002. *Life, Liberty and the Defense of Dignity: The Challenge for Bioethics*. Encounter Books.
- Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., and E. Fehr. 2005. "Oxytocin Increases Trust in Humans." *Nature* 435.
- Krakowski, M. 2003. "Violence and Serotonin: Influence of Impulse Control, Affect Regulation, and Social Functioning." *Journal of Neuropsychiatry and Clinical Neurosciences* 15.
- Miczek, K.A., Almeida, R.M.M., Kravitz, E.A., Rissman, E.F., De Boer, S.F., and A. Raine. 2007. "Neurobiology of Escalated Aggression and Violence." *The Journal of Neuroscience* 27.
- Morhenn, V.B., Park, J.W., Piper, E., and P.J. Zak. 2008. "Monetary Sacrifice Among Strangers Is Mediated by Endogenous Oxytocin Release after Physical Contact." *Evolution and Human Behavior* 29.
- Pacholczyk, A. 2011. "Moral Enhancement: What Is It and Do We Want It?" *Law, Innovation and Technology* 3(2).
- Persson, I., and J. Savulescu. 2012a. "Moral Enhancement, Freedom, and the God Machine." *The Monist* 95 (3).
- Persson, I., and J. Savulescu. 2012b. *Unfit for the Future*. Oxford University Press.
- Ramachandran, V.S., and L.M. Oberman. 2006. "Broken Mirrors: A Theory of Autism." *Scientific American* 295.
- Rilling, J.K., DeMarco, A.C., Hackett, P.D., Ditzen, B.,...Pagnoni, G. 2011. "Effects of Intranasal Oxytocin and Vasopressin on Cooperative Behavior and Associated Brain Activity in Men". *Psychoneuroendocrinology* 37.
- Sandel, M. 2004. "The Case Against Perfection." *Atlantic Monthly* (Apr. 2004): 51-62.
- Savaskan, E., Ehrhardt, R., Schulz, A. Walter, M., and H. Schächinger. 2008. "Post-learning Intranasal Oxytocin Modulates Human Memory for Facial Identity." *Psychoneuroendocrinology* 33.
- Savulescu, J. 2007. "Genetic Interventions and the Ethics of Enhancement of Human Beings." In *The Oxford Handbook of Bioethics*, edited by B. Steinbock. Oxford University Press.
- Schaefer, G.O. 2011. "What Is the Goal of Moral Engineering?" *AJOB Neuroscience* 2 (4).

- Shalvi, S. and C. De Dreu. 2014. "Oxytocin promotes group serving dishonesty." *Proceedings of the National Academy of Sciences USA* 111: 5503-5507.
- Shaw, W. 1999. *Contemporary Ethics. Taking Account of Utilitarianism*. Blackwell.
- Sheng, F., Zhou, B., Zhou, W., and S. Han. 2013. "Oxytocin modulates the racial bias in neural responses to others' suffering." *Biological Psychology* 92: 380-386.
- Stahl, S. 2006. *Essential Psychopharmacology: Neuroscientific Basis and Practical Applications*. Cambridge University Press.
- Stallen, M., De Dreu, C., Shalvi, S., Smidts, A., and A. Sanfey. 2012. "The herding hormone: oxytocin stimulates in-group conformity." *Psychological Science* 23: 1288-1292.
- Terbeck, S., Kahane, G., McTavish, S., Savulescu, J. Cowen, P.H., and M. Hewstone. 2012. "Propranolol Reduces Implicit Negative Racial Bias." *Psychopharmacology* 222 (3).
- Tse, W.S., and A. J. Bond. 2002. "Serotonergic Intervention Affects Both Social Dominance and Affiliative Behaviour." *Psychopharmacology* 161.
- Van Gompel, J.J., Chang, S.Y, Goerss, S.J., Kim, I.Y., Kimble, C., Lee, K.H. 2010. "Development of Intraoperative Electrochemical Detection: Wireless Instantaneous Neurochemical Concentration Sensor for Deep Brain Stimulation Feedback." *Neurosurg Focus* 29 (2).
- Walker, M. 2002. "Prolegomena to Any Future Philosophy." *Journal of Evolution and Technology* 10.
- Wallace, B., Cesarini, D., Lichtenstein, P., and M. Johannesson. 2007. "Heritability of Ultimatum Game Responder Behaviour." *Proceedings of the National Academy of Sciences* 104 (40).
- Wood, R.M., Rilling, J.K., Sanfey, A.G., Bhagwagar, Z., and R.D. Rogers. 2006. "Effects of Tryptophan Depletion on the Performance of an Iterated Prisoner's Dilemma Game in Healthy Adults." *Neuropsychopharmacology* 31.
- Zak, P. 2011. "The Physiology of Moral Sentiments." *Journal of Economic Behavior and Organization* 77.
- Zak, P.J., Kurzban, R., and W.T. Matzner. 2004. "The Neurobiology of Trust." *Annals of the New York Academy of Sciences* 1032.

- Zak, P.J., Stanton, A.A., and S. Ahmadi. 2007. "Oxytocin Increases Generosity in Humans." *Public Library of Science ONE* 2 (11).
- Zarpentine, C. 2013. "The Thorny and Arduous Path of Moral Progress: Moral Psychology and Moral Enhancement." *Neuroethics* 6.

Francisco Lara
Universidad de Granada
flara@ugr.es

This paper was received on November 23rd 2015, and was approved on May 12th May 2016.