

Moral Enhancement as a Collective Action Problem

WALTER GLANNON

Abstract

In light of the magnitude of interpersonal harm and the risk of greater harm in the future, Ingmar Persson and Julian Savulescu have argued for pharmacological enhancement of moral behaviour. I discuss moral bioenhancement as a set of collective action problems. Psychotropic drugs or other forms of neuromodulation designed to enhance moral sensitivity would have to produce the same or similar effects in the brains of a majority of people. Also, a significant number of healthy subjects would have to participate in clinical trials testing the safety and efficacy of these drugs, which may expose them to unreasonable risk. Even if the drugs were safe and effective, a majority of people would have to co-operate in a moral enhancement programme for such a project to succeed. This goal would be thwarted if enough people opted out and decided not to enhance. To avoid this scenario, Persson and Savulescu argue that moral enhancement should be compulsory rather than voluntary. But the collective interest in harm reduction through compulsory enhancement would come at the cost of a loss of individual freedom. In general, there are many theoretical and practical reasons for scepticism about the concept and goal of moral enhancement.

1. Introduction

Neuroscientist Donald Pfaff claims that altruistic behaviour is hard-wired into our brains.¹ Neurobiological mechanisms draw us out of self-interest and into the sphere of others by blurring the lines of identity. Cognitive psychologist Steven Pinker attributes the decline of violence over the centuries to a particular conception of morality, which 'is a consequence of the interchangeability of perspectives and the opportunity the world provides for positive-sum games'.² The ability to consider different perspectives is a function of our capacity to reason. Yet something clearly has gone awry in our altruistic brains and rationality when considering the magnitude

¹ Donald Pfaff, *The Altruistic Brain: How We Are Naturally Good* (New York: Oxford University Press, 2015).

² Steven Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (New York: Penguin, 2011), 182.

Walter Glannon

of harm among humankind. Another comment from Pinker provides some insight into the apparent loss of human pro-sociality:

The indispensability of reason does not imply that individual people are always rational or are unswayed by passion and illusion. It only means that people are *capable* of reason, and that a community of people who choose this faculty and *exercise* it openly and fairly can collectively reason their way to sound conclusions in the long run.³

Pinker's second comment suggests that reason itself has motivating power and that 'passion' and 'illusion' prevent people from acting on it. The tendency of certain emotions and irrational beliefs to interfere with reason may be part of an explanation for why individuals and groups harm others at local and global levels.

Interpersonal harm is not so much the result of a failure to recognise moral reasons as a failure to be motivated to act on them. Since the question at issue is how to reduce harm by improving moral behaviour, one would expect moral philosophers to offer guidance and provide the impetus for the action necessary to achieve this end. With arguably a few exceptions, however, they have not been successful in this regard.⁴ Philosophical arguments are generally ineffective in moving people and governments to act morally in respecting the rights, needs, and interests of all people. Some of these arguments rely on objective moral truths that presumably exist independently of our beliefs and attitudes.⁵ Many moral philosophers focus primarily on showing why one moral theory is superior to others in purportedly establishing these truths and often use implausible thought-experiments to defend their positions.⁶ The overemphasis on theoretical aspects of morality leaves their arguments motivationally inert.

Educational and political institutions have also failed to instill the motivation for moral behaviour. Noting this failure, Ingmar Persson

³ Pinker, *The Better Angels of Our Nature*, 183. Second emphasis added.

⁴ Peter Singer, 'Famine, Affluence and Morality', *Philosophy & Public Affairs* 1:3 (1972), 229–243, and *The Most Good You Can Do: How Effective Altruism is Changing Ideas About Living Ethically* (New Haven: Yale University Press, 2015).

⁵ T. M. Scanlon, *Being Realistic About Reasons* (Oxford: Oxford University Press, 2014).

⁶ Derek Parfit, *On What Matters*, Volumes I and II (Oxford: Oxford University Press, 2011).

Moral Enhancement as a Collective Action Problem

and Julian Savulescu emphasise the need to enhance our moral sensitivity and agency through biological means:

Our knowledge of human biology, in particular of genetics and neurobiology, is now beginning to supply us with the means of directly affecting the biological or physiological bases of human motivation [...]. We shall suggest that there are in principle no philosophical or moral objections to the use of such biomedical means of moral enhancement – moral bioenhancement, as we shall call it – and that the current predicament of humankind is so serious that it is imperative that scientific research explore every possibility of developing effective means of moral bioenhancement, as a complement to traditional means.⁷

As the last part of this passage indicates, Persson and Savulescu do not believe that moral bioenhancement (henceforth MB) alone will move us to engage in the type of moral behaviour necessary to meet the challenges of the ‘predicament of humankind’. Still, they suggest that MB would be the most critical intervention in addressing and ideally resolving this predicament. While they focus mainly on the threat from weapons of mass destruction (WMDs) and the effects of climate change, the problem also includes the harm resulting from genocide, civil wars, economic inequality, and smaller-scale criminal behaviour.

The concept of MB is fraught with scientific, ethical, and political challenges. I discuss MB as a set of collective action problems. Psychoactive drugs or other forms of neuromodulation designed to enhance moral sensitivity would have to produce the same or similar effects in the brains of a majority of people. This is a questionable hypothesis, given structural and functional differences in people’s brains and differences in how their neural networks mediate behaviour. A significant number of healthy subjects would have to participate in clinical trials testing the safety and efficacy of the drugs, which may expose them to unreasonable risk. Concerns about the ethical justification of the research alone may prevent such a project from getting off the ground. Even if research demonstrated that the drugs were safe and effective in improving moral motivation, a majority of the population would have to co-operate in a moral enhancement programme for such a project to succeed. This goal could be thwarted if enough people opted out and refused to enhance. It could also be thwarted if a minority who cause much of the world’s harm not only refused to enhance but

⁷ Ingmar Persson and Julian Savulescu, *Unfit for the Future: The Need for Moral Enhancement* (Oxford: Oxford University Press, 2012), 2.

also took advantage of the co-operation of others. To avoid either of these outcomes, Persson and Savulescu argue that moral enhancement should be compulsory rather than voluntary. But the collective interest in harm reduction through compulsory enhancement would come at the cost of a loss of individual freedom.

After analysing and discussing these issues, I conclude on a sceptical note. Although there is an urgent need to promote moral behaviour and reduce harm among humans, there would be too many problems with implementing and enforcing a programme of moral enhancement on a grand scale. Traditional indirect methods of moral enhancement, such as education, parenting, and political institutions, have fallen far short of this goal and at best can only approximate it. Yet even if these methods are limited in what they can achieve, they may be more viable and defensible alternatives to voluntary or compulsory alteration of thought and behaviour by manipulating the brain through pharmacological or other means.

2. The Neurobiology of Moral Decision-Making

The goal of any programme of moral enhancement would be to provide the motivation to act and facilitate the execution of this motivation in action. Although John Martin Fischer and Mark Ravizza develop and defend their theory of reasons-responsiveness to account for moral responsibility, it can be applied to moral enhancement. Reasons-responsiveness consists in the capacity to *recognise* reasons for or against actions and the capacity to *react* to these reasons in morally appropriate behaviour.⁸ Reasons-responsiveness involves not only the cognitive capacity to recognise moral reasons but also the emotional and volitional capacity to translate these reasons into right actions. These actions ideally would balance deontological considerations of treating individuals as ends in themselves and not merely as means⁹ with consequentialist considerations of bringing about outcomes that reduce or prevent harm and increase human welfare.¹⁰ Unlike the Kantian view, however, the crucial issue is

⁸ John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998), 62–91.

⁹ Immanuel Kant, *Groundwork of the Metaphysics of Morals* (1785), trans. and ed. by M. Gregor (New York: Cambridge University Press, 1998).

¹⁰ John Stuart Mill, *Utilitarianism and Other Essays* (1863), ed. by A. Ryan (London: Penguin, 1987).

Moral Enhancement as a Collective Action Problem

not whether one acts from duty or inclination in treating others as ends in themselves, but that one treats them in this way. It is the action itself that matters. The moral worth of the action depends more on how it affects others than the mental states behind it.

There has been disagreement in the debate on moral enhancement as to whether its goal would more likely be achieved by focussing on cognitive or on emotional processing. In a seminal paper on this topic, Thomas Douglas identifies what he calls ‘counter-moral emotions’ as targets for neuromodulation and a more general programme of moral enhancement.¹¹ These emotions include ‘a strong aversion to certain racial groups’ and ‘the impulse towards violent aggression’.¹² Douglas argues that the most promising way to correct these tendencies would be ‘an enhancement that will expectably leave the enhanced person with morally better motives than she had previously’.¹³ Emphasising the emotional aspect of behaviour, Douglas says that ‘the distinctive feature of emotional moral enhancement is that, once the enhancement has been initiated, there is no further need for cognition: emotions are modified directly’.¹⁴ In contrast, John Harris argues that moral enhancement is an extension of cognitive enhancement and that ‘emotional moral enhancement is simply ethically otiose’.¹⁵ Harris claims that ‘it is to rationality and its evolutionary origins, rather than to the emotions that we should look’.¹⁶ He further claims that ‘it is human imagination that is one of the most potent of cognitive faculties that enables us to put ourselves into, if not the shoes of relevant others, at least into an understanding of the nature and consequences of our acts and decisions, and of the effects of those decisions on others and the world’.¹⁷

Douglas’ and Harris’ analyses of moral behaviour in terms of separate cognitive and emotional faculties are symptomatic of a discredited dualistic model of explaining this behaviour. This model offers at best an incomplete explanation of moral reasoning and action and thus fails to include all the capacities that need to be targeted for moral enhancement. Increased imagination and decreased

¹¹ Thomas Douglas, ‘Moral Enhancement’, *Journal of Applied Philosophy* 25:3 (2008), 228–245.

¹² Douglas, ‘Moral Enhancement’, 231.

¹³ Douglas, ‘Moral Enhancement’, 231.

¹⁴ Thomas Douglas, ‘Moral Enhancement via Direct Emotion Modulation: A Reply to Harris’, *Bioethics* 27:3 (2013), 160–168, at 162–163.

¹⁵ John Harris, *How to Be Good: The Possibility of Moral Enhancement* (Oxford: Oxford University Press, 2016), 115.

¹⁶ Harris, *How to Be Good*, 125.

¹⁷ Harris, *How to Be Good*, 131.

counter-moral emotions would strengthen the recognitional component of reasons-responsiveness. But they would not automatically strengthen the reactive component of responsiveness and move one to perform the right actions in different circumstances.

Cognition and emotion are not segregated but integrated processes whose interaction enables moral sensitivity and rational and moral decision-making.¹⁸ These normative capacities are mediated by a distributed network of neural circuits. The prefrontal cortex (PFC) is particularly important in regulating the capacity to reason. It also has a critical role in emotion regulation. This is not only because of its projections to and connectivity with the amygdala and other limbic structures such as the cingulate cortex, but also because reason and emotion are both partly processed within the PFC itself. The complex function of the PFC shows that it is oversimplified and inaccurate to divide the brain into distinct regions mediating separate cognitive and emotional processing at the mental level.

Psychopathy is instructive in this regard. This is a behavioural disorder characterised by impaired capacity for empathy, impaired responsiveness to fear-inducing stimuli, and failure to conform to social norms. Because of these impairments, many psychopaths cause a substantial amount of harm to others. They have difficulty representing outcomes of actions. This can preclude or weaken their moral judgement by precluding or weakening their ability to foresee how their actions can adversely affect others.¹⁹ Functional imaging studies of psychopaths' brains have shown dysregulation in PFC-amygdala pathways, and this finding is part of an explanation of their behaviour. It is the result of dysfunction of not only cognitive or emotional processes but of dysfunctional interaction between

¹⁸ Antoine Bechara, Hanna Damasio, and Antonio Damasio, 'Emotion, Decision-Making and the Orbitofrontal Cortex', *Cerebral Cortex* **10**:3 (2000), 295–307. See also Antonio Damasio, *Descartes' Error: Emotion, Reason and the Human Brain* (New York: Grosset/Putnam, 1994), Jean Decety, Kalina Michalska, and Katherine Kinzler, 'The Contribution of Emotion and Cognition to Moral Sensitivity: A Neurodevelopmental Study', *Cerebral Cortex* **22**:1 (2012), 209–220, and Luiz Pessoa, *The Cognitive-Emotional Brain: From Interactions to Integration* (Cambridge, MA: MIT Press, 2013), 107–134.

¹⁹ R. J. R. Blair, 'Neurobiological Basis of Psychopathy', *British Journal of Psychiatry* **182**:5–7 (2003), 5–7, and R. J. R. Blair, 'Psychopathy: Cognitive and Neural Dysfunction', *Dialogues in Clinical Neuroscience* **15**:2 (2013), 181–190. See also Andre Glenn and Adrian Raine, *Psychopathy: An Introduction to Biological Findings and Their Implications* (New York: New York University Press, 2014).

Moral Enhancement as a Collective Action Problem

them. Any improvement in a psychopath's behaviour would require enhancing his cognitive-emotional processing by enhancing how his neural circuits generate and sustain this processing. Still, it is not clear to what extent psychopaths' brain abnormalities influence their wrongful and harmful behaviour. One must be cautious in drawing inferences from their brain abnormalities to their impaired moral reasoning and immoral behaviour because the first is not necessarily the cause of the second. Claims about the power of structural MRI and functional PET and fMRI scans to validate these inferences warrant caution as well. Scans of people's brains are only indirect measures of brain activity. They are visualisations of statistical analyses based on averaging large numbers of images. Brain scans are more appropriately described as scientific constructs than direct "real-time" pictures of the brain.

Empathy is a critical component of moral sensitivity. Harris' comment about moral enhancement depending on the cognitive capacity for imagination in 'putting ourselves into the shoes of relevant others' reflects a narrow conception of empathy. While imagination may promote concern for others that extends spatially and temporally beyond local interpersonal bonds held together by emotion, empathy is not a purely cognitive disposition. Nor is empathy separable from reason but can influence it in guiding action. Jean Decety and Jason Cowell explain:

Empathy [...] is not always a direct avenue to moral behavior. Indeed, at times it can interfere with morality by introducing partiality, for instance by favoring in-group members. But empathy can provide the motivational fire and push toward seeing a victim's suffering end, irrespective of group membership and culturally determined dominance hierarchies, preventing rationalization of injustice and derogation.²⁰

Decety and Cowell further state that empathy consists of cognitive, affective, and motivational components. The cognitive component enables one 'to consciously put oneself into the mind of another and imagine what that person is thinking or feeling'.²¹ The affective component enables one to become emotionally aroused by others'

²⁰ Jean Decety and Jason Cowell, 'The Equivocal Relationship Between Morality and Empathy', in Jean Decety and Thalia Wheatley (eds), *The Moral Brain: A Multidisciplinary Perspective* (Cambridge, MA: MIT Press, 2015), 297–302, at 279.

²¹ Decety and Cowell, 'The Equivocal Relationship Between Morality and Empathy', 284–285.

condition. The motivational component involves the urge to care for another's welfare. These mental capacities and their neural underpinning are not domain-specific but domain-general and involve an interconnected network of circuits distributed throughout cortical and limbic brain regions. This network comprises the 'moral brain'.²² It is similar to what Andrea Glenn, Adrian Raine, and Robert Schug describe as the 'moral neural circuit', which they point out is dysfunctional in psychopaths and may at least partly explain the failure to align their behaviour with moral and legal norms.²³ This circuit would have to be modulated in order to generate or improve moral sensitivity in the psychopath or others who lack this disposition. MB involving neuromodulation would have to influence not just one but all or most of the components in the moral neural circuit and how they interact with each other in regulating behaviour. Still, what Decety and Cowell describe as the 'urge' to care for another's welfare by itself is not sufficient to act in a way that exemplifies this care. One must translate this urge into morally appropriate behaviour. Any form of moral bioenhancement would have to facilitate the exercise of this motivational capacity in action.

Three hypothetical drug interventions have been proposed as possible ways of modulating neural networks associated with moral reasoning. Theoretically, they would enhance the capacity to translate the urge to care for others into action and thereby enhance moral behaviour. But there are neurobiological, psychological, and social reasons to be sceptical of the idea that taking a psychotropic drug targeting these networks would have this effect.

One study conducted by Molly Crockett and co-investigators showed that the selective serotonin reuptake inhibitor (SSRI) citalopram increased harm aversion in healthy subjects.²⁴ They claimed that 'these findings have implications for the use of serotonergic agents in the treatment of antisocial and aggressive behaviour' in promoting pro-social behaviour.²⁵ Crockett is more circumspect about

²² Decety and Cowell, 'The Equivocal Relationship Between Morality and Empathy', 283.

²³ Andrea Glenn, Adrian Raine, and Robert Schug, 'The Neural Correlates of Moral Decision-Making in Psychopathy', *Molecular Psychiatry* **14** (2009), 5–6.

²⁴ Molly Crockett, Luke Clark, Marc Hauser, and Trevor Robbins, 'Serotonin Selectively Influences Moral Judgment and Behavior Through Effects on Harm Aversion', *Proceedings of the National Academy of Sciences* **107**:40 (2010), 17433–17438.

²⁵ Crockett, Clark, Hauser, and Robbins, 'Serotonin Selectively Influences Moral Judgment', 17437.

Moral Enhancement as a Collective Action Problem

the behaviour-modifying potential of these drugs in a more recent contribution to a symposium on moral enhancement:

Most neurotransmitters serve multiple functions and are found in many different brain regions [...] serotonin plays a role in a variety of other processes [than harm aversion] including (but not limited to) learning, emotion, vision, sexual behavior, sleep, pain and memory, and there are at least 17 different types of serotonin receptors that produce distinct effects on neurotransmission. Thus, interventions [...] may have undesirable side effects, and these should be considered when weighing the costs and benefits of the intervention.²⁶

Differences in how neurotransmitters influence the activity of neural circuits, and how increasing levels of neurotransmitters might also influence this activity, suggest that there would be different effects on the behaviour of healthy people taking SSRIs, including no effects at all. It is unclear whether or how these drugs would influence cognitive-emotional processing. Variable responses to psychotropic drugs could mean that there would not be the uniformity of effects on neural circuitry necessary to reach a threshold of enhanced moral behaviour among a majority of the population. Also, while neurotransmitters have a critical role in regulating the activity of these circuits and the mental processes they sustain, these are not a function of these substances alone but of the influence of genetic, endocrine, immune, and environmental factors on neural circuitry as well. The capacity for moral sensitivity and moral reasoning, and impairments in these capacities, depend on more than the function or dysfunction of a particular neurotransmitter. It is not known how increasing levels of serotonin would affect all of the serotonergic receptors in all of the circuitry of the distributed neural network that partly regulates moral reasoning. So it is unclear whether or to what extent increasing levels of serotonin would make one more responsive to moral reasons when acting. Increasing harm aversion would not imply a corresponding increase in moral sensitivity and make one

²⁶ Molly Crockett, 'Moral Bioenhancement: A Neuroscientific Perspective', *Journal of Medical Ethics* 40:6 (2014), 370–371, at 370. See also Molly Crockett, 'Morphing Morals: Neurochemical Modulations of Moral Judgment and Behavior', and Ricardo de Oliveira-Souza, Roland Zahn, and Jorge Moll, 'The Neuropsychiatry of Moral Cognition and Social Conduct', both in S. Matthew Liao (ed.), *Moral Brains: The Neuroscience of Morality* (Oxford: Oxford University Press, 2016), 237–245, 203–236.

more attentive to the needs of others. It would not necessarily enhance the motivation to act in response to these needs. A small number of people with enhanced moral behaviour could have some positive effects on interpersonal relations. But unless a majority did this, the effects on reducing collective harm and increasing collective welfare would likely be negligible. Regarding the neurobiological basis of moral behaviour, because no two people's brains are alike and probably would not respond to the drugs in the same way, it is unlikely that broad use of an SSRI such as citalopram would move people to act morally and reduce harm.

The beta-adrenergic receptor antagonist propranolol has been used as an indirect means of cognitive enhancement. The drug inhibits the release of the stress hormones adrenaline and noradrenaline in response to perceived threatening stimuli. This in turn dampens the autonomic response to stress in the form of increased heart rate, palpitations, sweating and other symptoms. By dampening this response, the drug may enable one to avoid being distracted by these symptoms and remain focussed on a demanding cognitive task such as surgery, musical performance, or public speaking. The drug could also be used as one component of moral bioenhancement. One study has shown that propranolol can reduce implicit negative racial bias.²⁷ As with drugs used to raise levels of serotonin, however, one cannot infer that a reduction in a fearful or negative perception of groups and harm aversion will result in an increase in morally justifiable actions. Reducing bias would not necessarily translate into pro-social behaviour. The cognitive and affective capacities necessary for co-operation require much more than harm aversion or an absence of bias. Racial bias detected in lab experiments is not predictive of immoral behaviour in real life. In addition, like other pharmacological agents used for moral enhancement, trade-offs between positive and negative effects of propranolol in the body and brain would have to be considered.

Some neuroscientists have claimed that the neuromodulating effects of the neuropeptide oxytocin may have the greatest potential as a morality-enhancing agent. Oxytocin plays a critical role in social cognition.²⁸ Its highest levels are found in the hypothalamus,

²⁷ Sylvia Terbeck, Guy Kahane, and Sarah McTavish, 'Propranolol Reduces Implicit Negative Racial Bias', *Psychopharmacology* 222:3 (2012), 419–424.

²⁸ Heather Ross and Larry Young, 'Oxytocin and the Neural Mechanisms Regulating Social Cognition and Affiliative Behavior', *Frontiers in Neuroendocrinology* 30:4 (2009), 534–547. See also, Jennifer

Moral Enhancement as a Collective Action Problem

and it influences activity in the hypothalamic-pituitary-adrenal (HPA) axis by inhibiting the fear response to social stimuli in the amygdala. Increasing levels of oxytocin in the brain through intranasal administration could reduce fear and increase trust and social co-operation. Yet any positive social effects of oxytocin may be more local than global and limited to particular groups, with negative effects more likely to occur outside the compass of these groups. Studies show that this neuropeptide facilitates social bonding but also produces non-prosocial effects that may have evolved to promote offspring survival.²⁹ Oxytocin may promote antisocial rather than prosocial behaviour on a broad scale by strengthening a person's bonding and identification with an in-group and the perception of those in out-groups as competitors or threats. It could promote rather than prevent or reduce aggression between groups and individuals.³⁰

It would be an oversimplification to claim that increasing the level of a psychoactive substance would make one less self-regarding and more other-regarding. Moral behaviour is a function of multiple biological, psychological, and environment factors, not just neural chemicals and circuits. It is a disposition that depends on factors both inside and outside of the brain. The idea that pharmacological neuromodulation alone could enhance moral behaviour fails to appreciate the complexity of human moral psychology.³¹

Research in the form of randomised placebo-controlled clinical trials would be the only empirically verifiable way to determine the effects of psychotropic drugs or other brain- and mind-altering interventions on the neural networks mediating moral reasoning. These trials would be necessary to determine whether or to what extent the drugs could enhance moral behaviour. The trials would be

Bartz, Jamil Zaki, Niall Bolger, and Kevin Ochsner, 'Social Effects of Oxytocin in Humans: Context and Person Matter', *Trends in Cognitive Sciences* 15:7 (2011), 301–309, and Patricia Churchland, *Braintrust: What Neuroscience Tells us About Morality* (Princeton: Princeton University Press, 2011), 63–94.

²⁹ Rene Hurlemann and Dirk Scheele, 'Dissecting the Role of Oxytocin in the Formation and Loss of Social Relationships', *Biological Psychiatry* 79:3 (2016), 185–193.

³⁰ Andreas Bartels, 'Oxytocin and the Social Brain: Beware the Complexity', *Neuropsychopharmacology* 37:8 (2012), 1795–1796.

³¹ Among those who make this point are Nicholas Agar, 'A Question about Defining Moral Bioenhancement', *Journal of Medical Ethics* 40:6 (2014), 369–370, and Harris Wiseman, *The Myth of the Moral Brain: The Limits of Moral Enhancement* (Cambridge, MA: MIT Press, 2016).

necessary to establish a ratio of potential benefits to risks for research subjects taking these drugs. But there would be methodological and ethical problems with conducting this research.

3. Research Challenges

There are two questions about research into MB that are especially ethically fraught. The first question is whether any risk to which human subjects participating in clinical trials testing the drugs would be exposed could be ethically acceptable. Would the potential of the drugs to produce toxic levels of the relevant neurotransmitters or cause other adverse effects influence the permissibility of conducting these trials? The second question is how the outcome would be assessed with regard to the research question driving the trial. How would we know whether a drug actually enhanced moral behaviour?

In medical research, clinical trials are necessary to determine the safety and efficacy of any potential drug therapy. The aim of a Phase I trial for a drug is to determine its safety by determining its toxicity in human subjects following tests in animal models. Toxicity is measured in terms of the highest dose a human can tolerate without serious side effects. This is necessary in order to move to Phase II and III trials testing the efficacy of a drug or procedure in treating a disease or condition. In an MB trial, the main concern about safety would arise when researchers increased normal levels of a neurotransmitter with a drug to determine whether this could increase moral sensitivity. What the research would have to determine is how much of an increase in the relevant neurotransmitter would be optimal for achieving this effect. Again, though, not all people's brains are alike. Different people have different optimal levels of serotonin and other substances in their brains regulating mood, motivation, and behaviour. When researchers tried to determine the dose of a drug necessary to increase harm aversion, some subjects could be harmed by an excess of the neurotransmitter in their brains, which could rise to a toxic level. For example, increasing normal levels of serotonin in the brain has resulted in the serotonin syndrome, which can be caused by the adverse interaction of two or more drugs, one of which is an SSRI. This syndrome occurs in approximately 14–16% of persons who overdose on a drug in this class.³² Its symptoms may include euphoria, rapid muscle

³² Edward Boyer and Michael Shannon, 'The Serotonin Syndrome', *New England Journal of Medicine* **352**:11 (2005), 1112–1120.

Moral Enhancement as a Collective Action Problem

contractions, hyperthermia, and in severe cases coma and death. Similarly, in an attempt to determine the optimal level of oxytocin to promote trust and pro-social behaviour, it is possible that some researchers would unwittingly administer a dose of this neuropeptide that would raise its level above what the brain could tolerate. Like the serotonin syndrome, this could result in deleterious effects in the brain.

In response to the concern about risk of drug toxicity in an MB trial, some might point out that healthy subjects participating in drug trials for medical treatments are also exposed to risk. But the second type of trial is necessary to test drugs for treating disease. Healthy subjects in an MB trial would be exposed to risk in testing a drug for a condition that is not a disease in the generic medical sense of mental or physical dysfunction. The medical or non-medical purpose of the trial could influence the acceptability of risk. There would be no grounds for medicalising immoral or amoral behaviour because such behaviour is not simply a manifestation of dysfunction in the brain, mind, or body. Immorality is not a disease in the medical sense relevant to the ethical criteria necessary to conduct clinical trials. More precisely, while some types of neurophysiological dysfunction may correlate with some types of immoral behaviour, many factors in addition to processes in the body and brain are needed to explain this behaviour. Others might object that it would be paternalistic to deny healthy subjects the opportunity to participate in an experiment testing the effects of drugs intended to improve behaviour. But investigators conducting the trial would have a duty of nonmaleficence not to expose healthy subjects to unreasonable risk. The potential for positive behaviour modification would probably not be enough to justify exposing these subjects to any risk of neurobiological sequelae. These considerations of risk underscore the difficulty investigators would have in obtaining ethics approval from the appropriate regulatory body to conduct the trials.³³ The categorical difference between the physiological risk and potential behavioural benefit would make it difficult to calculate a risk-benefit ratio and judge that the risk was justifiable. Assuming that the risk of adverse neuropsychiatric outcomes from drugs increasing neurotransmitter levels was low, it would still be difficult to justify exposing subjects to *any* risk in testing the effects of a drug used for a condition that was not a disease.

³³ Ezekiel Emanuel, Christine Grady, Robert Crouch, Reidar Lie, Franklin Miller, and David Wendler, *The Oxford Textbook of Clinical Research Ethics* (New York: Oxford University Press, 2008).

Many people would argue that any improvement in moral behaviour would not offset the risk of neuropsychiatric sequelae from psychoactive drugs used in an MB trial. This would limit voluntary enrolment. Altruistic individuals might agree to be included in the earliest phase of the trial testing drug safety. They would already have a high level of moral sensitivity and would not need moral enhancement. In these respects, they would be exposing themselves to some risk of harm without any compensatory benefit. Self-interested individuals with less moral sensitivity might agree to be included in a later phase of a trial testing drug efficacy. They might benefit from an increase in their moral disposition if the drugs were deemed effective. This could generate unfair inequality between altruistic and self-interested subjects in the distribution of risks to benefits. Such an unequal distribution of potential negative and positive effects from the drugs is an additional reason why research necessary to make MB available to all would have difficulty meeting ethics standards.

Pharmaceutical companies have decreased development of new antidepressant and antipsychotic drugs and funding the research necessary to market them.³⁴ The estimated profit from developing and testing these drugs may not be enough of a return on their investment to fund the studies. This may limit the development of new therapies that could treat psychiatric diseases more effectively. These companies would be even more reluctant to underwrite drug trials for moral enhancement because the outcomes of the trials could not be measured by the standard quantitative and qualitative methods used in research designed to control or prevent diseases. Lack of funding from pharmaceutical companies or other private sources would mean that much if not all of the funding for MB research would have to come from publicly funded health care institutions such as the US National Institute of Mental Health (NIMH). In addition to the challenge of recruiting enough subjects for enough trials to determine the drugs' safety and efficacy, the research would be costly. With limited public resources to conduct research, this would raise the question of whether research into drugs that could enhance moral behaviour should be treated on a par with research aimed at developing safer and more effective treatments for and prevention of neuropsychiatric and other disorders. Some might argue that the harm resulting from behaviour symptomatic of our limited moral

³⁴ Steven Hyman, 'Psychiatric Drug Development: Diagnosing a Crisis,' *Cerebrum*, April 2013: http://www.dana.org/Cerebrum/2-13/Psychiatric_Drug_Development_Diagnosing_a_Crisis/.

Moral Enhancement as a Collective Action Problem

outlook is as significant as the harm resulting from disease. But the salience and global burden of neurological, psychiatric, and other diseases, as documented by empirical data, give priority to treating or preventing these diseases over trying to improve moral behaviour.

In her discussion of the ‘urgency of moral enhancement’, Elizabeth Fenton points out that ‘if we do not continue scientific research into enhancement, if we halt it out of concern for the consequences, then we have no hope of achieving the great moral progress that will ensure the survival of our species’.³⁵ This comment raises three issues. First, it is questionable whether scientifically sound clinical trials testing drugs for moral bioenhancement have begun or could begin, much less whether they should continue. These trials would be fundamentally different from studies of harm aversion and racial bias because the research question driving them would be different. Second, if the consequences of the research included significant neurological and psychiatric risks, then these risks could make the research ethically unacceptable. Third, we cannot assume that any evidence of the drugs enhancing moral behaviour would be evidence of moral progress.

Among the problems in trying to compare research into treating disease with research into moral enhancement is that the outcomes in the first type are measured *empirically* while outcomes in the second type would be measured *normatively*. The outcome of an MB study would be assessed in terms of social expectations and whether people’s actions met these expectations and displayed respect for others. This is part of a broader problem of defining what constitutes moral and immoral behaviour and how public consensus on a definition of “moral behaviour” could be reached. Research into MB would involve very crude measures that may not tell us much or anything about whether it actually increased or improved people’s moral sensitivity. Most medical research relies on quantitative measures of outcomes, such as the five-year survival rate of subjects receiving a new oncology drug for cancer. Some of the research relies on qualitative measures, such as reports from subjects about symptom relief in a psychiatric disorder or chronic pain. Whether a drug improves people’s moral behaviour cannot be measured by these models. The lack of empirical measures of outcomes of a prospective MB study would leave uncertainty about how to assess the efficacy of the intervention. It is not clear how we could judge the success or

³⁵ Elizabeth Fenton, ‘The Perils of Failing to Enhance: A Response to Persson and Savulescu’, *Journal of Medical Ethics* 36:3 (2010), 148–151, at 148.

failure of research other than observing people's behaviour after they had taken a drug for moral enhancement. But which criteria would we use to judge what we observe and whether the behaviour displayed greater moral sensitivity than before and led to a reduction in harm?

One proposed way of confirming whether the drugs enhanced moral behaviour would be to present hypothetical cases or thought experiments of moral conflict to subjects and assess their responses to them. The Trolley Problem is the most well-known of these methods. Philosopher Philippa Foot introduced this problem as a way of distinguishing what we owe to people in the form of aid from what we owe to them in the form of non-interference.³⁶ Psychologists and cognitive neuroscientists have adopted this problem in attempting to gain a better understanding of the neurobiological basis of moral judgement.³⁷ In a number of studies, subjects have been presented with two scenarios. In the first scenario, an out-of-control trolley is headed toward five workers on a track. The driver could turn the trolley onto another track on which there is one worker. Turning the trolley would save the five on the first track but would kill the one on the second. Subjects are asked whether it would be permissible to turn the trolley and kill the one worker rather than the five. In a variant of this case, subjects are asked if it would be permissible to push a fat man off a bridge and onto the track to stop the trolley. This action would stop the trolley

³⁶ Philippa Foot, 'The Problem of Abortion and the Doctrine of the Double Effect', in Foot, *Virtues and Vices* (Oxford: Oxford University Press, 2002), 19–32. This problem has been discussed by many other philosophers, most notably Judith Jarvis Thomson, 'Killing, Letting Die and the Trolley Problem', *The Monist* 59:2 (1976), 204–217, and F. M. Kamm, Judith Jarvis Thomson, Thomas Hurka, and Shelly Kagan, *The Trolley Problem Mysteries* (New York: Oxford University Press, 2016).

³⁷ Joshua Greene, Leigh Nystrom, Andrew Engell, John Darley, and Jonathan Cohen, 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment', *Neuron* 44:2 (2004), 389–400; Joshua Greene, 'Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains', *Trends in Cognitive Sciences* 11:8 (2007), 322–333; Michael Koenigs, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, and Antonio Damasio, 'Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments', *Nature* 446:7138 (2007), 908–911; and Carla Harenski, Olga Antonenko, Matthew Shane and Kent Kiehl, 'A Functional Imaging Investigation of Moral Deliberation and Moral Intuition', *NeuroImage* 49:3 (2009), 2707–2716.

Moral Enhancement as a Collective Action Problem

from killing the five workers but would kill the fat man. Whereas most subjects responded that it would be permissible to turn the trolley in the first case, there was disagreement about whether it would be permissible to push the fat man off the bridge to stop the trolley in the second case. Some investigators have claimed that the impermissibility of pushing the man reflects a deontological judgement, while the permissibility of pushing him reflects a consequentialist judgement.³⁸ Some of these subjects have been put through fMRI scanners when presented with the question about what to do with the trolley or fat man. Recorded differences in neural activity in different brain regions during their responses presumably correspond to deontological or consequentialist judgements about the right action in these hypothetical cases.

These judgements are not accurate reflections of moral reasoning because they involve only quick intuitive responses that fall short of the consideration and argumentation necessary to justify an action or policy.³⁹ This, combined with the fact that fMRI images are scientific constructs rather than actual pictures of brain activity, provides a good reason to be sceptical of combining brain imaging and thought-experiments to assess moral reasoning and moral judgement after taking a psychotropic drug. “Trolley cases” involve unavoidable harm and the question of whether it is permissible to cause a lesser harm to prevent a greater harm. Most of the harm committed by humans is not the result of actions they perform or fail to perform in times of moral conflict but from the more general failure to recognise and react to reasons for performing certain actions. Thought-experiments like trolley cases are generated in artificial settings and tell us nothing about how people act in the real world. There is often a gap between moral judgement and action. One may judge that A is the right thing to do in a situation but choose B when they are actually in that situation. Whether a drug effectively enhanced moral behaviour could not be determined by

³⁸ Greene, ‘Why are VMPFC Patients More Utilitarian?’, and Koenigs, Young, Adolphs, Tranel, Cushman, Hauser, and Damasio, ‘Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments’.

³⁹ For criticism of these experiments and what they presumably tell us about moral judgement, see Kwame Anthony Appiah, *Experiments in Ethics* (Cambridge, MA: Harvard University Press, 2008), Selim Berker, ‘The Normative Insignificance of Neuroscience’, *Philosophy & Public Affairs* 37:4 (2009), 293–325, and F. M. Kamm, ‘Neuroscience and Moral Reasoning: A Note on Recent Research’, *Philosophy & Public Affairs* 37:4 (2009), 330–345.

asking questions about what one would do in hypothetical cases but by observing people's actual behaviour.

Unlike the outcome of a clinical trial in medical research, whether a psychotropic drug enhanced moral sensitivity would not be amenable to assessment by quantitative or qualitative methods. The end-point of an MB trial would be open to varying interpretations of whether people displayed greater attention to and respect for others after taking a drug. There would be considerable variability in the behaviour of subjects after taking a brain-altering substance because of the complex interaction between and among biological, psychological, and environmental factors and the unique effects these factors have in each person's brain. It is thus unlikely that there could be a definitive answer to the research question of whether a psychopharmacological intervention enhanced moral behaviour. The outcome would likely be too crude to measure by any scientific model of research.

4. Co-operate or Defect?

Although particular philosophical arguments fail to provide the motivation to act morally, we need a general moral theory to serve as a framework in which to spell out the goal of moral enhancement and the process by which it might be achieved. This would require public consensus on which theory to adopt, a theory that ideally would accurately reflect human moral psychology. It would be overly simplistic to expect one theory to be sufficient for this purpose. Insofar as the goal of moral enhancement is to reduce harm, some form of consequentialism would be needed. Some form of deontology would also be necessary to recognise people's rights, especially the negative right to non-interference. Still, given that most people are mainly self-interested, something more than recognition of others' rights and the value of preventing or reducing harmful outcomes of actions or omissions would be needed to move people to act appropriately. An additional theory would have to complement consequentialism and deontology to ground the requisite motivational force. Most people are not naturally altruistic and do not typically act from other-regarding reasons. Some of us act altruistically at times, but not often enough and not to the extent necessary to have sustained beneficial effects on other people and their lives. The most realistic theory would be one consistent with a lower common behavioural denominator – rational self-interest. Accordingly, some version of social contract theory would be the most plausible moral complement to consequentialism and

Moral Enhancement as a Collective Action Problem

deontology as a theoretical basis for a moral enhancement programme.⁴⁰ It would be based on rational choice and promote social co-operation for mutual benefit. Each person would give up some of her self-interest by co-operating; but all would be better off by doing this. Such a model would be imperfect. In particular, future people who would be more adversely affected by the consequences of our present policies and actions or omissions could not participate in such a contract. A more demanding moral theory would be necessary to respect their claims. Given our limited moral compass and tendency to discount the future, though, a social contract model of moral enhancement, agreed upon by the present generation, would probably have the best chance of achieving a reasonably modest goal.

A social contract model of moral bioenhancement to promote co-operation would involve another collective action problem. A critical number of people would have to agree to take psychotropic drugs enhancing their moral disposition to produce the collective effect of reducing harm and promoting general welfare. Many would rely on this reasoning and assume that others would reason in the same way in deciding to enhance. They would calculate that the threshold effect of a large number of individuals co-operating in an enhancement programme would make co-operating and sacrificing some self-interest mutually beneficial and better for them than if they chose not to enhance. The psychology of this reasoning is similar in some respects to what is involved in the classic prisoner's dilemma. Each of two prisoners accused of a crime must decide whether they will co-operate by confessing or defect by remaining silent or testifying against the other. Defecting would be the most rational decision for each independently of how the other reasons and decides what to do. But each knows that the outcome is a function of what both decide, and that they would have a worse outcome if both defected. Each chooses the second best but all-things-considered most rational option of co-operating.⁴¹

This game-theoretic strategy would be of limited value in a programme of moral enhancement, however. The number of players is too large to be accommodated by these strategies, which involve

⁴⁰ Thomas Hobbes, *Leviathan*, ed. by M. Oakshott (Oxford: Blackwell, 1987/1651); David Gauthier, *Morals by Agreement* (New York: Oxford University Press, 1986); T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998); and Harris, *How to Be Good*, 156–171.

⁴¹ Richard Dawkins, *The Selfish Gene*, 30th Anniversary Edition (Oxford: Oxford University Press, 2006), 202–233.

Walter Glannon

cooperation 'in cozy little local enclaves'.⁴² While these enclaves may evolve into larger entities, they would still be local or relatively small in scale. The collective action problem of MB would be more difficult to resolve because it would be global and involve a very large number of people. The main problems in these scenarios would be having enough people enhancing to produce a positive collective effect and preventing some from free riding on the co-operation of others.⁴³ If less than a critical mass of people decided to cooperate, then the goal of reducing harm would not be achieved. Even if a drug could safely and effectively enhance moral behaviour, some would refuse to enhance. They might not trust others to voluntarily enhance and might not want to sacrifice any self-interest for fear of becoming worse off than they would be if they pursued self-interest in an unconstrained way. In a scenario where many people enhanced and a threshold of co-operation had been reached, some would calculate that they could refuse to enhance and benefit from the cooperation of others without sacrificing any self-interest. A few of these who engaged in the most harmful behaviour and needed MB the most would not only refuse to co-operate but would also take advantage of the co-operation of others for their own malevolent ends. In these scenarios, defectors would be making a rational choice. Given the sheer number of people involved, it would be difficult to prevent free riding or manipulation and implement and enforce measures that would punish people who did this.

The number of people and the behaviour that would thwart the realisation of the goals of MB may depend on which moral problem one focussed. Climate change would involve the largest number of players, and the harmful outcome would result from many individual omissions as much as from many individual actions. Civil wars and ethnic conflicts typically involve a smaller number of decision-makers and their more numerous followers adversely impacting the lives of others. The threat of WMDs lies primarily with a few agents or even one individual. An irrational decision by what Martin Rees calls one 'village idiot' controlling these weapons could have catastrophic consequences.⁴⁴ In the second and third scenarios, even if a majority of people chose MB, a small number or even one person who refused to co-operate in a project to prevent or reduce the threat of harm could be enough to doom it.

⁴² Dawkins, *The Selfish Gene*, 219.

⁴³ Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press, 1965).

⁴⁴ Martin Rees, *Our Final Century* (London: Arrow Books, 2004), 61.

Moral Enhancement as a Collective Action Problem

To prevent the idiot's irrational decision and its consequences in Rees' hypothetical scenario, Harris would argue that what we need is not moral but cognitive enhancement aimed at improving rationality. Yet because reasons to prevent outcomes adversely affecting other people are moral reasons, and because moral reasoning involves cognitive and emotional processing, preventing catastrophic outcomes would depend not only on enhanced cognition but also on enhanced cognitive and emotional capacities.

Harris says that '[e]thics is for bad guys! The good don't need ethics'.⁴⁵ Further, he claims that 'ethics is for those occasions in which compassion, altruism, and basic decency fail'.⁴⁶ MB is Persson and Savulescu's way of providing the ethics we need. But too many of the bad guys among us would choose not to enhance. In light of this and the likely general failure of voluntary MB to achieve its normative goal, Persson and Savulescu state: '[i]f safe moral enhancements are ever developed, there are strong reasons to believe that their use should be obligatory [...]. That is, safe, effective moral enhancement would be compulsory'.⁴⁷ The argument for compulsory MB should not just be to get the bad guys on Harris' interpretation to improve their behaviour. It should be aimed not only at those whose actions and omissions incrementally contribute to climate change, for example, but also and indeed more so at the "worst guys", those whose actions directly cause substantial harm to others. On an ideal reading of the prisoner's dilemma and the standard collective action problem, "nice guys finish first" if they reason that co-operating is in their all-things-considered best interests and do in fact co-operate.⁴⁸ Yet the likely scenario of morally deficient persons taking advantage of those who voluntarily enhanced their moral sensitivity would mean that the worst guys who produced most of the world's harm would "finish first" in achieving their own self-serving ends. Incentives to nudge those with the greatest need of moral enhancement would do little to motivate them to act in accord with

⁴⁵ Harris, *How to Be Good*, 112.

⁴⁶ Harris, *How to Be Good*, 112.

⁴⁷ Persson and Savulescu, 'The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity', *Journal of Applied Philosophy* 25:3 (2008), 162–177, at 173. Also, Persson and Savulescu, 'Moral Enhancement, Freedom and the God Machine', *The Monist* 95:3 (2012), 399–421, *Unfit for the Future*, and 'Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement', *Bioethics* 27:3 (2013), 124–131.

⁴⁸ Dawkins, *The Selfish Gene*, 219, and Olson, *The Logic of Collective Action*.

moral norms.⁴⁹ Because of the extent of their moral deficiency, changing their behaviour would take much more than incentives or nudges. The refusal of the worst perpetrators of wrongdoing to do this voluntarily could make state-sponsored compulsory enhancement necessary to reduce interpersonal harm. Could this be justified?

5. Compulsory Moral Enhancement: The Cost to Freedom

Compulsory MB would depend on research demonstrating the safety and efficacy of the drugs used to enhance. If MB were compulsory, then would participation in this research also be compulsory? Any action regarding MB could in principle be obligatory only if it was proven safe and effective, which the research itself is designed to determine. Since the research would be necessary to determine the safety and efficacy of the drug for public administration and consumption, compulsory participation in the research itself would be unjustifiable. This would be significantly different from military conscription or jury duty, which are compulsory but based on established social and legal institutions that protect people's rights and liberties. It would also be different from requiring young adults to purchase health insurance to ensure a fair distribution of risk across populations.

Harris has argued elsewhere that we have a moral duty to participate in medical research.⁵⁰ This obligation is generated by the fact that people in the present generation have benefited from medical treatments resulting from the sacrifices of people in the past who participated in medical research. These subjects exposed themselves to some risk without any direct therapeutic benefit. We in the present generation have benefited from the actions of those in the past in the form of better treatment and prevention of many diseases. Because of this, we have an obligation to participate in medical research that could benefit future people. This argument about a fair intergenerational distribution of burdens and benefits does not apply to MB, for three reasons. First, morally deficient behaviour is not a disease in the sense of physical or mental dysfunction and thus involves a different assessment of risk of psychotropic drugs intended to improve it. Second, none of us in the present generation has benefited from the results of previous research into MB, since there was no such research. Third, having a duty to participate in research

⁴⁹ Cf. Wiseman, *The Myth of the Moral Brain*, 8–9, 253ff.

⁵⁰ John Harris, 'Scientific Research is a Moral Duty', *Journal of Medical Ethics* 31:4 (2005), 242–248.

Moral Enhancement as a Collective Action Problem

is different from being compelled to participate in it. One can voluntarily fail to discharge a duty, which is not an option in compulsory action. The potential harm from being exposed to any risk of tinkering with neurotransmitters, hormones, or neuropeptides in the brain for the purpose of possibly improving moral behaviour undermines any claim for an obligation to participate in moral bioenhancement research. Harris does not extend his argument about an obligation to participate in medical research to research into MB because of his rejection of the idea of compulsory MB.

Even if one believed that safe and effective MB should be compulsory for all citizens in a liberal democratic society, there would remain the daunting task of enforcing compliance. This would require different levels of co-ordinated social and political action. It would also assume the moral integrity and public acceptance of those empowered with overseeing these tasks. This may assume too much. Harris appropriately asks, '[w]ho guards the guardians?'⁵¹ More fundamentally, even if compulsory MB significantly reduced harm, it could come at an unacceptable cost: it would leave no space for freedom.⁵² For some, the magnitude of the actual and potential harm resulting from voluntary action might be significant enough to justify imposing limits on action. For others, no amount of harm from our actions could justify such limits. Compulsory MB would undermine the free choice necessary for moral responsibility, praise, blame, and other normative concepts and practices on which our social and political institutions are based and which define us as human agents. By ensuring that everyone always acted in a certain way, MB would eliminate the capacity to choose between different courses of action – good and bad, beneficial and harmful.⁵³ Without this capacity, we would no longer be moral agents. As Harris puts it: '[a]gents are quintessentially actors: to be an agent is to be capable of action. Without agency in this sense, decision-making is [...] morally and indeed practically barren'.⁵⁴ The

⁵¹ Harris, *How to Be Good*, 105.

⁵² Harris, *How to Be Good*, 92. Also, Harris, 'Moral Enhancement and Freedom', *Bioethics* 25:2 (2011), 102–111, and Harris, 'Moral Progress and Moral Enhancement', *Bioethics* 27:5 (2013), 285–290. Cf. David DeGrazia, 'Moral Enhancement, Freedom and What We (Should) Value in Moral Behaviour', *Journal of Medical Ethics* 40:6 (2014), 361–368.

⁵³ For discussion, see Michael Hauskeller, *Better Humans? Understanding the Enhancement Project* (Durham: Acumen, 2013), 35–54, and Hauskeller, 'Is it Desirable to Be Able to Do the Undesirable? Moral Bioenhancement and the Little Alex Problem', *Cambridge Quarterly of Healthcare Ethics* 26:3 (2017), 365–376.

⁵⁴ Harris, *How to Be Good*, 94.

Walter Glannon

collective interest in preventing or reducing harm could never justify violating what is inviolable. Harris adds, 'I, like so many others, would not wish to sacrifice freedom for survival'.⁵⁵

What he calls the 'freedom to fall' has its limits, however.⁵⁶ Free actions often have consequences that adversely affect other people. We should have the freedom to fall provided that we do not cause others to fall along with us. This is consistent with John Stuart Mill's principle of liberty. Mill states that 'over himself, over his own body and mind, the individual is sovereign'.⁵⁷ He qualifies this statement in saying that 'the only purpose for which power can be rightfully exercised over any member of a civilized society, against his will, is to prevent harm to others'.⁵⁸ The question for MB is whether and in which circumstances the freedom to fall could be overridden by the collective interest in preventing harm. At one end of the harm spectrum, it would take only one "village idiot" to cause or initiate a process resulting in the extinction of the human species through the use of nuclear weapons. At the other end of the harm spectrum, extinction could also eventually be our fate from the collective effects of individual actions and omissions causing climate change. If none of us survived, then freedom to fall would have no value because there would no longer *be* any human agents, free or unfree.

Some might equate moral progress with moral enhancement insofar as enhancement resulted in a reduction of harm among humankind. But harm reduction alone would not be sufficient for a robust conception of moral progress. Indeed, depriving people of their autonomous agency through compulsory MB would eliminate the "moral" in moral progress. If being a moral agent presupposes the ability to choose between alternative courses of action, and if there were no choice under compulsory MB, then there would be no moral agents. Our actions would not be our own but the products of brain-altering interventions. There is something paradoxical about the idea of moral progress achieved through the elimination of or limitations on moral agency. Any plausible concept of moral progress presupposes responsibility, praise, blame, and other normative concepts, as well as the associated psychological properties of conscientiousness, remorse, and regret. These depend on the capacity to decide and act on our own

⁵⁵ Harris, *How to Be Good*, 74–75.

⁵⁶ Harris, *How to Be Good*, 74, 81.

⁵⁷ John Stuart Mill, *On Liberty*, ed. by G. Himmelfarb (London: Penguin, 1974/1859), 119.

⁵⁸ Mill, *On Liberty*, 119.

Moral Enhancement as a Collective Action Problem

considered desires, beliefs and reasons, to overcome weakness of will through our own efforts and to learn from our mistakes. These capacities are anathema to the idea of forcibly taking or undergoing a neuro-modulating drug or technique to prevent us from performing actions with potentially harmful consequences. Still, the value of reducing harm in promoting the survival of the human species cannot be minimised. Having choice is necessary for moral agency and responsibility. At the same time, reducing or preventing harm resulting from bad or evil choices is necessary for us and especially future generations to have the space for agency. There may be an intractable conflict between individual liberty and collective interest, between the freedom to fall and compulsory MB to reduce or prevent harm.

6. Conclusion

Moral bioenhancement would be a way of increasing people's moral sensitivity in motivating them to act in ways that respected the rights, needs, and interests of others. The aim would be to reduce actual harm and prevent future harm caused by humans at individual and collective levels. Persson and Savulescu's argument for MB as a response to the magnitude of global harm is well-motivated. But it is fraught with scientific, ethical, social, and political problems that would probably doom it as a viable normative project. Voluntary MB would likely fail to reach a critical level to have a positive collective effect because a significant number of people would refuse to enhance. Among those refusing to enhance would be individuals who contributed most to global harm and with respect to whom there would be the most compelling reasons for enhancement. Yet making it compulsory would threaten to undermine freedom of choice and cognitive liberty. While there is an urgent need to protect the collective interest in avoiding harm and promoting the survival of the species, it is doubtful that compulsory MB eliminating free choice would be justifiable. Questions about voluntary or compulsory MB also depend on the scientific question of whether research could determine that drugs designed to enhance moral behaviour were safe and effective. Yet clinical trials designed to test these drugs for this purpose would likely fail to meet standard ethical criteria for an acceptable risk-benefit ratio in medical research. There are also questions about whether drugs intended to modulate brain circuits and networks mediating moral reasoning and decision-making alone would strengthen moral motivation and enhance moral behaviour.

Allen Buchanan claims that ‘biomedical intervention might be one aspect of a multifaceted effort to extend concern and respect to all human beings, not just those who are like us’.⁵⁹ This is consistent with what Persson and Savulescu say about MB ‘as a complement to traditional means’.⁶⁰ The other aspects of the effort Buchanan describes would be psychological and social. Yet even if biomedical interventions were proven to be safe and effective, it is unclear how all three aspects of a bio-psycho-social model would be integrated and how such a model would be implemented and enforced. Acknowledging the hypothetical nature of MB and the obstacles that such a programme would face, Persson and Savulescu state that the development and perfection of moral enhancement is ‘not likely to be possible in the near future’.⁶¹ Harris makes a stronger claim against MB, stating ‘I believe it will never be possible to the extent the Persson/Savulescu thesis requires, or indeed that Tom Douglas believes [...] because moral enhancement has little prospect of preventing idiocy – but of course I could be wrong’.⁶² Perhaps the strongest claim among sceptics of moral bioenhancement is from Harris Wiseman, who says that ‘an explicit project of state-sponsored moral improvement of the general public is *unthinkable* in liberal states’.⁶³ This is in reference to what he calls ‘hard, fine-grained, moral enhancement’ of the compulsory type defended by Persson and Savulescu.⁶⁴ Yet what Wiseman calls a ‘soft, coarse-grained, moral enhancement’ project consisting of nudging and incentives would also likely fail.⁶⁵ These strategies would not provide the necessary motivational force for enough people to co-operate in a moral enhancement project. Nor would they prevent some from taking advantage of this co-operation for their own selfish ends. Education aimed at improving moral behaviour might reduce interpersonal harm to some extent but would not eradicate it. Moreover, not all states and their guardians would be interested in such a plan, much less implement and enforce it.

⁵⁹ Allen Buchanan, *Better Than Human: The Promise and Perils of Enhancing Ourselves* (New York: Oxford University Press, 2011), 170. See also Wiseman, *The Myth of the Moral Brain*, 280.

⁶⁰ Persson and Savulescu, *Unfit for the Future*, 2.

⁶¹ Persson and Savulescu, ‘The Perils of Cognitive Enhancement’, 175.

⁶² Harris, *How to Be Good*, 75.

⁶³ Wiseman, *The Myth of the Moral Brain*, 79.

⁶⁴ Wiseman, *The Myth of the Moral Brain*, 12, 129.

⁶⁵ Wiseman, *The Myth of the Moral Brain*, 129.

Moral Enhancement as a Collective Action Problem

Derek Parfit expresses a positive view on the possibility of moral progress: '[L]ife can be wonderful as well as terrible, and we shall increasingly have the power to make life good. Since human history may be only just beginning, we can expect that future humans, or supra-humans, may achieve some great goods that we cannot now even imagine'.⁶⁶ The goods of which Parfit writes cannot be separated from or achieved without an increase in the level of our moral sensitivity. They would be contingent on preventing or reducing harm. Moral progress requires enhanced moral sensitivity. The key issue is whether this sensitivity and progress can be made through any means. There are indeed many theoretical and practical reasons for scepticism about the concept and goal of moral enhancement. If there is no viable alternative project that could move people to act morally and resolve the most pressing issues of current and future generations, then a bad end may be in store for us all.*

University of Calgary
wglannon@ucalgary.ca

⁶⁶ *On What Matters*, Volume II, 618. I do not take human existence as such to have intrinsic value. What makes nuclear catastrophe and global warming harmful is not that they would extinguish the human species but that they would cause suffering in people who exist now and will exist in the future.

* I am grateful to the other participants in the University of Exeter conference on moral enhancement for discussion of these issues, and especially to Michael Hauskeller for very helpful comments on an earlier version of this essay.