Autonomy and the Ethics of Biological Behaviour Modification

Julian Savulescu, Thomas Douglas, Ingmar Persson

[This is a pre-publication version. The final version is available in A Akabayashi (ed) *The Future of Bioethics: International Dialogues*, Oxford University Press.]

Much disease and disability is the result of lifestyle behaviours. For example, the contribution of imprudence in the form of smoking, poor diet, sedentary lifestyle, and drug and alcohol abuse to ill-health is now well established. More importantly, some of the greatest challenges facing humanity as a whole – climate change, terrorism, global poverty, depletion of resources, abuse of children, overpopulation – are the result of human behaviour. In this chapter, we will explore the possibility of using advances in the cognitive sciences to develop strategies to intentionally manipulate human motivation and behaviour. While our arguments apply also to improving prudential motivation and behaviour in relation to health, we will focus on the more controversial instance: the deliberate targeted use of biomedicine to improve moral motivation and behaviour. We do this because the challenge of improving human morality is arguably the most important issue facing humankind (Persson and Savulescu, forthcoming). We will ask whether using the knowledge from the biological and cognitive sciences to influence motivation and behaviour erodes autonomy and, if so, whether this makes it wrong.1

1. COGNITIVE SCIENCE AND BEHAVIOUR MODIFICATION

One of the emerging subdisciplines of the cognitive sciences is the cognitive science of motivation and behaviour. Advanced techniques in neuroscience, such as functional magnetic resonance imaging, together with sophisticated pharmacological, psychological and economic experiments have begun to shed light on the subtle neural and psychological bases of motivation and behaviour. Perhaps the most controversial area of such research is investigating those features of psychology which are characteristic of our humanity: rationality and morality. Though the main controversies surrounding this research have focussed on what it tells us about the nature of rationality and morality, there is another, more neglected way in which progress in the cognitive sciences may have even more morally significant implications: it may yield new means of *modifying* morally significant aspects of motivation and behaviour (henceforth, 'moral motivation and behaviour'). In this chapter, we will explore not what neuroscience tells us about the nature of rationality and morality, but rather how ethics could justify the use of radical advances in the neurosciences for the purposes of modifying moral motivation and behaviour.

Behavioural manipulation, or "mind control" as it is often loosely put, has a bad name. Crude forms of mind control, such a brain washing or torture, have been around for millennia. Electrical brain implants were in the past used to 'treat' homosexuality, while radical psychosurgery was used to control aggression (Greely,

[1] This chapter is based on the article… forthcoming in the Monist

2008). Such forms of mind control were either used in the service of misguided goals or were performed without adequate protections for those subjected to them. However, safe and allegedly ethical means of influencing motivation and behaviour are being employed or advocated. For example, psychological research is affording strategies to influence behaviour by manipulating unconscious stimuli (Kiesel et al. 2006). One prominently discussed technique is the 'nudge' strategy, which harnesses knowledge about cognitive biases that may influence voluntary choice (Thaler and Sunstein 2008). To date, nudges have largely been advocated as means to improve health (see, for example, Charkrabortty 2008), but they can also be used to influence moral behaviour. For example opt-out systems for organ donation take advantage of a form of status quo bias to encourage registration for post-death organ donation.

'Nudging' is an example of an scientifically-informed *institutional* strategy that can be used to alter moral motivation and behaviour. Some have raised autonomy-based concerns about nudge techniques (see, for example, Bovens 2008). But in this article we will focus on what many find more concerning: techniques to alter moral motivation and behaviour that operate by directly manipulating our *biology*, not our social environment.

A number of commonly employed antidepressants and antihypertensives (Terbeck et al. Under Review b) affect moral behaviour as a side effect. Indeed, a number of drugs are already prescribed specifically for their behaviour-altering effects, some of which are morally significant: the anti-alcohol abuse drug disulfuram, the weight loss drug orlistat, and anti-libidinal agents sometimes used to reduce sexual re-offending. Neuropsychology is beginning to provide more robust evidence for biological correlates of morally relevant traits such as aggression, trust and empathy. For example, Ramachandran and colleagues have begun to identify neural loci of empathic responses in humans and animals (Ramachandran and Oberman 2006). This research may lead to pharmacological interventions to alter empathy, cooperation and trust (e.g. De Dreu et al. 2011). Indeed, our own empirical research has already shown that propranolol can reduce implicit racial bias (Terbeck et. al. 2012) and produce less utilitarian judgement (Terbeck et al. under review (b)).

Other possible techniques for biologically influencing choices include transcranial magnetic stimulation, deep-brain stimulation, transcranial direct current stimulation and optogenetics, offering the prospect of profound manipulation using genetic manipulation and targeted optic stimulation of precise areas of the brain. These technologies can directly modify behaviours, perhaps including addictive behaviour (Carter et al. 2009). Indeed, transcranial magnetic stimulation can affect behaviour without subjects' awareness (Brasil-Neto et al. 1992).

The Prospect of Moral Bioenhancement

Could such cognitive science be used to biologically improve decisions about good and bad, or right and wrong? Could it be used to biologically reduce weakness of will? At present there is no one drug or other biological manipulation that improves moral behaviour in all people in all circumstances. However, there are reasons to believe that manipulation of biology could be used to influence moral behaviour.

As discussed, hormonal manipulation to reduce libido in sexual offenders such as paedophiles is a crude form of moral enhancement. Improving impulse control with Ritalin and Adderall in children with Attention Deficit Disorder reduces imprudent and immoral behaviour, such as violence against others and may also, in some cases, be an example of crude moral bioenhancement.

The possibility also exists of moral bioenhancement amongst the general population. A number of studies have shown clear effects on moral behaviour through the hormone and neurotransmitter oxytocin. Oxytocin is known to play a key role in birth and breastfeeding. Higher levels of the hormone have been linked to maternal care, pair bonding, and other pro-social attitudes, like trust, sympathy and generosity (Insel et al. 2004).

Oxytocin levels vary and can be increased through certain external stimuli, such as sex and physical contact. It can also be elevated by a simple nasal spray, which is the delivery method of many of the experiments measuring its effects. In addition, several commonly used drugs are also thought to affect the release or metabolism of oxytocin. Over 100 million women world wide use the conbined oral contraceptive pill. An estimated 300 million people world wide suffer from asthma, for which one of the most effective treatments is glucocorticoids.
Both these are associated with an effect on oxytocin levels, with the contraceiptive pill associated with elevated baseline oxytocin levels and an increase in oxytocin secretion (Stock et al. 1994; Silber et al. 1987), and glucocorticoids thought to modulate both the release of oxytocin and the expression of oxytocin receptors in some parts of the brain (Link et al. 1993; Liberzon et al. 1997).

The effects of increased levels of oxytocin have been shown to be significant. For example, Kosfeld and collaborators investigated the relationship between oxytocin and trust in a simple game of cooperation (Kosfeld et al. 2005). In the experiment, subjects were randomised and given a spray containing either a placebo or oxytocin. They were allocated into pairs consisting of an investor and a trustee. The investor is given a sum of money and allowed to choose how much to pass to the trustee, who will receive 3 times the investor's amount . The trustee must then decide how much to return to the investor. The investor's initial decision indicates the level of trust, and the trustee's return an indication of trustworthiness and gratitude. The greater the initial payment chosen by the investor, the better the potential return for both players- but this can only be realised by the investor if the trustee responds accordingly- if they are trustworthy. The study showed that investors who had received oxytocin instead of the placebo exhibited significantly more trusting behaviour, Oxytocin has also been shown to facilitate increased co-operation in a number of other co-ordination problems (see, for example, Declerk et al 2010).

However, oxytocin's effects are complicated. Other research has shown that the pro-co-operation effects of oxytocin may be restricted to others perceived as members of the same group (De Dreu et al 2010 & 2011) Indeed, it may even reduce co-operation with out-group members. (De Dreu et al. 2010).

Another neurotransmitter implicated in moral behaviour is serotonin. Again, serotonin is naturally occurring, varying according to external stimuli, which is involved in mood regulation. It can also be affected by Selective serotonin reuptake inhibitors (SSRIs), which are commonly prescribed for depression, anxiety, and obsessive compulsive disorder. In 2003, in the UK alone, 19 million prescriptions for SSRIs were issued. SSRIs slow the reabsorption of serotonin, thereby making more of it available to stimulate receptors. But SSRIs have a side effect: they seem to make subjects more fair-minded and willing to cooperate. Tse and Bond (2002) measured its effects by observing subjects given the SSRI citalopram as they played the Dictator game. In this game, some subjects are assigned the roles of dictator and are given a sum of money to divide between themselves and another participant. Tse and Bond found that subjects who had injested citalopram made the division more equally than the control group. Conversely, studies have shown that depletion of a precursor of serotonin (tryptophan), which would in turn lead to reduced levels of serotonin, leads to lower rates of cooperation in the Prisoner's dilemma game (Wood et al. 2006). Crockett and colleagues (2008) found that lowered levels of tryptophan led to a greater propensity in subjects to reject offers perceived as unfair, relative to controls. This suggests that SSRIs, by increasing serotonin levels, affect their assessment of what counts as (unacceptably) unfair, possibly leaving them more vulnerable to exploitation.

2. THE ETHICS OF MORAL BIOENHANCEMENT

While the technology to biologically influence moral motivation and behaviour is still in its infancy, or even pre-embryonic stages, it seems likely that science will afford ever more powerful interventions. Douglas (2008) has argued that it might be permissible for individuals to use these interventions to bring about more moral motivation and behaviour in themselves, though he also raises concerns about the possible misuse of technology; concerns which may militate against seeking to develop them. By contrast, Persson and Savulescu have argued that the development of these technologies should be prioritised and aggressively pursued, such is the need for moral enhancement (Persson and Savulescu Forthcoming; Persson and Savulescu 2011; Persson and Savulescu 2010; Savulescu and Persson 2008; Savulescu and Persson 2011; Savulescu and Persson forthcoming).

One objection that is frequently raised against mind control and behavioural manipulation, even for a person's own benefit but especially when it is for the purposes of promoting more moral behaviour, is that it would compromise our freedom and autonomy. It is this objection that we wish to spell out and address in this paper. First, however, we should offer some thoughts on the likely nature of moral bioenhancement.

3. THE NATURE OF MORAL BIOENHANCEMENT

A moral enhancement is, we will assume, an intervention that makes it more likely that you will act morally, in some future period, than would have been the case if it were not used. One acts morally when one does the right thing, and for the right

reason(s). In many circumstances there would be disagreement about what actions, and reasons for acting, are right. What constitutes moral enhancement will depend on the what accounts of right action and right motivation are correct.

What constitutes right action is contested. Kantians, utilitarians, virtue theorists, deontologists and religious ethicists may all disagree on what the right action is. There is also significant disagreement about what sorts of motivations are the right ones to have and act on. Kant famously claimed that to act in a way that has true moral worth—rather than merely conforming to morality—one must act from the motive of duty (Kant 1964). This view has sometimes been taken to imply that the right motive for action is *moral reasoning*. One must reason about what duty requires and then act accordingly. But others would allow that emotions such as sympathy can produce genuinely moral action (Mill 1979, Arpaly 2003), and some would question whether moral reasoning is even capable of producing action with the support of emotions or desires that are external to reasoning itself (Hume 1978).

But despite this significant disagreement, there are areas of agreement. For example, almost every ethical theory says it is wrong to kill an innocent person in non-extra-ordinary circumstances. And on any plausible ethical theory, certain capacities will be necessary to act for the right reasons. For example, right motivation surely requires the capacity to act for the sake of others, or for morality itself, rather than in self-interest. This requires that one can conceptualise and be motivated by morality or the interests of others.

3.1 Self-sacrifice and Altruism

It is characteristic of morality, as opposed to prudence or self-interest, that it requires the sacrifice of one's own interests for the sake of others, or at least for the sake of some moral code. It is a prerequisite of moral action that one should sacrifice/constrain one's own self-interest for the sake of others or of morality.

For example, proponents of most moral theories could accept a principle of easy rescue stating that, when (i) the harm to A of Fing is small, (ii) the benefit to another/others, B, is great, and (iii) there is no harm to third parties, then A should F. Even this undemanding and relatively uncontroversial principle requires some minor sacrifices of self-interest.

Perhaps, some might argue, there is no duty of easy rescue. But any morality requires that we do not kill innocent persons, or at very least, innocent in-group members for no good reason. In some cases, it will be against our interests to refrain from killing someone. Perhaps it is frequently so. A controlling spouse, a demanding boss, a person who blocks your chances of advancement. In all these cases, morality requires that you set aside your interests and not kill.

Thus, a willingness to sacrifice one's own interests is required by even undemanding moralities. Yet it is something which, like all human characteristics, varies from person to person. Some will be less inclined to make sacrifices, or less often or of very small magnitude.

Various factors predictably increase self-sacrifice. For example, if one derives pleasure from self-sacrifice, this will increase willingness to sacrifice one's interests.

The praise or esteem of others increases self-sacrifice. Rituals, dances, induction ceremonies, et cetera have all been used increase self-sacrifice of members of groups. In the future, it will become possible to not only manipulate the situational and social determinants of self-sacrifice, but also the biological determinants.

This can be seen by comparing the sexes. It is plausible to think that in general women have a greater capacity for self-sacrifice than men. Baron-Cohen (2003) argues that women have a greater capacity for empathy than men. If men could be made more like typical women with respect to empathy, this might increase their willingness to sacrifice one's own interests for the sake of others at least in some circumstances. This could qualify as a moral enhancement on almost any widely accepted account of morality.

Of course, the acquisition of any single trait which contributes to more moral behaviour can be used for immoral purposes. For example, greater willingness to self-sacrifice will not always qualify as a moral enhancement. Some Nazis described having to overcome their revulsion at killing Jews for the sake of others (future generations of Aryan Germans). Heinrich Himmler reportedly told his masseur that "It is the curse of greatness that it must step over dead bodies to create new life. Yet we must . . . cleanse the soil or it will never bear fruit. It will be a great burden for me to bear" (Kersten 1956, p. 120). Himmler would not have been morally enhanced by an intervention that made him even more willing to kill Jews for the sake of future Germans, for he accepted a mistaken view according to which Jews were not persons, and their interests should be attached no weight. But for those whose descriptive and moral views are less mistaken, increasing the willingness to sacrifice one's own interests may result in a stronger disposition to act morally.

3.2 Violence and Aggression

The opposite of promoting another's interests is damaging another's interests. Since harming others is often immoral, traits which increase harmful behaviour tend thereby to increase immoral behaviour. The reduction in these tendencies would thus often qualify as a moral enhancement. An obvious example is the treatment of psychopathy. But more common are personality disorders.2 Especially dangerous amongst these is Anti-Social Personality Disorder (75% of prison inmates have this) and Borderline Personality Disorder.

Personality disorder (PD) affects 5-10% of the population, placing heavy demands on psychiatric, social and forensic services (NIMH 2003): 64% of male and 50% of female offenders have PD (NOMS 2011). Traits associated with many personality disorders include criminal behaviour, addiction, self-harm, violence, selfishness, recklessness, impulsivity, lack of empathy and remorse, poor anger management, and willingness to exploit others. PD has an inherently moral component: traits are moral failings that harm self and others (Charland 2004; Pickard 2009; 2011).

Alongside genetic predisposition (Lang and Vernon 2001), the strongest predictor of PD is early-environment psychosocial adversity. PD is associated with parental psychopathology, institutional care, sexual, emotional, and physical abuse (Paris

[2] Thanks to Hannah Pickard for contributing details on Personality Disorder.

2001). The chaotic/violent behaviour and emotional instability diagnostic of PD mirrors early environment. Patients with PD did not have the opportunity to learn moral skills.

There is increasing evidence that PD can be treated pharmacologically and psychologically. Antidepressants are recommended for depressive symptoms and impulsivity (NIMH 2003); sedatives for short-term crises (NICE 2009). There are specific psychological therapies: cognitive-behavioural therapy (Davison 2008), dialectical behavioural therapy (Linehan and Dimeff 2001) and STEPPS (Blum et al. 2008); mentalization-based therapy (Fonagy et al. 2004); and Therapeutic Communities (Lees et al. 1999). These develop theory of mind skills and self-control as well as promoting personal and social responsibility (Pickard 2011). Psychiatric interventions are acting as moral enhancers (Pearce and Pickard 2009).

In addition to treating PD, it may be possible to reduce aggression by modifying more subtle psychological factors. Baron-Cohen notes that empathy can act as 'brake on aggression' (2003, 35). Thus, we should expect that a lesser male capacity for empathy could go with the greater display of male aggression, which is borne out by the statistics of crimes like murder (see e.g. Baron-Cohen 2003, 36). If women do have a lower tendency to harm others overall, it seems that in principle we could make men more moral by biomedical methods by making them more like women, or rather, more like the men who are more like women in respect of empathy and aggression.3

3.4 Racial and Sexual Bias

It would be fairly uncontroversial that freedom from certain biases—such as racial and sexual bias—is conducive to acting morally in many contexts. Such biases are frequently impediments both to right motivation and to right action. Though there is disagreement about precisely what motives are the right motive action, it would be widely accepted that these motives should be free from racial and sexual bias. Biased reasoning and conative states driven by biases are *not* the right motives for action. Similarly, while there is considerable disagreement about right action, it would be widely accepted that sexually and racially discriminatory conduct is often wrong.

It is tempting to think that racism and sexism are, at least in Western societies, largely a thing of the past. But the evidence suggests not. Though racial bias is notoriously difficult to measure, most research suggests that, though it has declined since 1960, it remains present. Regression analyses typically find that Black US men earn less than their White counterparts even after correction for alternative explanatory factors such as educational attainment and age (Darity, Guilkey & Winfrey 1996; Rodgers & Spriggs 1996; Gottschalk 1997). Darity and Mason (1998, p. 71) estimate that in 1980 and 1990 black men in the United States were paid 12-15% less than white men as a result of racial discrimination. Additionally, Black males with darker skin appear to fare worse in the labour market than Black males with lighter skin, again, after correction for other explanatory variables (Ransford 1970; Keith & Herring 1991; Johnson, Bienenstock & Stoloff 1995). Further direct evidence of bias comes from

court proceedings (successful suits for racial discrimination remain frequent) and audits, in which pairs of actors who differ in race but are trained to perform equally well at interview apply for the same position with matched curricula vitae. A series of such audits in the United States found that black male actors were three times more likely to be turned down for a job than white male actors (Fix, Galster & Struyk 1993, pp. 79-81). Similar evidence is available for sexual bias (Neumark, Bank & Van Nort 1996). In one interesting study, Goldin and Rouse (2000) found that where symphony orchestras move from auditioning candidates in the view of auditioners to 'blind' auditions, the average likelihood of women being selected increases by fifty percent.

Tendencies to favour the members of a particular sex or racial group may not always qualify as biases, since in some cases the preference may be morally permissible. For example, most of us would find it permissible for a person to favour members of a particular sex or race when considering who to invite on a date, or who to accept for the roles of Othello or Desdemona in a play. But in many cases, racial and sexual preferences do amount to biases.

There is an emerging understanding of the biological bases of racial bias. Several neuroimaging studies have suggested that activation of the amygdala – part of the brain that has been implicated in emotion – may underpin this bias. Lieberman and collaborators found that both Black and White subjects exhibited greater amygdala activation on functional magnetic resonance imaging when presented with photos of Black Americans as compared to photos of White Americans (Lieberman et al 2005). Other studies have identified a correlation between these amygdala responses and implicit racial attitudes revealed by psychological tests (Phelps et al 2000; Amodio 2008). These findings are consistent with the view that negative implicit evaluative reactions to certain racial groups are mediated by differences in amygdala activity, plausibly due to the amygdala's role in emotion. There is some further support for this hypothesis. For example, the only persons known to lack a consistent tendency to discriminate on the basis of race are the victims of Williams syndrome, a rare chromosomal abnormality associated with reduced fear in social situations (Santos et al 2010). This suggests a possible role for the emotion of fear in mediating racial bias.

Though this research is very far from yielding biological interventions capable of selectively and reliably attenuating racial bias, one might expect that further scientific progress will ultimately lead to the development of such techniques. At least, it would be bold to rule this out. These techniques might well constitute moral enhancements, at least in some individuals and circumstances.

3.4 Other Morally Relevant Traits

There are other traits which are conducive to acting morally in many circumstances. Willingness to co-operate with other people is one. As we have seen, SSRIs increase willingness to co-operate (though they may have other undesirable moral effects). Another trait is impulse control. If one cannot withstand temptation and delay gratification, one will be less likely to sacrifice one's own interests for the interests of others or a moral code. Drugs which increase impulse control thus contribute to more moral behaviour. Ritalin, Adderall and other drugs improve impulse control in children with attention deficit disorder, indeed reducing violence and antisocial behaviour.

Of course, modification of these traits could be done for nefarious purposes making someone, for example, a more effective criminal. Moreover, even when done with good intentions, attempts at moral bioenhancement might misfire. It is easy to imagine circumstances in which an isolated enhancement of any single one of the traits we have suggested would produce moral disenhancement rather than enhancement—recall the case of Himmler above—though this risk might be mitigated by enhancing a combination of traits. Similarly, in some people, enhancing one or more of the traits we have discussed might fail to produce moral enhancement because that individual already possesses the trait(s) in question to an optimal (or supra-optimal) degree: it is possible to be too self-sacrificing from the point of morality, so enhancing those the willingness to self-sacrifice of those who already possess this trait to a high degree might lead to moral deterioration. Our point is merely that, in many people, enhancing one or more of the traits we have discussed would, in many circumstances, result in that individual being more likely to act morally than would otherwise have been the case.

4. MORAL BIOENHANCEMENT AND FREEDOM

We now turn to consider an objection to moral bioenhancement raised recently by John Harris. Harris traces the objection to Milton.

> Famously, in Book III of Paradise Lost Milton reports God saying to his "Only begotten Son" that if man is perverted by the "false guile" of Satan he has only himself to blame:
>
> ……whose fault?
> Whose but his own? Ingrate, he had of me
> All he could have; I made him just and right,
> Sufficient to have stood, though free to fall.4
>
> (Harris, 2011).

It is not immediately clear that there is anything in this passage that should concern a proponent of moral bioenhancement. If we read the claim that humans are "sufficient to have stood" as implying that there is no *need* for moral enhancement – that humans already have sufficiently moral motives and behaviour – then it will look clearly false. It will also look inconsistent with Harris' own admission that we often succumb to temptation,[5] and often have purposes other than, and in conflict with, moral goals.[6] If, on the other hand, it is understood as holding merely that humans have the *capacity* for sufficiently moral motives and behaviour (henceforth, the capacity to be moral), then it seems quite consistent with the thought that moral bioenhancement would be

[4] Ibid Line 96ff.

[5] Harris, op. cit. n. 10, pp. 103-104.

[6] Ibid: 104.

morally permissible and indeed desirable. We could undergo moral bioenhancements that further enhanced this capacity, or that disposed us to exercise it more effectively.

Why, then, does Harris appeal to Milton? Harris explains:

> God was, of course, speaking of the fall from Grace, when congratulating herself on making man "sufficient to have stood though free to fall", she was underlining the sort of existential freedom … which allows us the exhilaration and joy of choosing (and changing at will) our own path through life. And while we are free to allow others to do this for us and to be tempted and to fall, or be bullied, persuaded or cajoled into falling, we have the wherewithal to stand if we choose. So that when Milton has God say mankind "had of me all he could have", he is pointing out that while his God could have made falling impossible for us, even God could not have done so and left us free. Autonomy surely requires not only the possibility of falling but the freedom to choose to fall, and that same autonomy gives us self-sufficiency; "sufficient to have stood though free to fall. (Harris 2011)

And then,

> part of Milton's insight is the crucial role of personal liberty and autonomy: that sufficiency to stand is worthless, literally morally bankrupt, without freedom to fall. . . . [M]y own view is that I, like so many others, would not wish to sacrifice freedom for survival. I might of course lack the courage to make that choice when and if the time comes. I hope however that I would, and I believe, on grounds that have more eloquently been so often stated by lovers of freedom throughout history, that freedom is certainly as precious, perhaps more precious than life.[7]

Moral enhancement thus, according to Harris, is wrong because it restricts the freedom to do wrong and thereby reduces personal autonomy.

Harris' argument has a theological parallel in the free will defence of theism. The argument from evil holds that there can be no omnipotent, omniscient and benevolent God, since that God would not allow evildoing to occur. The free will defence maintains, in reply, that evildoing is a consequence of our possessing the freedom to do evil, which is, all things considered, good. Though the freedom to do evil possesses the great instrumental disvalue of allowing evildoing, it also possesses some other, greater value. Thus, God rightly bestowed on us the freedom to do evil despite the risk of wrongdoing that this created.

Similarly, Harris would say, we would be right to retain our freedom to act immorally by declining moral bioenhancement, despite the risk of wrongdoing that this entails. His argument implies both that moral enhancement would deprive us of our freedom

[7] Harris, op. cit. n. 10, pp. 110-111.

to act immorally, and that the disvalue of this loss of freedom would exceed any value associated with a reduction in the rate of wrongdoing. In what follows we respond to these claims.

4.1 Must Moral Bioenhancement Restrict the Freedom to Act Immorally?

Moral bioenhancement, and biological behaviour manipulation more generally, need not restrict freedom. It could simply make us more like the most morally virtuous individuals already among us. To see this, suppose that women are, in at least some circumstances, more disposed to act morally than men because their greater empathy leads them to make greater personal sacrifices in certain circumstances where self-sacrifice is what requires. Then men might be morally enhanced by being made more like women in their capacity for empathy. Plainly, this would not make the men less free to do wrong. For women are not less free to act immorally than men. A certainly they are not barred from acting wrongly by their greater empathy.

This result holds regardless of whether human action is determined. Suppose, first, that our behaviour is fully determined but that our freedom is compatible with it being fully determined whether or not we shall do what we take to be good. In this case, effective moral bioenhancement will not reduce our freedom; it will simply bring about circumstances where we are more often, or always, determined to do what we take to be good. The actions would be those of someone today who is morally perfect.

However, if we are free only because, by nature, we are not fully determined to do what we take to be good, then moral bioenhancement can never be fully effective because its effectiveness is limited by our indeterministic freedom. So, irrespective of whether determinism or indeterminism reigns in the realm of human action, moral bioenhancement will not curtail our freedom.

Some critics of moral bioenhancement seem to think that we risk becomingautomatons who do not act for reasons. Harris writes that moral bioenhancement will ‘make the freedom to do immoral things impossible, rather than simply making the doing of them wrong and giving us moral, legal and prudential reasons to refrain’ (2011). However, the morally bioenhanced could still act for the same reasons as un –enhanced humans who act morally. The sense in which it is ‘impossible’ for morally bioenhanced people to do what they regard as immoral will be the same as it is already for the virtuous person: it is psychologically or motivationally out of the question. People who are morally good and always try to do what they regard as right are not necessarily less free than those who sometimes fail to do so.

To take a final parallel, consider someone who reads a good novel. Such a person might be brought to vividly imagine what it is like to be another person to a much finer and deeper degree. As a result, he empathises with that character and develops sympathy for him. Such a moral enhancement does not rob freedom. If anything it facilitates richer imagination of what life is like and its alternatives. If a pill were to do the same thing, it would be no different, regarding its effects on freedom, to a novel. If a pill were to make people more open to the experiences and lives of others, this would no less erode freedom than reading Tolstoy. Consider the following example.

4.2 Beggar in the street

Sarah is a lawyer at a London firm. She is asked to take on an extra hour per week on pro bono work, but prefers to spend the time relaxing with friends in a bar,

Sarah takes a drug which makes her more interested in the suffering of others, more empathetic, more capable of vividly imagining what it would be like to be in another person's shoes. The drug is like a pair of moral spectacles, clarifying her vision ofother people's experiences. She sees pro bono clients not as extra one hour of her time spent working, but as people caught in a complex legal system, confused, stigmatised and lacking the funds to pay for justice. She sees how their lives will go with her expert helpl and how they will go without it. She decides to give up her time for the clients.

In this case, Sarah retains the same deliberation and judgement. Sarah acts for reasons, in the same way that anyone does. She has simply viewed the original circumstances in a different way. Sarah's giving of her time was not unfree, it was virtuous. Imagine that Sarah, when she took the drug, always behaved in the morally correct way. She would not be unfree. She would be the most virtuous person.

Consider now James. James is a district court judge in a multi-ethnic area. He was brought up in a racist environment and is aware that emotional responses introduced during his childhood still have a biasing influence on his moral and legal thinking. For example, they make him more inclined to counsel jurors in a way that suggests a guilty verdict, or to recommend harsher sentencing, when the defendant is African-American. James recognises this, and dislikes it. A drug is available that would help to reduce his aversion African-Americans, thus mitigating his bias. It would help him to do the right thing. And, since it would remove one inappropriate motive—racial aversion—it would help him to do it for the right reasons.

Taking this drug might plausibly be said to increase rather than decrease, James' freedom. For the aversive reaction that it attenuates might itself be thought to be a constraint on his freedom. Suppose we draw a distinction between the true or authentic self, and the brute self. An agent acts freely, let us say, when her true self determines what she does. If James' racial aversion is part of his brute self—which seems plausible—then the drug helps to free his true self from brute constraints. It helps to enhance his freedom.

It may be that as our understanding develops, that moral bioenhancement will be most effective in children during their early development,. Perhaps by giving them drugs or other biological manipulation we will be able to increase their ability to more easily learn to behave morally, just as cognitive enhancement may enable them one day to learn to study more easily and effectively. Moral bioenhancement will of course rest a conventional moral education: children would still need to be taught correct values, and the importance of acting on values, et cetera, just as cognitive enhancers do not work without education and study. But the mora bioenhancement may allow the education we routinely give our children to be more effective.

For example, most parents would aim to encourage children to recognise suffering in other people, and to respond to try to ameliorate it. By engineering this biologically, some would argue that this would restrict the child's open future. But we do this all the time through education, stories, literature, and punishment and we do not believe it to restrict the freedom to act immorally. Why should it make a difference if we should we do this biologically?

It might be objected that we have painted an unwarrantedly rosy picture of the likely nature of moral bioenhancement. Such enhancements might be unlikely to take the form of drugs that enhance empathy or imagination, reduce unwanted racial aversion, or aid childhood moral development. More likely, it might be thought, they would simply remove the option of acting immorally. For example, neurofeedback might be used to condition irresistible disgust reactions at the thought of harming others. Such interventions surely would restrict the freedom to act immorally. But would they thereby diminish our *autonomy*? And if so, would this render them morally unjustified, all things considered?

4.4 Perfect Mind Control: Phone Hacking

It is 2100. The mobile phone evolved into a brain-computer interface allowing a wide range of communications under direct mind control. One could communicate just by thinking and directing one's thoughts to a target person or artificial intelligence. The iPhone 10EEE was so useful and successful that every parent implanted one into the earlobe of their children, like an ear ring. People without the iPhone 10EEE came to be seen as disabled because they could not communicate sufficiently. They were the deaf and blind of their generation. Governments soon implanted these cheap devices into all newborns as a way of enabling their lives and securing their human rights.

Technology progresses relentlessly and exponentially in other directions. It becomes possible to evaluate and intervene in human intention by hacking into the iPhone 10EEE communications network. A small government spin-off perfects, MT, or moral technology. They can pick up intentions to perform grossly immoral actions and intervene to change these. Traditional government realises the potential and implements MT.

MT is only designed to prevent gross immorality. It only intervenes in human action to prevent great harm, injustice or other deeply immoral behaviour from occurring. For example, murder of innocent people no longer occurs. As soon as a person forms the intention to murder, and it becomes inevitable that this person would, without intervention, act to kill, MT would intervene. The would-be murderer would 'change his mind'. MT does not intervene in trivial immoral acts, like minor instances of lying or cheating. Only when a threshold insult to some sentient being's interests is crossed is MT deployed.

Humans are still free to act morally, since if they chose to do so, MT does not intervene. They are only unfree to do grossly immoral acts, like killing or raping. This was seen as preferable to physical incarceration, which physically restricted the freedom to be immoral. It was seen as preferable that would-be murderers changed their minds, than that an innocent person was killed and then the murderer incarcerated for life. A would-be murderer never knows that her intentions have been

changed by an authority outside of herself. It seems to her that she has "changed her mind" – she experiences a life of complete freedom, though she has not been free. And no one is ever wrongfully killed.

There had been quite a bit of controversy over what should be classified as "grossly immoral action" which should be within the purview of MT. Should cheating in exams be extinguished? Marital infidelity? The cognitively and morally enhanced government decides that only those acts which would have resulted in imprisonment of a person should be modified. Thus prisons are abolished.

It is this kind of world which objectors to moral enhancement like Harris fear. Human beings are no longer 'free to fall' or at least not free to fall a long way. It might be wondered what is so bad with such a world after all? It is true that people are in one way less free in the world with MT. But plausibly everyone is much better off for the absence of evil. There is no physical incarceration or great harm wrought by one human being on another.

We return to the question whether restricting the freedom to do wrong might be morally justified, all things considered, in section 2.6 below. However, for the moment, we focus on the more limited question whether it would reduce *autonomy*. Plausibly, the reason for caring about the freedom to do wrong is that we have have reason to protect our autonomy—roughly speaking, our control over our lives. If a restriction on our freedom to do wrong would thereby restrict our autonomy, it might be of moral concern. But if it would not, it is not clear that it should trouble us.

Though MT does compromise the freedom to do wrong, there are at least two circumstances in which it might be thought not to compromise autonomy or self-government. One of these circumstances is where the immoral action prevented by MT would have been the result of an inauthentic or irrational desire. The other is where a person had voluntarily chosen to be connected to MT as a form of precommitment contract. The case in which autonomy could most plausibly be said to have been preserved is where both of these circumstances obtain: where an individual forms a precommitment contract to prevent himself from acting on an irrational (or inauthentic) desire. The paradigm example of such a case is Ulysses and the Sirens.

2.4 Ulysses and Sirens

The story of Ulysses and the Sirens provides an example of what can be called an obstructive desire. Ulysses was to pass "the Island of the Sirens, whose beautiful voices enchanted all who sailed near. [They] ... had girls' faces but birds' feet and feathers ... [and] sat and sang in a meadows among the heaped bones of sailors they had drawn to their death", so irresistible was their song. Ulysses desired to hear this unusual song, but at the same time wanted to avoid the usual fate of sailors who succumbed to this desire. So he plugged his men's ears with bees' wax and instructed them to bind him to the mast of his ship. He told them: "if I beg you to release me, you must tighten and add to my bonds." As he passed the island, "the Sirens sang so sweetly, promising him foreknowledge of all future happenings on earth." Ulysses

shouted to his men to release him. However, his men obeyed his previous orders and only lashed him tighter. They passed safely (Graves 1960).8

Before sailing to the Island of the Sirens, Ulysses made a considered evaluation of what was best for him. Thinking clearly, with all the facts before him, he formed a plan which would enable him to both hear the song of the Sirens and live. His order that he should remain shackled was an expression of his autonomy.

Moreover, his order prevented him from acting on an irrational desire. In the grip of the Sirens' song, Ulysses' strongest desire was that his men release him. At the time, this may have been his only desire. But it was an irrational desire. The song of the Sirens was irresistible. It is plausible to assume that this song so preoccupied those who heard it that they could think of nothing else. It consumed the listener's attention and so prevented vivid imagination of other alternatives. The desire to move closer to the Sirens was irrational because it was not the result of vivid imagination of all alternatives and because it prevented the satisfaction of a rational desire (to hear the Sirens' song and stay alive). There are thus two reasons why Ulysses' precommitment did not restrict his autonomy: because it was itself the result of an autonomous choice, and because the desires it frustrates were irrational and thus, plausibly, impediments to autonomy.9

Where an individual voluntarily connects to MT to prevent acting on an irrational desire, MT achieves the same thing as wax and lashings in Ulysses' case.

[Have moved an altered version of this to later.]

2.5 Rationalist Autonomy

We have suggested that MT may not compromise autonomy where it is a precommitment contract or where the desire to do wrong is, or is based on, an irrational or inauthentic desire. But thus far we have defended this suggestion only by offering a case design to pump intuitions. In this section, we seek to give some theoretical backing to our suggestion. To do this, it is necessary to say something more about the nature of autonomy.

Autonomy is not mere choice by a competent agent. The word, "autonomy", comes from the Greek: autos (self) and nomos (rule or law).i Autonomy is self-government or self-determination. Being autonomous involves freely and actively making one's own evaluative choices about how one's life should go. But autonomous choice is not merely intentional. What distinguishes autonomous choice from mere choice is that it is evaluative, employing a person's normative capacities. It is based on full (or at least

---

8 All quotations in this paragraph are from this work.

9 The notion that some desires can frustrate the expression of our autonomy is also described by Young (*Op. Cit.*, especially pp. 9, 14, 50, 56), Frankfurt (*Op. Cit.*, especially pp. 68-71) and Watson (*Op. Cit.*, especially pp. 109-110, 117). The last two writers use the term freedom rather than autonomy. Feinberg gives a detailed list of the kinds of states which can interfere with autonomy (1973)

as full as possible) appreciation of the nature of the options on offer, and so requires both information and rational deliberation to form rational beliefs. One of us has elsewhere called this a rational choice and defended a rationalist account of autonomy. It could be called a Kantian account.

> P rationally desires some state of affairs, that q, if and only if P desires that q while in possession of all relevant, available information, without making relevant, errors of logic and while vividly imagining what each alternative course of action and resulting state of affairs would be like.

Arguably, one necessary condition for a choice to be autonomous is that it be based on or satisfy a rational desire If choosing to do immoral things would be based on an irrational desire, then preventing a person from so acting would not constrain the persons autonomy. . [cite rational desires and limitation papers]

The paradigm of a person P autonomously choosing between A and B is that, having appreciated the nature of A and B, this P judges that one is better than the other. Why is appreciation of the nature of A and B important? It will not do when imagining what A is like, to imagine some state of affairs which is more like B, or some other state of affairs. If this person were to choose A under these circumstances, what P would really want is B, or something else entirely.

To appreciate A and B as they are, this person must know what each is like. She needs relevant, available information. For example, in considering whether to go on some diet, a person needs to know what all its effects, will be, on weight, health, her financial and temporal assets.

In processing information, it is important not to make any errors of logic. Logic enables us to transfer information and belief into the widest web of rational belief. Steve Jobs, the Apple pioneer who died recently of pancreatic cancer was initially diagnosed with a neuroendocrine cancer and it is rumoured that was possibly treatable. However, Jobs allegedly elected complementary/alternative medicine and delayed definitive surgical treatment for 9 months, by which time the cancer had metastasized.10 Assuming these rumours to be true for the purposes of argument, if Jobs had had access to accurate information, his choice may have been explained by a failure of logic, resulting in irrational beliefs. Suppose that a person is provided with information and reasons in the following way.

(1) There is a chance surgery will have serious adverse effects. (true)
(2) Alternative medicine has no risk of serious adverse effects. (true)
Therefore, if I want to best survive my cancer, then I should have alternative medicine. (false)[think it's the argument that's invalid, not the conclusion]

All that can be concluded from (1) and (2) is that if I want the treatment with the least side effects, then I should choose alternative medicine. These premises say nothing about the effectiveness of surgery or alternative therapy.

---

[10] http://www.dailymail.co.uk/news/article-2049019/Steve-Jobs-dead-Apple-CEO-shunned-conventional-cancer-medicine.html

Logic is important so that a person can utilize available facts properly. False beliefs which arise from correctable errors of logic corrupt a person's appreciation of the nature of the options, and so reduce the autonomy of his choice.

Importantly, autonomous choice also requires "vivid imagination" of the alternatives if one is to appreciate fully their nature as options. Here is an argument to that effect. The concept of choice entails that at least two alternatives are available. But it is necessary to distinguish between subjectively and objectively available alternatives. Two objective alternatives may exist with only one subjective alternative.

Consider the following example, after Locke.ii A person in a room is led to believe that the room is locked, when in fact one door is open. This person has two objective alternatives (leave or stay) but only one subjective alternative (stay).
It is only after a person has presented herself with subjective alternatives that she can choose the one which she judges is best. One's choice cannot be fully self-determining if one believes that the path one sets upon is the only path available. As far as demonstrating that a choice is autonomous, it is not enough to show that objective choices exist. There must be some evidence that subjective choice exists.iii In order to be self-determining, then, it is necessary to present at least two alternatives to oneself. However, being autonomous requires more than this. Imagine that P wants to do A. P believes that he could also do B. However, it is A that P wants to do, and P does not think about B. In one sense, it can be said that P has chosen to do A, but is doing A an expression of P's self-determination? Self-determination is an active process of actually determining the path of one's life. In order to judge what is best for himself, P must think and imagine what it would be like for her if A and B obtained, and what the consequences, at least in the short term, of each of these would be for her. Thus, not only must P know what A and B are like, but she must also imagine what A and B would be like for her. This is vivid imagination.

One's rational objects need not be egoistic. One can autonomously care about morality and moral ends. But to be autonomous, these must have been rationally evaluated. It is true that irrationality, spontaneity and impulsivity can be a part of an autonomous life, but only if rationally endorsed at some time, at some higher level. A completely irrational unreflective life is not an autonomous life, even if it is whole heartedly endorsed, on this Kantian account.

What compromises autonomy? On a full blooded rationalist account, many things will compromise autonomy: relevant false beliefs, invalid or incomplete logic, lack of vivid imagination. The choice to do wrong may often be irrational for one or more of these reasons. Where this the case, it would not, on the rationalist account, compromise the agent's autonomy. It may be preferable to correct false or irrational belief, errors of logic, and facilitate imagination. But this may not be possible or practicable. In these cases, employment of MT, even in cases of competent adults who have not consented to its use, may not offend autonomy and may actually open the door to a more autonomous life preventing incarceration and so promoting a wider range of options.

2.6 The Value of the Freedom to Fall

The objector to moral enhancement by means of MT might respond, "It is fine to allow voluntarily use of MT to prevent oneself from acting in immoral and non-autonomous ways. But it is not fine to coercively people to adopt MT. Indeed, this is problematic even where the behaviour that MT prevents would be non-autonomous. For the act of forcing MT on someone against their will or without their consent is itself an infringement of autonomy. We should not restrict other's autonomy, even where doing so will in the long run increase their autonomy. Thus, for example, it would be wrong to use MT on a child who cannot consent to it, or to use it on a competent adult who refuses consent. Moreover, MT would almost certainly have to be used in this way. Criminals would be unlikely to voluntarily request MT so that, in order to eradicate crime, it would need to be involuntarily employed"

But would it be wrong to infringe the autonomy of a child or adult to prevent grossly immoral action? Autonomy is only one value. We would be negligent if we did not physically restrain a child we knew to be about to commit murder. We work very hard to develop moral education for children, aimed at shaping theirdesires. MT would only remove the most harmful desires, leaving the child free to develop without the taint of serious murder or other serious harm that would be a burden for the rest of his life . And without imprisonment.

The criminal justice system also restrains autonomy in many, varied and serious ways, largely in order to prevent grossly immoral action. And these restraints on autonomy are widely thought to be justified.

It might be objected that there is a difference between restricting someone's autonomy externally (say through incarceration) and restricting someone's autonomy internally (as with MT). Incarceration impedes autonomy by limiting our freedom of movement. But MT directly changes our bodily (brain) and mental states. Arguably, we have a stronger claim to bodily and mental non-interference than we do to freedom of movement. Thus, MT might seem to be a more serious restriction on autonomy than incarceration.

However, even if the distinction between internal and external restrictions has moral significance, it is not clear that this objection can be sustained, since incarceration and other putatively justified restrictions on autonomy designed to prevent gross immorality *do* in fact cause internal changes. A person's brain and mind are not unaffected by being imprisoned for 20 years, say.

Moreover, even if MT *is* a more serious restriction on autonomy than widely accepted restraints currently imposed by the criminal justice system, it may still be justified. For those existing restraits have proven rather poor at preventing gross immorality. Crime remains prevalent. MT, as we described it, would be much more effective at preventing gross immorality than is, say, incarceration. So it might be justified even if it constitutes a more serious restriction on autonomy.

In the world of MT, serious crime would be non-existent. There would be great benefits to society in general, but MT would also be of great benefit to potential criminals: they would no longer risk losing their freedom by imprisonment or capital punishment by committing serious crime. In the absence of effective moral

enhancements that do not restrict autonomy, the loss of freedom in one domain of our lives, to commit evil deeds, would surely be worth the benefits. We would be otherwise free. Even in those cases in which MT does undermine autonomy, the value of human well-being and respect for the most basic rights of others plausibly outweighs the value of autonomy.

Of course, there might be other objections to MT. It could be argued, for example, that MT would be too prone to misuse, or that it's acceptance would be the beginning of a slippery slope to unjustified restrictions on mental autonomy. Our point is that the very use of MT would not always constitute an unjustified restriction on autonomy. In some cases its use would not restrict autonomy at all. And in others, it would restrict autonomy, but justifiably.

5. CONCLUSION

Moral bioenhancement (or its opposite) may already be occurring in small ways when drugs like SSRIs are taken for psychiatric indications. However, there has been no strategic programme to use knowledge from the science of morality to deliberately and effectively improve moral motivation and behaviour through biological means. But such enhancement seems possible and, in many ways, desirable.

In this chapter, we have addressed the objection that moral bioenhancement is wrong because it would compromise the freedom to act immorally and thereby undermine personal autonomy. We have argued that moral bioenhancement need not restrict freedom, and that where they do they may, as with Ulysses, nevertheless preserve personal autonomy. Moral bioenhancements which work by enhancing capacities for empathy or imagination, by removing xenophobic aversions, or by aiding childhood moral development do not diminish autonomy. They may enhance it.

There might be instances of biological intervention to produce moral action that do control the moral agent, subjugating that person to the will of another thus diminishing her autonomy. But such interventions might nevertheless be justified by the benefits that they produce, as are certain existing crime prevention measures. Even if they do reduce freedom and autonomy, the value of behavioural control may outweigh this loss.

Note that, although we have focused on the case of *biological* interventions to enhance moral motivation and behaviour, our arguments also have implications for other varieties of moral enhancement—varieties that are perhaps more likely to be widely adopted. The cognitive sciences have already given rise to social-institutional reforms capable of altering human moral behaviour, for example, nudge techniques. One objection that has sometimes been raised to the use of these techniques is that they restrict autonomy. Though a full discussion of these objections lies beyond the scope of this article, we believe that our discussion of moral bioenhancement suggests reasons to doubt that the objections will count decisively against using nudge techniques, at least when the aim is to prevent gross immorality.

Amodio, D. M., 'The Social Neuroscience of Intergroup Relations', European Review of Social Psychology 19, no. 1 (2008): 1-54.
Arpaly, Nomy (2003). Unprincipled Virtue: An Inquiry Into Moral Agency (New York: Oxford University Press).
Baron-Cohen, Simon 2003. The Essential Difference: Male and Female Brains and the Truth about Autism, New York: Basic Books.
Baumeister, R. F., Bratslavsky, E., Muraven, M. & Tice, D.M. 1998. Ego-depletion: Is the active self a limited resource? Journal of Personality and Social Psychology, 74: 1252-1265.
Baumeister, R. F. 2002. Ego Depletion and Self-Control Failure: An Energy Model of the Self's Executive Function. Self and Identity, 1: 129-136.
Blum N, St. John D, Bruce P, Stuart S, McCormick B. Allen J, Arndt S, and Black DW 2008. "Systems Training for Emotional Predictability and Problem Solving (STEPPS) for Outpatients With Borderline Personality Disorder: A Randomized Controlled Trial and 1-Year Follow-Up," American Journal of Psychiatry, 165: 468 – 478.
Boettiger, Charlotte A., Jennifer M. Mitchell, Venessa C. Tavares, Margaret Robertson, Geoff Joslyn, Mark D'Esposito, and Howard L. Fields 2007. "Immediate Reward Bias in Humans: Fronto-Parietal Networks and a Role for the Catechol-O-Methyltransferase 158Val/Val Genotype," The Journal of Neuroscience, 27(52):14383-14391.
Bovens, L., 2008. "The ethics of nudge." In Till Grüne-Yanoff and S.O. Hansson (eds) Preference Change: Approaches from Philosophy, Economics and Psychology, Berlin and New York: Springer, Theory and Decision Library A, Chapter 10.
Brasil-Neto, J.P., Pascual-Leone, A., Valls-Sole, J., Cohen L.G.. and Hallett, M 1992. "Focal transcranial magnetic stimulation and response bias in a forced-choice task," Journal of Neurology, Neurosurgery, and Psychiatry, 55: 964-966.
Caria A, Sitaram R, Veit R, Begliomini C, and Birbaumer N 2010. "Volitional control of anterior insula activity modulates the response to aversive stimuli: A real-time functional magnetic resonance imaging study," Biological Psychiatry, 68; 5: 425–432.
Carter A, Hall W, Nutt D 2009. "The treatment of addiction," in Addiction Neurobiology: Ethical and Social Implications, Carter A, Capps B, Hall W eds. Office for Official Publications of the European Communities.
Casebeer, W. And Churchland, P. S. 2003. "The neural mechanisms of moral cognition," Biology and Philosophy 18: 169-194.
Charkrabortty A 2008. "From Obama to Cameron:Why do so many politicians want a piece of Richard Thaler?" The Guardian. 8 July 2008.
Charland L 2004. Moral treatment and the personality disorders," in The philosophy of psychiatry: a companion Radden J ed. 64-77. Oxford: OUP.
Churchland, P.S. 2011. Braintrust: What Neuroscience Tells Us about Morality. Princeton: Princeton University Press.
Cohen Kadosh R, Soskic S, Luculano T, Kanai R, Walsh V 2010. "Modulating neuronal activity produces specific and long-lasting changes in numerical competence," Current Biology, 20: 2016-20.
Crockett M. J., Clark L, Tabibnia G, Lieberman MD, Robbins TW 2008. "Serotonin modulates behavioral reactions to unfairness," Science, 320:1739.

Crockett MJ, Clark L, Hauser MD, Robbins TW 2010. "Serotonin selectively influences moral judgment and behavior through effects on harm aversion." Proceedings of the National Academy of Sciences, 107(40): 17433-8.
Darity Jr, W. A., D. K. Guilkey and W. Winfrey, 'Explaining Differences in Economic Performance Among Racial and Ethnic Groups in the USA: The Data Examined', American Journal of Economics and Sociology 55, no. 4 (1996): 411–25.
Darity, W. A., and P. L. Mason, 'Evidence on Discrimination in Employment: Codes of Color, Codes of Gender', Journal of Economic Perspectives 12, no. 2 (1998): 63–90.
Davison K 2008. Cognitive therapy for personality disorder, 2nd edition. London: Routledge.
de Dreu, Carsten et al, 2010. "Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflicts among Humans," Science, 328: 1408-11.
De Dreu CKW, Greer LL, Van Kleef GA, Shalvi S, Handgraaf MJJ 2011. "Oxytocin promotes human ethnocentrism," Proceedings of the National Academy of Sciences, 108 (4): 1262 -6
de Waal, Frans, 2010. The Age of Empathy, London: Souvenir Press
Declerck, C. H., Boone, C., Kiyonari, T. 2010. "Oxytocin and cooperation under conditions of uncertainty: The modulating role of incentives and social information," Hormones and Behavior 57(3): 368-374.
Douglas, T. 2008. "Moral enhancement," Journal of Applied Philosophy, 25(3): 228-45.
Feinberg, J. Social Philosophy, Englewood Cliffs (N.J.): Prentice-Hall, 1973, p. 13.)
Fix, M., G. C. Galster and R. J. Struyk, 'An Overview of Auditing for Discrimination', in Clear and Convincing Evidence: Measurement of Discrimination in America, ed. M. Fix and R. J. Struyk (Washington, DC: Urban Institute Press, 1993), 1–68.
Fonagy P, Gergely G, Jurist EL, Target M 2004. Affect regulation, mentalization, and the development of the self. London: Karnac.
Gazzaniga, M. S 2005. The Ethical Brain. Dana Press, New York.
Goldin, C., and C. Rouse, 'Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians', American Economic Review 90, no. 4 (2000): 715–41.
Gottschalk, P., 'Inequality, Income Growth, and Mobility: The Basic Facts', Journal of Economic Perspectives 11, no. 2 (1997): 21-40.
Graves, R. The Greek Myths, Volume 2, London: Penguin, 1960, 361.
Greene, J. D. 2008. "The secret joke of Kant's soul," in W. Sinnott-Armstrong ed., Moral Psychology: Volume III – The Neuroscience of Morality, 35-79. MIT Press.
Harris, John 2011. "Moral Enhancement and Freedom," Bioethics, 25: 102–111.
Hume, D., *A Treatise of Human Nature*, Second Edition, ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1978).
Insel TR, Fernald RD 2004. "How the brain processes social information: searching for the social brain," Annual review of neuroscience, 27697.
Johnson, J. H., E. J. Bienenstock and J. A. Stoloff, 'An Empirical Test of the Cultural Capital Hypothesis', Review of Black Political Economy 23, no. 4 (1995): 7–27.
Kant, I. (1964) Groundwork of the Metaphysic of Morals (New York: Harper & Row).
Keith, V. M., and C. Herring, 'Skin Tone and Stratification in the Black Community', American Journal of Sociology 97, no. 3 (1991): 760–78.

Kersten, F. 1956. The Kersten Memoirs, 1940-1945, trans. C. Fitzgibbon and J. Oliver, introduced by H. R. Trevor-Roper. Hutchinson, London.
Kiesel, A., Wagener, A., Kunde, W., Hoffmann, J., Fallgatter. A.J. & Stöcker, C 2006. "Unconscious manipulation of free choice in humans," Consciousness and Cognition, 15: 397-408.
Kosfeld M., Heinrichs M., Zak P. J., Fischbacher U. & Fehr E 2005. "Oxytocin increases trust in humans," Nature, 435(7042): 673-6.
Lang KL, Vernon PA 2001. "Genetics," in Handbook of personality disorders,
Lieberman, M. D., A. Hariri, J. M. Jarcho et al., 'An fMRI Investigation of Race-Related Amygdala Activity in African-American and Caucasian-American Individuals', Nature Neuroscience 8, no. 6 (2005): 720-2.
Livesley WJ ed. 231-241. New York: Guildford Press.
Lees J, Manning N, Rawlings B. 1999. Therapeutic community Effectiveness: A Systematic International Review of Therapeutic Community Treatment for People with Personality Disorders and Mentally Disordered Offenders, York: York Publishing.
Liberzon; E A Young 1997. "Effects of stress and glucocorticoids on CNS oxytocin receptor binding," Psychoneuroendocrinology, 22(6): 411-22.
Linehan M and Dimeff L 2001. "Dialectical Behavioural Therapy in a Nutshell,".The California Psychologist 34:10-13.
Link H, Dayanithi G, Gratzl M 1993. "Glucocorticoids rapidly inhibit oxytocin-stimulated adrenocorticotropin release from rat anterior pituitary cells, without modifying intracellular calcium transients," Endocrinology, 132:873–877.
Mill, John Stuart 1859. On Liberty. Oxford University, 21–22. Retrieved 2008-02-27.
J. S. Mill (1979) *Utilitarianism* (Indianapolis, Hackett), pp. 27-28
Morgan, D., Grant, K. A., Gage, H. D., Mach, R. H., Kaplan, J. R., Prioleau, O., Nader, S. H., Buchheimer, N., Ehrenkaufer, R. L. & Nader, M.A 2002. "Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration," Nature Neuroscience, 5: 169-74.
National Institute of Mental Health in England (NIMH(E)) 2003. Personality disorder: no longer a diagnosis of exclusion, London: NIMH(E).
National Offender Management Strategy (NOMS) 2011. Working with personality disordered offenders: a practitioner's guide, London: NOMS.
National Institute of Clinical Excellence (NICE) 2009. Borderline personality disorder: treatment and management, London: NICE.
Neumark, D., R. J. Bank and K. D. Van Nort, 'Sex Discrimination in Restaurant Hiring: An Audit Study', Quarterly Journal of Economics 111, no. 3 (1996): 915–41.
Hessel Oosterbeek & Randolph Sloof & Gijs van de Kuilen, 2004. "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis," Experimental Economics, 7(2): 171-188.
Paris, J. 2001. Psychosocial adversity. In Handbook of personality disorders. Livesley WJ ed. 231-241. New York: Guildford Press.
Pearce S, Pickard H. 2009. The moral content of psychiatric treatment. British Journal of Psychiatry 195:281-282.
Persson, Ingmar and Savulescu, Julian Forthcoming. Fit for the Future? Modern Technology, Liberal Democracy and the Need for Moral Enhancement,", Oxford: Oxford University Press.
Persson, Ingmar and Savulescu, Julian 2011. "Unfit for the Future? Human Nature, Scientific Progress and the Need for Moral Enhancement," in Savulescu, Julian, Ter

Meulen, Ruud, and Kahane, Guy., (eds.). Enhancing Human Capacities. Oxford: Wiley-Blackwell.
Persson, I. and Savulescu, J. (2010) 'Moral Transhumanism'. Journal of Medicine and Philosophy, thematic issue on Transhumanism and Bioethics. Published online November 2010 (forthcoming in print 2011).
Phelps, E. A., K. J. O'Connor, W. A. Cunningham et al., 'Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation', Journal of Cognitive Neuroscience 12, no. 5 (2000): 729-38.
Pickard H. 2009. Mental illness is indeed a myth. In Psychiatry as cognitive neuroscience. Broome MR, Bortolotti L eds., 83-101. Oxford: OUP.
Pickard H. 2011 Forthcoming. Responsibility without blame: empathy and the effective treatment of personality disorder. Philosophy, Psychiatry, Psychology. 18;3.
Ramachandran VS and Oberman LM. 2006. Broken mirrors: a theory of autism. Scientific American. 295:62-9.
Dominance and Affiliative Behaviour. Psychopharmacology, 161, 324-330.
Ransford, H. E., 'Skin Color, Life Chances, and Anti-White Attitudes', Social Problems 18, no. 2 (1970): 164–79.
Rodgers, W. M., and W. E Spriggs, 'What Does the AFQT Really Measure? Race, Wages, Schooling and the AFQT Score', Review of Black Political Economy 24, no. 4 (1996): 13–46.
Santos, A., A. Meyer-Lindenberg and C. Deruelle, 'Absence of Racial, but Not Gender, Stereotyping in Williams Syndrome Children', Current Biology 20, no. 7 (2010): R307-R308.
Savulescu, Julian and Persson, Ingmar 2008. "The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity," Journal of Applied Philosophy, 25(3): 162 - 167.
Savulescu, Julian and Persson, Ingmar 2011. "The Turn for Ultimate Harm: A Reply to Fenton," Journal of Medical Ethics published online February 2011 (forthcoming in print March 2011) 10.1136/jme.2010.036962
Savulescu, J. and Persson, I. (forthcoming 2011) 'Getting Moral Enhancement Right: The Desirability of Moral Enhancement'. Bioethics published online 29 July 2011.
Silber, M., Almkvist, O., Larsson, B., Stock, S. & Uvna$s-Moberg, K. 1987. "The effect of oral contraceptive pills on levels of oxytocins in plasma and on cognitive functions," Contraception, 36:641-650.
Singer, Peter 2005. "Ethics and intuitions," Journal of Ethics. 9: 331–52.
Sitaram R, Caria A, Veit R, Gaber T, Rota G, Kuebler A, Birbaumer N 2007. "fMRI brain-computer interface: a tool for neuroscientific research and treatment," Computational Intelligence and Neuroscience, Article ID 2548.
Sitaram R, Caria A, Birbaumer N 2009. "Hemodynamic brain-computer interfaces for communication and rehabilitation,". Neural Networks, 22;9:1320-1328.
Stock S, Karlsson R, von Schoultz B. 1994. "Serum profiles of oxytocin during oral contraceptive treatment," Gynecological Endocrinology, 8(2): 121-6.
Sunstein, Cass 2005. "Moral heuristics," Behavioral and Brain Sciences, 28: 531-542.
Terbeck S, Kahane G, McTavish S, Savulescu J, Cowen P, Hewstone M. Under review a. "Beta-Adrenergic Blockade Reduces Implicit Negative Racial Bias"
Terbeck S, Kahane G, McTavish S, Savulescu J, Cowen P, Hewstone M. Under review b. "Emotion in moral decisionmaking: Beta adrenergic blockade increases deontological moral judgments".
Thaler RH, Sunstein Cass 2008. "Nudge: Improving Decisions about Health," Health,and Happiness. Yale: YUP.

Tse, W.S. & Bond, A.J. 2002. “Serotonergic intervention affects both Social Dominance and Affiliative Behaviour,” Psychopharmacology,161: 324–330.
Wallace B, Cesarini D, Lichtenstein P, Johannesson M. “Heritability of ultimatum game responder behaviour,” Proceedings of the National Academy of Sciences,104(40):1 5631-4.
Wang, B., Shaham, Y., Zitzman, D., Azari, S., Wise, R. A. &You Z. B. 2005. “Cocaine experience establishes control of midbrain glutamate and dopamine by corticotropin-releasing factor: a role in stress-induced relapse to drug seeking,” Journal of Neuroscience, 25: 5389-96.
Wood R M, Rilling J K, Sanfey A G, Bhagwagar Z, Rogers RD 2006. “Effects of tryptophan depletion on the performance of an iterated Prisoner's Dilemma game in healthy adults,” Neuropsychopharmacology 31 (5): 1075–84.
Westen, D. 2007. The Political Brain. PublicAffairs books.
Zak P, Kurzban R, Matzner W 2004. “The Neurobiology of Trust,” Annals of the New York Academy of Sciences, 1032:224-227.

---

[i] Dworkin G. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press, 1988: 12.

[ii] In Locke's case, the person believes the door is open when in fact it is locked. (Locke, J. *An Essay concerning Human Understanding* (ed. Pringle-Pattinson, A.S.) Clarendon Press, Oxford, 1924, Book II, Chapter xxi, sec. 10.)

[iii] Kahneman and Varey note: "A basic tenet of psychological analysis is that the contents of subjective experience are coded and interpreted representations of objects and events. An objective description of stimuli is not adequate to predict experience because coding and interpretation can cause identical physical stimuli to be treated as different and different ones to be treated as identical..." (Kahneman, D. and Varey, C. "Notes on the psychology of utility," in Elster, J. and Roemer, J.E. *Interpersonal Comparisons of Well-Being*, Cambridge University Press, Cambridge, 1991, p. 141.)