

Moral Enhancement and Moral Freedom: A Critique of the Little Alex Problem

By John Danaher

[Forthcoming in Hauskeller, M. and Coyne, L (eds) *Moral Enhancement: Critical Perspectives*, Philosophy Supplement (Cambridge University Press)]

Abstract: A common objection to moral enhancement is that it would undermine our moral freedom and that this is a bad thing because moral freedom is a great good. Michael Hauskeller has defended this view on a couple of occasions using an arresting thought experiment called the 'Little Alex' problem. In this paper, I reconstruct the argument Hauskeller derives from this thought experiment and subject it to critical scrutiny. I claim that the argument ultimately fails because (a) it assumes that moral freedom is an intrinsic good when, in fact, it is more likely to be an axiological catalyst; and (b) there are reasons to think that moral enhancement does not undermine moral freedom.

1. Introduction

Humanity faces a series of existential challenges over the next century: climate change, malevolent artificial intelligence, the proliferation of nuclear and bio-weaponry, to name but a few. According to Ingmar Persson and Julian Savulescu, one of the reasons why we find ourselves on the existential precipice is because our technological capacities far exceed our moral ones.¹ They argue that moral bioenhancement technologies might be needed to redress the balance and pull us back from the brink.

¹ Persson, I. and Savulescu, J. *Unfit for the Future: The Need for Moral Enhancement* (Oxford: Oxford University Press 2012)

Their argument has attracted a lot of criticism. One of the most persistent is that using drugs and other biomedical interventions to improve moral behavior will undermine our ‘freedom to fall’, i.e. our freedom to do good or evil (what I call, from here on out, our ‘moral freedom’). John Harris was the first to launch this criticism² and he and Persson and Savulescu have slogged it out on the topic ever since.³ Others have occasionally entered the fray (e.g. Sparrow 2014) to pass judgment and offer new insights.⁴

One of the more interesting contributions to the ‘freedom to fall’-debate has come from Michael Hauskeller.⁵ Using an arresting thought experiment derived from Anthony Burgess’s novel *A Clockwork Orange*, Hauskeller draws our attention to the axiological intuitions underlying the ‘freedom to fall’ objection, and makes a passionate plea for the view that a world with moral freedom and the occasional bad deed is better than a world in which people do good things but lack the freedom to fall.

In this chapter, I carefully reconstruct and evaluate Hauskeller’s arguments. I start by discussing his ‘Little Alex’ thought experiment and formalizing the argument that he derives from it. I then proceed to critically

² Harris, J. ‘Moral Enhancement and Freedom’ *Bioethics* 25 (2011): 102-111.

³ Harris, *How to Be Good: The Possibility of Moral Enhancement* (Oxford: Oxford University Press, 2016) and Harris, J. ‘Moral Blindness – The Gift of the God Machine’ 9 (2016): 269-273; Persson, I. and Savulescu, J. ‘Enharrishment: A Reply to John Harris about Moral Enhancement’ *Neuroethics* 9 (2016): 275-77; and

Persson, I. and Savulescu, J. ‘Moral Bioenhancement, Freedom and Reason’ *Bioethics* 9 (2016): 263-268.

⁴ Sparrow, R. ‘Better Living Through Chemistry? A Reply to Savulescu and Persson on ‘Moral Enhancement’” *Journal of Applied Philosophy* 31 (2014): 23-32.

⁵ Hauskeller, M. ‘The “Little Alex” Problem’ *The Philosophers’ Magazine*, 62 (2013): 74-78; and Hauskeller, M. ‘Is it desirable to be able to do the undesirable? Moral Bioenhancement and the Little Alex Problem’ *Cambridge Quarterly of Healthcare Ethics*, forthcoming – available at: https://www.academia.edu/18809315/Is_It_Desirable_to_Be_Able_to_Do_the_Undesirable_Moral_Bioenhancement_and_the_Little_Alex_Problem

evaluate the key premises of that argument. I first evaluate the moral intuition that Hauskeller uses to motivate his claim, namely: that a world with good outcomes but no moral freedom would be worse than a world with moral freedom and the occasional bad deed. Drawing upon lessons learned in the debate about the existence of God and the problem of evil, I argue that moral freedom lacks intrinsic value: its mere presence doesn't make the world better or worse. Instead, moral freedom is an axiological catalyst: something that makes good deeds better and bad deeds worse. I then turn to the question of whether moral enhancement would in fact undermine our moral freedom. Looking at common theories of free will and responsibility, I argue that moral enhancement may not undermine our moral freedom and could even on some occasions increase our moral freedom. Then I look at the more political dimension to freedom, which Hauskeller draws upon in his argument, and suggest that moral enhancement need not undermine political freedom and may, in fact, increase it.

Before I get underway, I need to say a word or two about the terminology I use in this chapter. As will become clear later on, one of the major disputes in the moral enhancement debate is about what exactly counts as a 'moral' enhancement. On one interpretation, a *moral* enhancement would involve improving an individual's moral virtue and moral reasoning. It would involve creating "people who *are* good and *do* what is right, for the *right* reasons" .⁶ On other interpretations, moral enhancement is largely about securing preferred moral outcomes, irrespective of the reasoning or virtues of the people implicated. In other words, it involves ensuring enhanced *conformity* with moral norms, not necessarily enhanced moral virtue and reasoning. In this chapter, I will favour the latter interpretation of moral enhancement. It is more inclusive and fits better with the argument Hauskeller defends.

⁶ De Melo Martin, I. and Salles, A. 'Moral Bioenhancement: Much Ado About Nothing' *Bioethics* 29 (2014): 223-232, 224.

On top of this, there is always some uncertainty in the enhancement debate as to what interventions count as ‘enhancements’. On a broad interpretation, an enhancement is any intervention that improves the human condition relative to its pre-existing state.⁷ On this view, the invention of the wheel, literacy, political reform, and better nutrition would all count as enhancements. On a narrow interpretation, an enhancement is a biomedical or technological intervention that directly targets and tries to improve some function of the human brain or body. On this view, drugs, brain implants, and other therapeutic interventions would count as enhancements. Proponents of enhancement frequently stress the broader interpretation⁸ in order to highlight the continuity between new and old forms of enhancement, and to reduce opposition to newer interventions. Opponents typically favour a narrower interpretation, trying to draw principled distinctions between narrow and broad enhancements. In this chapter, I will favour the narrower interpretation. The primary reason for this is that if one is to understand the debate about the freedom to fall objection one must be willing, if only for the sake of argument, to draw a distinction between biomedical forms of moral enhancement and more traditional forms such as moral education.

Finally, I will use the term ‘moral freedom’ in two distinct senses. For the majority of the article, I will use it to refer to our capacity for free will and moral responsibility. In other words, I will assume that in order to have moral freedom we must have the capacity to exercise our free will (whatever that requires) and be held morally responsible for what we do (whatever that requires). In the penultimate section, however, I will adopt a more politicized sense of the term ‘moral freedom’, which focuses on liberal and republican conceptions of freedom.⁹ In other words, I will hold that we are morally free if we are free from

⁷ Buchanan, A. *Beyond Humanity? The Ethics of Biomedical Enhancement*. (Oxford: Oxford University Press 2011).

⁸ See for example Harris, J. *Enhancing Evolution* (Princeton, NJ: Princeton University Press, 2007) and Buchanan, *Beyond Humanity*.

⁹ List, C. and Vallentini, L. ‘Freedom as Independence’ *Ethics* 126 (2016): 1043: 1074

interference and/or domination by others. There is some overlap between these conceptions of freedom, but they are distinguishable in certain instances.

2. Understanding the Little Alex Problem

Hauskeller uses a thought experiment to introduce his version of the freedom to fall objection. The thought experiment comes from Anthony Burgess's (in)famous novel *A Clockwork Orange*. The novel is set in an unspecified, dystopian future. It tells us the story of 'Little' Alex, a young man prone to exuberant acts of ultraviolence. Captured by the authorities, Alex undergoes a form of aversion therapy in an effort to rid him of this tendency towards ultraviolence (the therapy is known as 'Ludovico's Technique' in the novel). He is given medication that makes him feel nauseous and then repeatedly exposed to violent imagery. His eyes are held open in order to force him to view the imagery. The therapy works. Once he leaves captivity, he still feels violent urges but these are quickly accompanied by feelings of nausea. As a result, he no longer acts out in violent ways. The therapy has enhanced his moral conformity.

The novel takes an ambivalent attitude towards this conformity (and eventually Alex relapses into his ultraviolent exuberance after a suicide attempt). One of the characters (a prison chaplain) suggests that Alex is not truly good as a result of the therapy. In order to be truly good, Alex would have to *choose* to do the good. But due to the aversion therapy, this choice is taken away from him. The induced nausea every time he has a violent thought effectively *compels* him to do the good. Indeed, the chaplain goes further and suggests that Alex's induced goodness is not as valuable as his natural badness. It is better if a person can choose to do the bad than be forced to do the good. This is what Hauskeller calls the 'Little Alex' problem. And he describes it like this:

This is what I call the "Little Alex" problem... it invites us to share a certain moral intuition (namely that it is in some unspecified way bad or wrong or inhuman to force people into goodness) and thus to accept the ensuing

paradox that under certain conditions the bad is better than the good — because it is not only suggested that it is wrong to force people to be good (which is fairly uncontroversial) but also that the resulting goodness is somehow tainted and devaluated by the way it has been produced.¹⁰

This description of the problem hints at an argument, one that can be expressed in more formal terms. It starts with a premise stating the core moral intuition and uses this to critique the practice of moral enhancement. This is how I would reconstruct that argument:

- (1) It is better to have moral freedom, i.e. the freedom to do the bad (and to occasionally act on that freedom), than to be forced to do the good.
- (2) Moral enhancement takes away our moral freedom.
- (3) Therefore, moral enhancement is, in some sense, a morally inferior way of ensuring moral conformity.

This formulation is a little bit loose (the derivation of the conclusion from the premises is not straightforward or watertight), but I think it captures the gist of Hauskeller's interpretation of the freedom to fall objection. Over the remainder of the chapter I will evaluate the two premises of this argument.

3. Is Moral Freedom Intrinsically Valuable?

The first premise of the argument is the most interesting. It makes a seemingly paradoxical and contentious axiological claim. It states that the

¹⁰ Hauskeller, The 'Little Alex' Problem, 75.

freedom to do bad is such an important good that a world without it is worse than a world with it. To be more precise, it states that on some occasions, and under certain conditions, we should prefer it when people do bad things than when they do good things.

One of the more important features of Hauskeller's contribution to the freedom to fall debate is the way in which he draws attention to this axiological claim. Now, I happen to think that this axiological claim is false and I'm going to defend my view by first showing that in order to accept premise (1) you probably need to believe moral freedom is an intrinsic good (i.e. that its mere presence adds value to the world), and then by arguing that it is implausible to suppose that moral freedom is an intrinsic good. Instead, I argue that we should view it as an *axiological catalyst*, i.e. something that adds to both the moral value and disvalue of the world and hence something whose value cannot be assessed independently from the way in which it is used.

Why think that the argument presupposes that moral freedom is an intrinsic good? Hauskeller has a particular conception of the value hierarchy of different possible worlds that he uses to explain his view. In his original article on the topic, Hauskeller suggests that any proponent of the 'freedom to fall'-argument must accept something like the following value hierarchy as between different possible worlds:

Best World: A world in which we are free to do bad but choose to do good (i.e. a world in which there is both moral conformity and moral freedom)

2nd Best World: A world in which we are free to do bad and (sometimes) choose to do bad (i.e. a world in which there is moral freedom but not, necessarily, moral conformity)

3rd Best World: A world in which we always do good but are not free to do bad (i.e. a world in which there is moral conformity but no moral freedom)

Worst World: A world in which we are not free and do bad (i.e. a world in which there is neither moral conformity nor moral freedom).

In his more recent paper, he proposes a similar but more complex hierarchy featuring six different levels (the two extra levels capture differences between 'sometimes' and 'always' doing good/bad). In that paper he notes that although the proponent of the 'freedom to fall' argument must place a world in which there is moral freedom and some bad above a world in which there is no moral freedom, there is no compelling, watertight argument in favour of this hierarchy of value. It is really a matter of moral intuitions and weighing competing values. Hauskeller's intuitions lead him to favour the world with the freedom to (sometimes) do bad over the world of moral conformity.

Can we understand this intuition at a deeper level? It seems that there might be much to learn here from the debate between atheists and theists over the problem of evil. As is well-known, the problem of evil is the most famous atheological argument. It comes in a variety of forms. These are usually broken down into two main families: (i) the logical problem of evil and (ii) the evidential problem of evil.¹¹ The essence of both versions is that the existence of God is incompatible with the existence of any gratuitous or unnecessary evil. The reasoning is as follows: God is a maximally powerful, morally perfect being. Given his moral perfection, he would not allow for evil to occur unless it was somehow necessary for a greater good. Given his maximal power, it is possible for him to intervene to prevent unnecessary evil from occurring. Atheists then claim that there is evidence (certain or highly probable, depending on how strong they want the argument to be) that gratuitous evils occur. They conclude, therefore, that God must not (or is highly unlikely to) exist.

¹¹ On the different problems of evil, see Mackie, J.L. (1955). Evil and Omnipotence. *Mind* 64 (1955): 200-212; Plantinga, A. *God, Freedom and Evil*. (Grand Rapids, MI: Eerdmans, 1977); and Rowe, W. 'The Problem of Evil and Some Varieties of Evil' *American Philosophical Association* 16 (1979): 335-341.

Theists have a variety of responses. They often highlight our epistemic ignorance about value on a grand, cosmic scale.¹² They then try to construct 'theodicies' or 'defences' that either justify how God might allow a seemingly gratuitous evil act to occur, or introduce prima facie justifications that cannot be easily second-guessed or challenged, given our epistemic limitations. The most popular of these theodicies/defences are those that focus on free will and moral responsibility (i.e. moral freedom). The view shared by proponents of these theodicies is that free will and moral responsibility are great goods and their being great goods is what justifies God in creating a world with some evil in it. According to one of the most influential formulations of this view, it is not logically possible for God to create a universe with the great good of moral freedom without also allowing for the possibility of that freedom sometimes (possibly many times) being used for ill effect.¹³ The ingenuity of this is that it allows for theists to remain committed to the view that they live in a universe created by a morally perfect being – i.e. in a universe that is, ultimately, the best of all possible worlds – while tolerating evil deeds. In essence then, they are ascribing to Hauskeller's intuition that a world with moral freedom in it is, somehow, better than a world without.

I would suggest that the most plausible way to make sense of this shared intuition is to believe that moral freedom is, somehow, intrinsically valuable, i.e. that its mere presence (irrespective of how it is used) adds to the value of the universe. But why is this the most plausible way to make sense of the intuition? Couldn't it simply be the case that moral freedom is necessary for some other goods? Perhaps, but I think it is difficult for Hauskeller or a theist proponent of the free will theodicy to sustain this view alongside a consistent bias towards worlds with moral freedom. Some theists have tended to view free will as a

¹² For example Bergmann, M. 'Skeptical Theism and Rowe's New Evidential Argument from Evil' *Nous* 35 (2001): 278-296.

¹³ Plantinga, *God, Freedom and Evil*.

necessary instrumental good for moral responsibility,¹⁴ but this does not get us away from the notion that moral freedom is an intrinsic good since responsibility is part and parcel of moral freedom. Others tend to be a bit cagey and simply suggest that, for all we know, it might be linked to other goods. But this is a difficult argument to make when you do not know what these other goods are and yet still believe they trump other known goods. If you are going to consistently rank a world in which there is moral freedom (and some evil) over a world in which there is no moral freedom (and much good) – in other words, if you are going to accept that there is some good in a world in which people exercise their moral freedom to do tremendous evil – then it seems like you simply must be committed to the view that there is some intrinsic moral magic to moral freedom: that when it is present it automatically ups the value to the world. To be clear, this does not mean that advocates of moral freedom need to believe that the intrinsic good of moral freedom trumps all other goods, or that there could be no other outweighing intrinsic goods. They might still believe that. But in order to maintain their consistent bias toward worlds in which there is moral freedom, it is most plausible to believe that it is because moral freedom has intrinsic value.

The problem I have with this is that moral freedom doesn't seem to function like an intrinsic good. There are a couple of ways to get to this conclusion. Some have tried to do so by arguing that it is intrinsically neutral. Derk Pereboom¹⁵ has defended this view in tandem with his general free will skepticism, and Steve Maitzen has defended it specifically in response to free will theodicies.¹⁶ The essence of both of their claims is that in many cases of moral evaluation, the freedom to do bad does not add to the assessed value of an action. It seems to function, instead, as a *morally weightless* consideration, and not just a

¹⁴ Swinburne, R. *The Existence of God* 2nd Edition. (Oxford: Oxford University Press, 2004).

¹⁵ Pereboom, D., Free Will, Evil and Divine Providence. In Dole and Chignell (eds). *God and the Ethics of Belief* (New York: Cambridge University Press., 2005)

¹⁶ Maitzen, S. 'Ordinary Morality Implies Atheism' *European Journal for Philosophy of Religion* 2 (2009): 107-126.

morally outweighed one. To give an example, suppose there is a violent criminal who has exercised his moral freedom and savagely murdered ten people. Surely, we don't think that the fact that he murdered them freely speaks in his favour? His act is very bad, pure and simple; it is not slightly good and very bad.

There is something to this argument. It definitely feels right to say that the murderer's freedom doesn't add any intrinsic good to his actions (it doesn't counterbalance some of the disvalue). But it also feels wrong to say that it is axiologically neutral. Compare two cases: (i) a murderer savagely kills ten people while exercising his moral freedom; and (ii) a murderer savagely kills ten people without exercising moral freedom (because he had some brain tumour that undermined his moral reasoning, or he was compelled to do so by others at gunpoint). Both outcomes are bad, but which case is worse? I think the answer is obviously (i) because there is no excusing factor present in that case. This suggests something interesting. It suggests that moral freedom makes a bad deed much worse. Thus, moral freedom seems to be making an axiological difference; it is not neutral. We can see this happening in the other direction too. Compare the following: (iii) a multi-millionaire donates 90% of his money to some charity by exercising his moral freedom; and (iv) a multi-millionaire donates 90% of his money to some charity because his accountant made a calculation error. The outcomes are both good (assume this, for the sake of argument), but I am confident in saying that case (iii) is clearly better than (iv). Again, the presence of moral freedom seems to be making the difference.

These thought experiments suggest to me that moral freedom is neither intrinsically good, nor intrinsically neutral. It is, instead, an *axiological catalyst*: it makes good things better and bad things worse. It exaggerates whatever value is already present (positive or negative). This, in turn, leads to the conclusion that premise (1) of the Little Alex Argument is flawed. A world with moral freedom and occasional bad is not *necessarily* better than a world without moral freedom. Because it is an axiological catalyst, we cannot use moral freedom to rank and order worlds relative to one another. The value of moral freedom can only be determined in connection with how it is exercised. If it is exercised to good

effect, then a world with it is, indeed, a better world; but if it is exercised to ill effect, it makes things worse. This significantly undercuts the 'freedom to fall' objection. If moral enhancement techniques work as advertised, they would simply be preventing moral freedom from being exercised in wicked ways. They would, consequently, be making the world a better place, without taking away something that is intrinsically good.

This argument might not be plausible to all. Fortunately, even if you cling to the view that moral freedom is an intrinsic good, it does not follow that you ought to reject all forms of moral enhancement on the grounds that they undermine moral freedom. As mentioned earlier, moral freedom could be an intrinsic good and it could be outweighed or countermanded by other goods/bads in certain cases. This is, again, something that has come to the fore in the debate about the problem of evil. When theists press the line about moral freedom providing some God-justifying excuse for the occurrence of evil, atheists push back by pointing to particularly egregious forms of evil that result from moral freedom. Take the case of a remorseless serial killer who tortures and rapes young innocent children. Are we to suppose that their freedom to do bad outweighs the child's right to live a torture and rape-free life? Is the world in which the serial killer freely does bad really a better world than the one in which he is forced to conform? It seems pretty unlikely. This example highlights the fact that moral freedom might be intrinsically valuable but that in certain 'high stakes' cases its intrinsic value is outweighed by other moral considerations, particularly the need to ensure greater moral conformity.

It is perfectly open to the defender of moral enhancement to argue that its application should be limited to those 'high stakes' cases. Then it will all depend on how high the stakes are and whether moral enhancement can be applied selectively to address those high stakes cases. The thing that is noticeable about Persson and Savulescu's case for moral bioenhancement is that it is linked directly to high stakes cases. They argue that the future of the human race could be on the line unless we enhance our moral conformity. If they are right, then the stakes might be high enough to outweigh the putative intrinsic value of moral

freedom. That said, the burden they face when making this case is significant since they are not simply arguing that we occasionally interfere with the moral freedom of specific individuals; they are arguing that we interfere with every agent's moral freedom.

4. Is Moral Enhancement Really Incompatible with Moral Freedom?

What about the second premise of the Little Alex Argument? This premise claims that moral freedom is incompatible with moral enhancement, i.e. that if we ensure someone's conformity through a technological intervention like moral enhancement, then they are not really free. How persuasive is this? It all depends on what you understand by moral freedom and how you think moral enhancement works.

Let's stick with idea that moral freedom is the capacity to exercise free will and to be a responsible moral agent. There are many different theories about what this requires. They vary depending on whether they think free will and responsibility are possible in a causally deterministic universe, and whether they think we live in such a universe. We can divide them into two main categories: (i) libertarian theories of free will (which hold that free will is only possible if humans are somehow exempt from causally deterministic laws) and (ii) compatibilist theories of free will (which hold that free will is possible even if humans are not exempt from causally deterministic laws). There is also a position known as 'free will skepticism', but we can safely ignore that here since a free will skeptic will have no truck with arguments about the good of moral freedom. Now, it is not going to be possible to review every theory of free will within the two main categories in the space of this chapter, but we can review some of them, and by doing so we can see that moral enhancement may be less likely to undermine moral freedom than you might first suspect.

Let us start with the libertarian accounts of free will. These can be fleshed out in a couple of different ways. One can adopt an *agent causalist* approach, which holds that the way in which *agents* cause something to happen is distinct

from the way in which *events* cause other events to happen. Agents are primary, *sui generis* causes of events in the real world. They do not sit directly within the ordinary chain of causation. Human beings are agents and so can cause things to happen without being causally determined to do so. This view obviously faces a number of conceptual and philosophical challenges. It is pretty difficult to reconcile the fact that certain aspects and features of human agency are clearly constituted in an event-causalist network of brain activity with the claim that human agency causes events in a distinctive, *sui generis* manner. Consequently, it is difficult to know exactly what an agent causalist would say about the Little Alex case. They might be inclined to share Hauskeller's view that Alex is being forced to be free (perhaps because the aversion therapy blocks the pathway to agent causation), but they might be more optimistic. After all, if all forms of moral enhancement will operate on the causal networks inside the human brain, it is possible that the 'agent' (who is separable from those networks) is unaffected by the enhancement process.

A similar analysis applies to event-causalist libertarian views, such as those advocated by Kane¹⁷ and Balaguer.¹⁸ These views do not posit a *sui generis* agent who sits outside the ordinary causal flow of the universe. They accept that we are (to a considerable extent) constituted by our brains and that our brains usually follow deterministic patterns of causation. They simply argue that there are occasional moments of indeterminism (what Balaguer calls 'torn decisions') and that our status as free and responsible agents is tied to these moments of indeterminism. Balaguer's view, for instance, is that on certain occasions, the causal environment in which the human agent operates (which includes their brain and surrounding context) is in a state of equilibrium at the moment prior to the choice – no ordinary causal factor can 'make a difference' to whether one option is chosen over the other. At those moments, the human 'will' operates and it is at these moments that we exercise free choice.

¹⁷ Kane, R. *The Significance of Free Will* (Oxford: Oxford University Press, 1996).

¹⁸ Balaguer, M. *Free Will as an Open Scientific Problem* (Cambridge, MA: MIT Press, 2010)

Again, this view has certain conceptual and philosophical problems, but let us set those to the side and consider what it means for moral enhancement. On the one hand, a proponent of an event-causal libertarianism might remain optimistic about moral freedom. Enhancement technologies might intervene in the causal networks of the brain but they may leave open the possibility of there being 'torn decisions' and our moral freedom could continue to be grounded in those moments. On the other hand, they might fear that enhancement technologies like Little Alex's aversion therapy *reduce the number* of torn decisions the typical human faces. It could be, for all we know, that before his aversion therapy, every time Little Alex was faced with the option of committing an act of ultraviolence he was facing a torn decision: his brain was perfectly equilibrated between committing the act and not. He was exercising his moral freedom on each such occasion. The aversion therapy then changed the causal equilibrium, tipping the balance decisively in favour of not committing violence. This would eliminate free will. Then again, the opposite could also be true. Perhaps before the aversion therapy the causal network within his brain was tipped decisively in favour of committing ultraviolence. The aversion therapy then worked by bringing it back to a more equilibrated state where his will could make the critical difference between doing good and doing evil. In other words, perhaps the enhancement therapy works by opening up more opportunities for torn decisions.

The overall point here is that it is very difficult to say, in the abstract, whether libertarian models of free will are undermined or constrained by moral freedom. Because they hold that the will (or the agent) is somehow separate from the causal mechanisms of the universe, it is possible that they are unaffected by moral enhancement. And, even if they are affected, they might be affected in a positive way. A lot depends on the particularities of the individual case.

Other theories of free will and moral responsibility are compatibilist in nature. They claim that moral freedom is possible within a deterministic causal

order. Deterministic causation by itself does not undermine freedom. What matters is whether an agent's actions are produced by the right kind of causal mechanism. There are many different accounts of compatibilist free will, but some of the leading ones argue that an agent can act freely if the causal mechanism producing their moral choices is *reasons-responsive* and/or produces actions that are consistent with their character and higher order preferences (Fischer and Ravizza 1998; Dworkin 1988; Frankfurt 1972).¹⁹

Moral enhancement *could* undermine compatibilist free will so understood. It may change the causal sequence of action from one that is compatible with moral freedom to one that is not compatible with moral freedom. In the Little Alex case, the aversion therapy causes him to feel nauseous whenever he entertains violent thoughts. This is definitely inconsistent with some versions of compatibilism. From the description, it seems like Alex's character is still a violent one and that he has higher-order preferences for doing bad things, it's just that he is unable to express those aspects of his character thanks to his nausea. He is thus blocked from acting freely according to higher-order preference accounts of freedom. Similarly, it could be that the aversion therapy makes him less responsive to certain kinds of reasons for action. This might undermine a reasons-responsive account of moral freedom.

But, again, the devil will be in the detail. The modality of the enhancement will be all important. Some modalities might be perfectly consistent with compatibilistic moral freedom: they might change an incompatible causal sequence into a compatible one. In this respect, aversion therapy is hardly the only game in town. Some agents might desire the good at a higher-order level

¹⁹ Dworkin, G. *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press 1988)

Fischer, JM and Ravizza, M. *Responsibility and Control* (Cambridge: Cambridge University Press, 1998).

Frankfurt, H. Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1971): 5-20

and be thwarted from pursuing it by lower-order impulses. A moral enhancement technology might work by blocking these lower order preferences and thereby enabling moral freedom. If we believe that chemical castration can count as a type of moral enhancement, then this is arguably how it works: by blocking lower-order preferences for sexual misdeeds. Other modalities of moral enhancement might work by changing an agent's ability to appreciate, process and respond to different reasons for action. This would improve their reasons-responsivity. Although not written with moral enhancement in mind, Maslen, Pugh and Savulescu's paper (2015) on using DBS to treat Anorexia Nervosa highlights some of these possibilities, showing how neurointerventions that are directly mediated through higher-order brain functions might be preferable to those that are not.²⁰

On top of this, those who claim that moral freedom is undermined by enhancement must deal with the case in which an agent freely decides to undergo an enhancement treatment at T1 that will compromise their moral freedom at T2. Most theories of moral freedom accept this as a case involving genuine moral freedom. They view it as a case involving a pre-commitment to being good. To use the classic example, the individual who chooses to undergo the process is like Odysseus tying himself to the mast of his ship: he is limiting his agency at future moments in time through an act of freedom at an earlier moment in time. The modality of enhancement doesn't matter then: all that matters is that he isn't forced into undergoing the enhancement at T1. Hauskeller acknowledges this possibility in both of his papers, but goes on to suggest that they may involve a dubious form of self-enslavement. At this point, the metaphysical account of moral freedom becomes less important. We must turn, instead, to a more political understanding of freedom.

5. Freedom, Domination and Self-Enslavement

²⁰ Maslen, H., Pugh, J. and Savulescu, J. 'The Ethics of Deep Brain Stimulation for Anorexia Nervosa' 8 (2015): 215-230.

Where metaphysical freedom is about our moral agency and responsibility, political freedom is about how others relate to and express their wills over us. It is about protecting us from interference and domination by others so as to meet the conditions for a just and mutually prosperous political community — one that respects the fundamental moral equality of its citizens (Gaus 2010; Pettit 2014).²¹ Consequently, accounts of political freedom are not so much about ensuring free will as they are about ensuring that people can develop and exercise their agency without being manipulated and dominated by others. So, for example, I might argue that I am politically unfree in exercising my vote, if the law requires me to vote for a particular party. In that case, others have chosen for me. Their will dominates my own. I am subordinate to them.

Although underexplored in the enhancement debate,²² this politicized account of freedom might provide a more promising basis for a defence of premise (2). After all, one critical problem with debates about moral enhancement – particularly those that appeal to ‘high stakes’ risks such as those that figure in Persson and Savulescu’s account – is that they implicitly or explicitly endorse cases in which others decide for us whether we should undergo moral enhancement therapy. Thus, our parents could genetically manipulate us to be kinder; our governments may insist on us taking a course of moral enhancement drugs to become safer citizens; it may become a conditional requirement for accessing key legal rights and entitlements that we be morally enhanced, and so on. The morally enhanced person would, consequently, be in a politically different position from the naturally good person. That, at least, is how Hauskeller sees it:

The most conspicuous difference between the naturally good and the morally enhanced is that the latter have been engineered to feel, think, and behave in a certain way. Someone else has decided for them what is evil and what is not,

²¹ Gaus, G. *The Order of Public Reason*. (Cambridge: Cambridge University Press, 2010); and Pettit, P. *Just Freedom*. (New York: WW Norton & Co, 2014)

²² Though see Sparrow, ‘Better living through chemistry?’.

*and has programmed them accordingly, which undermines, as Jurgen Habermas has argued, their ability to see themselves as moral agents, equal to those who decided how they were going to be. The point is not so much that they have lost control over how they feel and think (perhaps we never had such control in the first place), but rather that others have gained control over them. They have changed...from something that has grown and come to be by nature, unpredictably, uncontrolled, and behind, as it were a veil of ignorance, into something that has been deliberately made, even manufactured, that is, a product.*²³

There is a lot going on in this quote. But the gist of it is clear. The problem with moral enhancement is that it creates an asymmetry of power. We are supposed to live together as moral equals: no one individual is supposed to be morally superior to another. But moral enhancement allows one individual or group to shape the moral will of another. But this complaint doesn't quite get at the pre-commitment situation. What if there is no other individual or group making these decisions for you? What if you voluntarily undergo moral enhancement? Hauskeller argues that the same inequality of power argument applies to this case:

*...we can easily extend [this] argument to cases where we voluntarily choose to submit to a moral enhancement procedure whose ultimate purpose is to deprive us of the very possibility to do wrong. The asymmetry would then persist between our present (and future) self and our previous self, which to our present self is another. The event would be similar to the case where someone voluntarily signed a contract that made them a slave for the rest of their lives.*²⁴

What should we make of this argument? It privileges the belief that freedom from the yoke of others is what matters to moral agency — that we should be left to *grow* and *develop* into moral agents through natural processes — not be

²³ Hauskeller, 'The "Little Alex" Problem', 78-79

²⁴ Hauskeller, 'The "Little Alex" Problem', 79

manipulated and *manufactured* into moral saints (even if the manipulation and manufacturing is done by ourselves). But I'm not sure we should be swayed by these claims. Three critical points seem apposite to me.

First, I think we should be generally skeptical of the claim that it is better to be free from the manipulation of others than it is to be free from other sorts of manipulation or interference. The reality is that our moral behaviour is the product of many things: our genetic endowment, our social context, our education, our environment, and various contingent accidents of personal history. It is not obvious to me why we should single out causal influences that originate in other agents for particular ire. In other words, the presumption that it is better that we naturally grow and develop into moral agents seems problematic to me. Our natural development and growth — assuming there is a coherent concept of the 'natural' at play here — is not intrinsically good. It is not something that is necessarily worth saving or necessarily better than the alternatives. At the very least, the benefits of moral conformity would weigh (perhaps heavily) against the desirability of natural growth and development.

Second, I do not think that the claim that induced moral enhancement involves problematic asymmetries of power holds up under scrutiny. At the very least, it is not a claim that is generally or unquestionably true. If anything, I think moral enhancement could be used to correct for asymmetries of power. To some extent this will depend on the modality of enhancement and the benefits it reaps, but the point can be made at an abstract level. Think about it this way: The entire educational system rests upon asymmetries of power, particularly the education of young children. This education often involves a moral component. Do we rail against it because of the asymmetries of power? Not really. Indeed, we often deem education necessary because it ultimately helps to correct for asymmetries of power. It allows children to develop the capacities they need to become the true moral equals of others. If moral enhancement works by enhancing our capacities to appreciate and respond to moral reasons, or by altering our desires to do good, then it might help to build the capacities that correct for asymmetries of power. It might actually enable effective self-control and autonomy. Being

morally enhanced then does not mean that you are problematically enslaved or beholden to the will of others.

Third, and perhaps most controversially, I am not convinced that self-enslavement is a bad thing. Every decision we make in the present enslaves our future selves in at least some minimal sense. Choosing to go to school in one place, rather than another, enslaves the choices your future self can make about what courses to take and career paths to pursue. Is that a bad thing? If the choices ultimately shape our desires — if they result in us really wanting to pursue a particular future course of action — then I'm not sure that I see the problem. Steve Petersen has made this point in relation to the creation of robot slaves.²⁵ He argues that if a robot is designed in such a way that it really really wants to do the ironing, then maybe getting it to do the ironing is not so bad *from the perspective of the robot* (this last bit is important — it might be bad from a societal perspective because of how it affects or expresses our attitudes towards others, but that's not relevant here since we are talking about self-enslavement). Likewise, if by choosing to undergo moral enhancement at one point in time, I turn myself into someone who really really wants to do morally good things at a later moment in time, I'm not convinced that I'm living some inferior life as a result.

6. Conclusion

To sum up, the 'freedom to fall' objection is one of the more popular objections to the moral enhancement project. Hauskeller defends an interesting version of this objection, one that draws our attention to certain key axiological intuitions: that it is better to have moral freedom and do bad than it is to be forced into moral conformity; that it is better to naturally develop and grow into a moral person than it is to be manipulated and molded by others (including yourself). In this chapter, I have called these axiological intuitions into question. I argue that they rest on the implausible belief that moral freedom is an intrinsic

²⁵ Petersen, S. 'Designing People to Serve' in Lin, Abney and Bekey (eds) *Robot Ethics: The Ethical and Social Implications of Robotics*. (Cambridge, MA: MIT Press, 2012)

good and that manipulation/self-enslavement is an intrinsic bad. Neither of these views is correct or sufficient to justify the argument against moral enhancement. Moral freedom is an axiological catalyst, not an intrinsic good; self-enslavement and manipulation are not obviously inferior to other forms of causal determination or external molding. On top of this, it is not even obvious that moral enhancement undermines moral freedom. It all depends on what you mean by moral freedom and the modality of the moral enhancement. All the leading theories allow for the possibility that moral enhancement increases, rather than undermines, moral freedom.