Reply to Commentaries

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.]

In our chapter, we argued that moral bioenhancements could preserve autonomy even where they restrict the freedom to act wrongly. We suggested that this would most plausibly be the case where (I) the agent autonomously chooses to undergo the bioenhancement *and* (II) the bioenhancement operates by attenuating an autonomy-restricting desire. And we argued that even in those cases in which it does undermine autonomy, it could still be morally justified in certain circumstances.

Ibuki and Kodama invoke a Frankfurtian hierarchical account of autonomy to argue that Moral Enhancement Technology or MT could threaten autonomy even in the kinds of cases where we suggested it threaten only freedom. It could do this, they suggest, because it could alter the agent's 'authentic' or 'higher' self, which they associate with the agent's second-order desires—her desires regarding her other desires. Ibuki and Kodama suppose that we were imagining cases in which an agent has (i) a putatively contra-moral first-order desire, such as a desire to inflict harm, and (ii) a second-order desire to maintain this first-order desire. They then imagine two different ways in which MT might operate. First, it might directly attenuate only the contra-moral first-order desire. Second, it might directly attenuate the first-order desire *and* alter the second-order desire to align it with the agent's new first-order

desire. Regarding the first case, Ibuki and Kodama worry that the agent might, as a "defence mechanism", modify his second-order desire so as to bring it into line with his first-order desire. For example, if MT has weakened or eliminated an agent's desire to inflict harm, the agent might, over time, also eliminate his desire to possess this desire. In that case, MT would have indirectly influenced the agent's higher self. In the second case, the worry is more obvious: in this case, MT directly influences the agent's the agent's higher self. Ibuki and Kodama suggest that, because both of these kinds of MT would alter the agent's higher self, they might threaten autonomy. They would do so, for example, if imposed by one agent on another. The first agent would then, they suggest, be "manipulating" the higher self of the other.

An initial problem with this argument is that it is not clear why altering another person's higher self entails restricting that person's autonomy. On minimalist Frankfurtian accounts of autonomy, all that is needed for autonomy is that one's first-and second-order desires are aligned—that is, that one's first-order desires are endorsed by one's second order desires. Though both of the scenarios that Ibuki and Kodama imagine involve (directly or indirectly) altering second-order desires, neither ultimately leaves the agent with unaligned first- and second-order desires, thus neither would threaten autonomy, on the minimalist Frankfurtian view.

Admittedly, there are more sophisticated Frankfurtian accounts which posit more stringent conditions for autonomy. For example, on one view, autonomy requires not only alignment between one's first- and second-order desires but also that this alignment is not the result of inauthentic influences (Dworkin 1981). On this view,

MT might restrict autonomy even if it leaves the agent's first- and second-order desires well-aligned. However, it will do so only if MT itself qualifies as an inauthentic influence and Ibuki and Kodama provide no argument to show that it must.

A second and more serious problem with Ibuki and Kodama's argument is that it does not bear on cases in which our suggested conditions—(I) and (II)—hold. In those cases, the agent autonomously chooses to use MT, so we would not have one agent manipulating the higher self of another. Rather, in these cases, the agent would be autonomously altering *her own* higher self, and this would not threaten her autonomy. On any plausible Frankfurtian account, an agent can autonomously adopt actions which alter her second-order desires without thereby compromising her autonomy. Consider a person whose second order desires change because she chooses to read Anna Karenina. Surely there is no assault on autonomy in this case.

Perhaps Ibuki and Kodama's aim was merely to show that autonomy would be threatened in cases where (II) holds, but not (I)—cases, that is, where MT attenuates or blocks an autonomy-restricting desire, but is not undergone autonomously. This would be enough to establish that condition (II) is not sufficient for the preservation of autonomy, which would be an interesting result (though not one that we disputed).

Unfortunately, however, Ibuki and Kodama do not establish even this. This is because the scenarios they discuss are not ones in which MT is used to block or

attenuate a nonautonomous desire. In their scenarios, the agent has a contra-moral first-order desire and a secord-order desire *to maintain* that contra-moral desire. These cases are, if we accept a Frankfurtian account of autonomy, cases in which nothing is amiss at the outset, at least from the point of view of autonomy; the agent's first- and second-order desires are aligned.[1] In these cases, there is no autonomy-restricting desire for the agent to block or attenuate.

The cases that would most plausibly satisfy condition (II), on a Frankfurtian account, would be cases in which the agent has a contra-moral first order desire and a second-order desire *to be without* that first order desire. MT would could then be used to block or attenuate the first-order desire. Ibuki and Kodama do not consider such cases, and it is not clear why they would be autonomy-restricting, on a Frankfurtian account.

A final difficulty with Ibuki and Kodama's argument is that it presupposes that a Frankfurtian account of autonomy is correct. In fact such accounts are dogged with problems of how second-order desires are autonomously formed and regress problems (Thalberg 1978). There are other accounts of autonomy—such as the rationalist account we outlined—that are arguably more plausible and according to which it is irrelevant whether MT alters the agent's second-order desires.

[1] We assume here that the second-order desire to maintain the contra-moral first-order desire is not the result of inauthentic forces.

Ibuki and Kodama suggest that we are committed to accepting a Frankfurtian account; our "argument requires a certain understanding of 'the self'; that is, the hierarchy of individual desires, where upper-level desires (or reason) control lower-level ones". We do not see why our argument requires this. It does require that some of an agent's desires can be autonomy-restricting, but this is something that can be accommodated by many accounts of autonomy, including the rationalist account that we outline.

Morioka argues we should pursue social rather than biological means to moral enhancement. He points to the dramatic drop in homicide over the last 60 years which he attributes to increased prosperity and gun control laws. He then argues that future threats, such as the use of extremely powerful biotechnology to kill millions (Persson and Savulescu 2008) should also be addressed using social measures:

> "The only way to prevent them would be to strictly control the access to those problematic pharmaceutical substances and establish laws to punish individuals for possession of those drugs. Japan has succeeded in prohibiting the possession of guns among ordinary citizens."

Sadly, gun control models are unlikely to be effective in tackling the existential threats we face. Modified smallpox virus could potentially obliterate the human population and, within a decade or two, 100,000s of people may have the capacity

to create such viruses, thanks to progress in genetic engineering and synthetic biology. It may take only one of these individuals to wreak catastrophic havoc.

The situation is so urgent, Ingmar Persson and I argued in a recent book (Savulescu and Persson 2012), that we must pursue all avenues open to us. We 3 have NEVER argued against social means to moral improvement. Indeed, we have acknowledged social means will typically be the most desirable means. However, it is doubtful whether social means will reduce the risk of catastrophic harm to negligible levels. Thus, once social means have been exhausted, there will remain a case for exploring the possibility of moral bioenhancement (Persson and Savulescu 2008). In fact, we tend towards the view that even once all acceptable social *and biological* means to moral improvement have been pursued, there will remain a significant risk of catastrophic harm due to the malevolent use of new technologies. Gun control is important, but it is a vanishingly small part of the existential challenges we face.

Morioka raises another common objection: exploitation of the morally enhanced.

> "Imagine lifeboat ethics. There are six people on a lifeboat with a capacity for five. One of the six individuals is a morally bioenhanced person. Savulescu argues that self-sacrifice and altruism are the two central characteristics of morality, and that these traits can be enhanced by biological determinants. If Savulescu is right, this morally bioenhanced person in the lifeboat would think that she has to sacrifice herself to save her fellow passengers by her

> plunging into the sea. As a result, the other five greedy people would be saved."

Similarly, he sees people taking oxytocin as enabling others to "effectively dominate them, use them, and finally exploit them as slaves." The morally enhanced would be like indigenous people at the hands of "wild colonists."

And there would be no protection: "The police whose hearts are filled with empathy and generosity would never be able to complete their mission in emergency situations."

Morioka has a very narrow view of what moral enhancement would consist in. We gave altruism and empathy as examples of traits whose augmentation might produce, or be a component in, moral enhancement in some cases. But we explicitly acknowledged that *in some individuals and some circumstances* augmenting these traits would not produce a moral enhancement, and might even result in moral disenhancement. Morioka's examples illustrate this. It is not moral for an altruistic person to give up his life for a bad person. Arguably, it is wrong. It would be an altruistic act to give up his life for three innocent, normal children – and that would surely be laudable. Similarly, it is not moral to always follow the Christian ideal to 'turn the other cheek'. In some cases, aggression is warranted in the face of a grave injustice or wrong.

Moral enhancement is a complex and context-specific process to which many different factors might contribute: moral imagination, empathy, sympathy, altruism, general intelligence, strength of will, sense of justice, willingness to retaliate to moral wrongs, etc. Typical instances of moral enhancement will, we assume, involve altering a number of these traits. There may be some individuals, in some circumstances, who would be morally enhanced by augmenting only one of these traits. (Imagine a person whose only moral defect is a slight lack of empathy.) However, as we have been at pains to emphasize, there is no one of these traits whose augmentation would be sufficient for moral enhancement in all people and all circumstances.

A common strategy in the literature critical of moral bioenhancement has been to single out each of the traits suggested as possibly relevant to moral enhancement by its defenders and show that there are cases in which augmenting this trait would in fact produce moral deterioration (Harris 2012). However, this strategy misses the mark. To our knowledge, no-one who has defended moral bioenhancement has simplistically posited the augmentation of altruism, sympathy or any other single trait as a universal basis for moral enhancement.

Finally, Morioka worries that moral sensitivity would make us unhappy - instead of enjoying our expensive dinner we would think of starving people and it would spoil our enjoyment.

> "The reason why ordinary people can survive every day would be that they are not so morally sensitive as to worry about such "small" things."

People like Peter Singer would argue that morally should worry more about our expensive selfish tastes and our obligations to others. The fact that our immorality makes us unhappy is, if not a good thing, perhaps a necessary thing for moral improvement. As Singer has argued (and shown), the moral life is in fact perfectly compatible with a truly happy and fulfilled life (Singer 1997).

Morality will inevitably require self-sacrifice. So it will at some deeper level cut into well-being. However, as Sidgwick argued (he called it the Dualism of Practical Reason (Sidgwick 1884)), how the reasons of self-interest are to be weighed against those of morality is one of the deepest questions for ethics. One minimal answer to this question is a duty of easy rescue: when the cost to you is small of performing some action, and the benefit to others is great, then you should all things considered perform that action. Buying a £30 of wine instead of a £300 bottle for the sake of helping someone else is surely not too much to ask.

There may be a point beyond which moral enhancement is no longer morally required, because it will result in more self-sacrifice than we are morally required to bring about. There may also be a point beyond which augmenting one's disposition to self-sacrifice for the sake of others would no longer qualify as a moral enhancement, say, because it would leave one permanently sick and thus reliant on

others.

But most of us have scope to morally enhance ourselves a great deal before reaching either of these points.

In any case, concerns about self-sacrifice involved in moral enhancement are not specific to moral *bio*enhancement. If there are limits to how far we must or may go in augmenting our disposition towards self-sacrifice, those limits will apply as much to moral enhancement via introspective reflection, engagement with literature, or moral discussion with others as the will to biological interventions.

Rob Sparrow is in eloquent form, moving gracefully back and forth between incompatible objections. He starts by laying out some criteria for thought experiments and complaining that our thought experiment fails to meet them. “It is not too much of a stretch, then, to characterise their paper as a thought experiment in service of a thought experiment.” But this *is* too much of a stretch. We constructed a thought experiment about being able to change people’s intentions and behavior without their knowledge. Clearly, it is not currently possible to alter people’s intentions and behavior in the fine-grained way that we imagined. However, it is possible to biologically alter intentions and behavior—including morally relevant intentions and behavior—in much messier ways. For example, propranolol, oxytocin and selective serotonin re-uptake inhibitors have all been shown to have influence either morally significant behavior, or moral judgments that are likely to have behavioural effects (Levy et al. forthcoming). Propranolol and SSRIs

are widely used drugs, and oxytocin is an endogenously produced agent whose production and release is affected by widely used drugs including steroids. In addition, some drugs are already used in part for their effects on morally significant behavior: in several European and North American jurisdictions, some sex offenders are offered testosterone-lowering agents ('chemical castration') to help prevent re-offending, and methylphenidate (Ritalin) is widely used in part in order to control what might be regarded as immoral behavior in schools. Thus, our thought experiment is, like most thought experiments, simply a 'cleaned-up' version of something that is already possible, and indeed is already happening. It has relevance to the moral assessment of drugs that are used in part in order to control moral behavior (as in the case of methylphenidate and testosterone-lowering agents) and of drugs that are used for other purposes but are likely to have affects on moral behavior (such as propranolol and SSRIs).

Having criticized our argument for lacking practical relevance, Sparrow then takes the reverse tack, speculating that "there is a real danger that their [our] argument will license attempts to manipulate behavior through drugs and brain implants, which raise profound moral issues that they [we] barely mention." But if this is a thought experiment about thought experiment, with no real world application, how could it license attempts to use drugs or brain implants to manipulate behavior? Either we are engaging in armchair philosophical speculation with no practical application or we are discussing something that could be real. Sparrow criticizes us for both at the same time.

Moreover, he does not substantiate either accusation. On the one hand, it is not clear why mere armchair speculation on the topic of moral enhancement would be a bad thing. After all, armchair speculation is the modus operandi of most philosophers outside of practical ethics. Sparrow worries that our argument "does not illuminate a pressing moral dilemma" and suggests that "[t]he matter of how and why it has become the case that bioethicists feel compelled to discuss the ethics of every hypothetical technology that can't be shown to be impossible is worthy of an essay in its own right". But it is not clear what is positively wrong with such hypothetical speculation, and Sparrow seems to acknowledge that it does have some philosophical interest: "As a piece of philosophy their argument is indeed thought provoking and has significant merits". Would Sparrow raise similar concerns regarding the work of most metaphysicians and logicians, which also illuminates no pressing moral dilemma but is hopefully of some philosophical interest?

On the other hand, Sparrow does not adduce convincing evidence that our argument is likely to be misused to devastating effect. As I argue in my response to his chapter, any prediction that reasonable bioethical discussion will somehow lead to an atrocious outcome must be backed up my more than mere speculation.

We argued in our paper that some forms of moral enhancement would not undermine freedom because they could act, for example, by opening up someone to understanding the suffering of others, like reading Tolstoy. Sparrow objects, "Someone who reads Tolstoy arguably learns *reasons* to be less judgemental and in doing so develops greater understanding: someone who takes a pill has merely

caused their sentiments to alter. In so far as moral action requires acting for the right reasons, the person who has learned tolerance from Tolstoy has more and better reasons for action."

But learning is something that we can be more or less good at. Biological manipulations can enhance learning abilities. It is a common mistake to assume that all moral bioenhancements would directly alter sentiments and thereby directly alter behavior, leaving our deliberative capacities entirely out of the picture. Actually, it may be the case for many that they act by augmenting the normal processes by which we form moral motivations and learn to be moral. Indeed, one of us has previously characterized the most plausible examples of moral enhancement as interventions that alleviate barriers to (among other things) sound moral reasoning (Douglas 2008). Just as steroids do not make a person stronger without physical training, just as cognitive enhancements do not produce enhanced knowledge or cognitive skill without learning, so too moral enhancements may not produce more moral behavior without precisely the activities that Sparrow has in mind. They would just increase the magnitude or likelihood of the benefit from those experiences.

Sparrow suggests that "There is an obvious tension between their description of the naturally virtuous person as someone for whom it is psychologically or motivationally out of the question to do wrong and their later claim that autonomy requires the vivid imagination of alternatives." But we do not see this tension. A person can vividly imagine an option while nevertheless being so strongly motivated not to pursue that alternative that it is rightly described as 'out of the question'.

Indeed, the alternative might be 'out of the question' precisely because the agent imagines how horrible the consequences of that alternative would be. Moreover, it does not follow from the fact that it is motivationally out of the question for the agent to do an immoral act that he is unfree in any morally problematic way. It remains the case that the agent could have acted wrongly.

Of course, in our case of 'perfect mind control' the agent *is* unfree in a morally important way: he genuinely can't act wrongly. But we argued that even in these cases, the agent may still be fully autonomous. Regarding this case, Sparrow raises some interesting points. For example, he argues that "given that people who are subject to the magical "moral technology" are *not* free to do anything other than act morally this suggests that there is an important sense in which they do not act freely even when they choose to act in such a way as the technology does not intervene." This assumes incompatibilism about free will. We do not wish to enter this complex debate but according to compatibilism, freedom can exist even if determinism is true, that is, even if we could never have acted other than we did. If freedom is compatible with complete determinism, it is compatible with the moral technology we describe.

Sparrow closes with his political critique, similar in vein to his chapter in this volume. He argues that much choice is socially constructed. He refers to Angus Dawson's slides on the spread of obesity in the US. The origin of behavior – in individual free choice or through social construction – is indeed interesting. But it is not our target. We have not argued that "social problems [are] rooted in biology." We are

interested in how behavior can be biologically modified, whether it is biological, psychological or social in origin, or some combination of those. It is certainly true that social means can be effective at modifying behavior. And it it may even be true that all our problems are social in origin. We have sought to explore whether biological means can also be employed to deal with these problems and whether their employment would raise new or irresolvable ethical issues.

Sparrow worries that "the project of "moral bioenhancement" invites abuse: "it assumes that we know what moral behavior in various circumstances consists in, where in fact this is both, within limits, controversial and should remain so; it will almost certainly involve the powerful acting on the powerless."

There is certainly controversy about what is right and good. But there is also consensus (Smith 1994). Racism and sexism are wrong. Sexual abuse of young children is wrong. In our recent book, two of us (Ingmar Persson and Julian Savulescu) have focused on the collective action problem of climate change, gross global inequality and threat of annihilation of the human race (Savulescu and Persson 2012). These are all uncontroversially bad states of affairs and we have sought to understand what role knowledge of human biology might play in the future in addressing these.

The powerful have many ways already ready at hand to oppress the powerless. It is hard to see how they need the project of moral bioenhancement to exercise their power. It is precisely the kind of oppression that Sparrow fears which is the target of

our concerns: how could we use knowledge of the nature of the human animal prevent the kind of oppression that already has occurred with relentless frequency and in atrocious magnitude. History has not been rosy (Glover 2001). With the exponentially increasing power of technology together with globalization, the tendency of humans to oppress and harm each other reaches critical mass. More than ever, we require a project of moral enhancement using our knowledge from medicine and science in general.

References

Dworkin, R. 1981. The Concept of Autonomy. In: *Science and Ethics,* R. Haller (ed.). Amsterdam: Rodopi.

Douglas, T. 2008. Moral Enhancement. *Journal of Applied Philosophy* 25(3):228–245.

Glover, Jonathan. 2001. *Humanity: A Moral History of the Twentieth Century.* New Haven and London: Yale Nota Bene.

Harris, J. 2012. Moral Progress and Moral Enhancement. *Bioethics*. doi: 10.1111/j.1467-8519.2012.01965.x

Levy, Neil, Thomas Douglas, Guy Kahane, Sylvia Terbeck, Phil Cowen, Miles Hewstone, and Julian Savulescu. Forthcoming. Are You Morally Modified? The Moral Effects

of Widely Used Pharmaceuticals. *Philosophy, Psychiatry and Psychology*.

Persson, I., and Julian Savulescu. 2008. The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy* 25(3):162 – 177.

Savulescu, Julian and Ingmar Persson. 2012. *Unfit for the Future: The Need for Moral Enhancement.* Uehiro Series in Practical Ethics.Oxford: Oxford University Press.

Sidgwick, Henry. 1884. *The Methods of Ethics*. Macmillan and co.

Singer, Peter. 1997. *How Are We to Live?: Ethics in an Age of Self-Interest.* Oxford: Oxford University Press.

Smith, Michael. 1994. *The Moral Problem*. Malden MA: Blackwell Publishing

Thalberg, I. 1978. Hierarchical Analyses of Unfree Action. *Canadian Journal of Philosophy*, VIII: 2. As reprinted in: *The Inner Citadel: Essays on Individual Autonomy (*pp. 123-36), J. Christman (ed.)