



Should Moral Bioenhancement Be Covert? A Response to Crutchfield

Louis Austin-Eames

Received: 15 February 2023 / Accepted: 10 August 2023
© The Author(s) 2023

Abstract Crutchfield (Crutchfield in *Bioethics* 33:112–121, [4]) has argued that if moral bioenhancement (MBE) ought to be compulsory, then it ought to be covert. More precisely, they argue that MBE is a public health intervention, and for this reason should be governed by public health ethics. Taking from various public health frameworks, Crutchfield provides an array of values to consider, such as: utility, liberty, equality, transparency, social trust, and autonomy. Subsequently, they argue that a covert MBE programme does better than an overt one, in preserving or promoting said values, and hence, that a covert MBE is preferable. In this paper, I will provide novel reasons to doubt that the relevant values are in fact better promoted or preserved by a covert MBE programme. Additionally, I will provide a novel autonomy-based consideration which counts in favour of the MBE programme being overt, rather than covert. Given that as things currently stand it is unclear which kind of MBE programme is preferable, the upshot of my criticism of Crutchfield will be to provide some recommendations as to how we might proceed in establishing whether a covert or overt MBE programme fares better.

Keywords Moral bioenhancement · Crutchfield · Autonomy · Authenticity · Public health intervention

Introduction

In the human enhancement literature, it has been argued that we have an imperative to undergo moral bioenhancement (MBE).¹ There has been difficulty defining exactly what MBE is for a variety of reasons. For example, some take it that moral enhancement is aimed at humanity overall,² whereas others define moral enhancement such that the target is the individual.³ Moreover, there is disagreement regarding whether moral enhancement need be behaviour, or capacities oriented; and disagreement with respect to the means by which the enhancement must occur.⁴ Thus, the definition provided here will likely not capture everyone's understanding of MBE. However, for present purposes, a definition such as the following should suffice: The moral bio enhancement of an

¹ Persson, I. and Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy*, 25(3), pp.162–177.

² Ibid.

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

⁴ For more discussion regarding the disagreement in defining moral enhancement, see Raus, K., Focquaert, F., Schermer, M., Specker, J., & Sterck, S. (2014). On defining moral enhancement: A clarificatory taxonomy. *Neuroethics*, 7(3), 263–273 [1].

L. Austin-Eames (✉)
Philosophy, School of Humanities, University of Glasgow,
Glasgow, Scotland
e-mail: l.austin-eames.1@research.gla.ac.uk

individual consists in the improvement of their moral attitudes, dispositions, and motivations, by means of a biological intervention.

Persson and Savulescu [2, 3] have argued that we have an imperative to undergo MBE in virtue of the fast-approaching possibility to cognitively enhance ourselves, in addition to recent technological developments, which in conjunction make it easier for a single bad actor to cause ‘ultimate harm’ (UH). Moreover, they have argued that it is necessary that humans undergo MBE in order to prevent the greater risk of ultimate harm (UH) ([2], p 101–135.) UH can be understood as an event that causes ‘high rates of death, and for those that survive, immense suffering’ ([4], p1.). An example could include the engineering and spreading of a deadly infectious disease that kills millions.

Under the supposition that we ought to implement a compulsory MBE programme, to thwart UH, the following question has been the subject of recent discussion in the literature: Should such a compulsory MBE programme be covert or overt? In other words, assuming that humanity ought to undergo mandatory MBE, should we be aware of the MBE programme, or should it happen unbeknownst to us? Under the assumption that MBE can be implemented safely and is necessary to prevent UH, Crutchfield [4] has argued that a covert MBE programme is preferable to an overt one. Crutchfield does this in two stages. Firstly, they argue that MBE is a matter of public health, and as a result, that the permissibility of MBE ought to be governed by the ethics of public health interventions (henceforth referred to as public health intervention values, or PHVs). The second stage in Crutchfield’s argument is that a covert MBE programme does better than an overt one at preserving or promoting the PHVs. Therefore, a covert MBE programme is preferable [4], p1.). Zambrano [5] has responded to Crutchfield, raising two broad issues. Firstly, they deny the first stage of Crutchfield’s argument and argue that a MBE should not be evaluated relative to the PHVs, as MBE is not a public health intervention. Secondly, Zambrano [5] argues that there is an autonomy-based reason to prefer an overt MBE to a covert one. Crutchfield [6] has since responded to both of Zambrano’s objections.

Given both: the possible imperative we may have to undergo MBE and the disagreement over how such an MBE programme ought to be implemented, it is

of significant importance that we establish whether a covert or overt MBE programme is preferable. In this paper, I will argue, contrary to Crutchfield, that we have good reason to doubt that a covert MBE programme is preferable to an overt one. Moreover, I will argue that to establish whether compulsory MBE ought to be covert or overt, we plausibly require additional empirical and conceptual work. I will do this in four stages. In Section 2, I grant that MBE is a public health intervention and hence that whether a covert or overt MBE programme is preferable, ought to be evaluated relative to the PHVs. In Section 3, by way of response to Crutchfield, I will provide novel considerations which count against a compulsory MBE programme being covert, concluding that as things currently stand, it is unclear which kind of MBE programme is preferable. In Section 4, I will provide a novel autonomy-based reason to prefer an overt MBE to a covert one, namely that we are able to dissent to MBE only if the programme is overt. Finally, in Section 5, I will conclude with some recommendations as to how we might proceed in establishing whether a mandatory MBE ought to be covert, or overt. More precisely, I suggest that at least three things are required. Firstly, we require empirical research which investigates how people would respond to a mandatory MBE. Secondly, we must estimate how much more likely UH is to occur if the MBE programme is overt. Finally, we require a more plausible notion of authenticity, to that which is offered by Crutchfield [4], to act as a criterion by which to adjudicate the preferability of the different kinds of MBE. I then hint that such recommendations may have a broader scope in the bioethical literature, such that they would help settle other disagreements between bio conservatives and transhumanists.

MBE as a Public Health Intervention

It is worth starting by clarifying how the first stage of Crutchfield’s argument supports the second. That is, how Crutchfield’s view that MBE qualifies as a public health intervention supports the proposition that a covert MBE programme is preferable to an overt one. Roughly speaking, Crutchfield argues that since the PHVs are better preserved or promoted by a covert MBE programme, covert MBE is preferable to overt MBE ([4], p 4–6). Put differently, the PHVs are the

criteria by which we determine which kind of MBE programme is preferable. I offer the following argument intended to illustrate the relationship between the first and second stage of Crutchfield's argument:

P1) If mandatory MBE is a public health intervention, then the PHVs will be salient when adjudicating whether a covert or overt MBE programme is preferable.

P2) Mandatory MBE is a public health intervention.

C1) The PHVs will be salient when adjudicating whether a covert or overt MBE programme is preferable.

P3) If the PHVs are better preserved or promoted by a covert MBE programme, rather than an overt one, then a covert MBE programme is preferable, all else equal.

P4) The PHVs are better preserved or promoted by a covert MBE programme, rather than an overt one.

C2) A covert MBE programme is preferable, all else equal.

In support of P2, i.e., that MBE is a public health intervention; Crutchfield argues roughly as follows: MBE is a health intervention which seeks to prevent UH. UH is a harm suffered by *large groups of people*. Therefore, MBE is aimed at the *public's health*. Hence, MBE is a public health intervention. In other words, MBE is a public health intervention insofar as it primarily seeks to prevent harm to a large number of people, and only secondarily to the individual ([4], p3.). Prior to offering my response to P2, it is worth noting that there is disagreement between Zambrano and Crutchfield on this matter, as Zambrano has argued that compulsory MBE would not qualify as a public health intervention as it fails to meet certain conditions inspired by the Siracusa principles – legal principles that provide conditions under which the government can restrict individual liberties.⁵ Crutchfield has since responded arguing that the proposed conditions are too strong, and that compulsory MBE

remains a public health intervention.⁶ For present purposes, I am going to grant that MBE is a public health intervention. Whilst I agree with Crutchfield, that the conditions proposed by Zambrano seem too strong, this is not my primary motivation for granting that we ought to evaluate MBE programmes relative to the PHVs. This is in part because there may well be other plausible objections to MBE's status as a public health intervention and/or objections to Crutchfield's construal of the different public health ethics frameworks. Ultimately, however, it seems crucial to understand the relationship between the different kinds of MBE programmes and the PHVs, regardless of whether MBE is strictly speaking, a public health intervention. This is because values such as utility, liberty, and equality, among other of the PHVs, will play a significant role when evaluating the competing kinds of compulsory MBE programmes. In other words, we will want to know how different kinds of MBE programmes affect utility, liberty, equality, etc. Thus, going forward, I will be granting that MBE is a public health intervention and hence that the different kinds of MBE programmes ought to be evaluated relative to the PHVs.

With this in mind, we can now look at the PHVs themselves. Crutchfield [4] takes from various public health ethics frameworks, such as those proposed by [8–10], and provides a list of various values that must be considered when implementing a public health intervention. The primary values considered are utility, liberty, equality, fairness, transparency, social trust, and solidarity [4]. For more detail on the relevant public health ethics frameworks, see [4], p 3–4). Crutchfield [4] then proceeds to argue that a covert MBE programme does better than an overt one at balancing the PHVs. That is, a covert MBE programme better preserves or promotes the PHVs overall, all else equal. Therefore, Crutchfield concludes, a covert MBE programme is preferable to an overt one [4].

Is a Covert MBE Programme Preferable?

Having granted that the permissibility of mandatory MBE ought to be evaluated relative to the PHVs, we can now look at the reasoning Crutchfield provides as to why a covert MBE programme is preferable to an overt one. That is, we can examine the considerations Crutchfield provides in support of P4 i.e., that the

⁵ Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights, United Nations. Doc E/CN.4/1985/4, Annex (1985) [7]. For more on this, see: Zambrano ([5], p2).

⁶ See Crutchfield ([6], p2).

PHVs are better preserved or promoted by a covert MBE programme. Firstly, Crutchfield concedes that certain values, namely transparency, social trust, and perhaps solidarity, *are not* better preserved or promoted in a covert MBE programme. That is, Crutchfield concedes that with respect to the preservation or promotion of some of the PHVs, an overt MBE may be preferable. However, Crutchfield holds that utility, liberty, equality, and fairness, *are* better preserved or promoted in a covert MBE programme. Crutchfield offers no way to rank the competing values. However, he proceeds to argue that the latter values, which are better preserved or promoted by covert MBE, have played a more significant role in moral and political philosophy, and therefore, that those who wish to show that an overt MBE programme is preferable, have the burden of showing that the former values are more valuable than the latter ([4] p 4 – 6.).

Before providing an exposition of, and responding to, Crutchfield's support of P4, I will make my position on the matter clear. I am agnostic about whether a compulsory MBE programme ought to be covert, or overt. However, as I will argue, I believe that we have good reason to doubt Crutchfield's view that the PHVs are better preserved or promoted by a covert MBE programme.

In responding to Crutchfield, I want to start by pointing out that in order to establish that an overt MBE programme is preferable to a covert one, one need not establish that the former values, which are better promoted by overt MBE, are more valuable than the latter values, which are better promoted by covert MBE. Rather, one could just show that the latter values are actually better preserved or promoted by an overt MBE, not a covert one. Alternatively, as I will attempt here, one could argue that we have reason to doubt that a covert MBE better promotes the PHVs overall.

With this in mind, I do grant that it seems plausible that preserving or promoting utility, liberty, equality, and fairness, is more important than preserving or promoting transparency and social trust, all else equal. However, surely it depends upon the *degree to which the different PHVs are preserved/promoted or undermined* by each kind of MBE. To illustrate, if a given public health intervention resulted in the promotion of transparency, on a societal level, but simultaneously generated a negligible amount of disutility, say that which is equivalent to a dust spec

in one's eye, it seems that we ought to nevertheless implement the public health intervention. I assume Crutchfield would agree with this. After all, in establishing which kind of MBE programme is preferable, we must establish which kind does a better job of *balancing* the PHVs. Surely this balancing must take into account the degree to which the different PHVs are preserved/promoted or undermined. Given this, I will proceed to argue that we have reason to doubt that covert MBE better preserves or promotes utility and liberty to a degree that warrants undermining social trust and transparency, if at all. And secondly, that when it comes to social trust and transparency, it is clear that they are significantly better preserved or promoted by an overt MBE programme.

Utility

Crutchfield offers two considerations as to why a covert MBE is preferable with respect to promoting or preserving utility. Firstly, in the case of an overt MBE programme, there would be a greater risk of UH occurring, and therefore a greater risk of a vast amount of disutility being generated. The reason there would be a greater risk of UH occurring if the MBE programme is overt, is because there would likely be certain individuals who evade, or dissent to the MBE. The more individuals who remain unenhanced, the more likely UH will occur (at the hands of those unenhanced individuals). Conversely, if the MBE were covert, the individuals who desire not to be enhanced would have much more difficulty avoiding the enhancement, as they would not be aware it was happening in the first place. Thus, there would be fewer unenhanced individuals who could cause UH, and therefore a significant amount of disutility, if the MBE were covert [4].

The second consideration is roughly as follows. In an overt MBE programme, the individuals who do not want to be enhanced would suffer from *knowing* that their desire, not to undergo MBE, is being frustrated. Conversely, if the MBE were covert, no one would know about the MBE and hence no one would suffer from knowing their desire is being frustrated. Therefore, with respect to promoting or preserving utility, a covert MBE programme is preferable (Crutchfield 2018, p 5.).

In response to the first consideration, I want to start by granting that Crutchfield is plausibly correct

that overt MBE would increase the risk of UH occurring, and thereby increase the risk of a significant amount of disutility being generated. This is because I agree that if MBE were overt, it is plausible that more people would remain unenhanced, and that this would increase the risk of UH occurring (at the hands of those unenhanced individuals). However, it seems crucial to establish, at least roughly, *how much more at risk* we would be of suffering UH in the case of an overt MBE programme, to then be able to establish which kind of MBE programme is preferable.

For clarity, I agree that when we weigh up the different values that govern public health interventions, the disutility caused by UH would plausibly warrant undermining transparency and social trust. However, this is distinct to the question of whether *an increased risk* of UH warrants undermining transparency and social trust. Importantly, it is this latter question that we must answer in order to establish which kind of MBE programme is preferable. After all, it is not guaranteed that UH will occur if the MBE programme is overt, it is merely *more likely to occur*. It could be the case that the increase in risk of UH occurring is negligible and thus does not warrant undermining the competing values. That is, the best way to balance the PHVs may well be to preserve social trust and transparency, while accepting the increase in risk of UH occurring. Given the importance of transparency and social trust, and the consequences of undermining them, it is impermissible that we undermine such values without good reason. A mere increase in risk of UH is not sufficient. It must be a large enough increase in risk. Until we have reason to believe that the increase in risk would be large enough, we ought not undermine transparency, social trust (and as I will later argue, autonomy).

By way of analogy, consider a hypothetical covert Covid-19 vaccination programme, which disguises Covid-19 vaccinations as flu vaccinations. Let's assume that this would have decreased the risk of a great amount of disutility ensuing from Covid-19 by making the vaccine compliance rate in the UK 100%. Whether such a programme is permissible seems to heavily depend upon the answer to the following question: How much would the risk of disutility decrease if this programme were implemented? In

the case of Covid 19, given that only 3% of people in the UK refused vaccination in the summer of 2021,⁷ it seems plausible that it would have been impermissible to implement the covert vaccination programme given how severely it would have undermined transparency and social trust, and given how catastrophic the consequences could have been if the programme were discovered. The 3% increase in compliance, and the associated decrease in risk of disutility, would likely not warrant undermining transparency and social trust.

In response to the second consideration, i.e., that individuals would suffer from knowing about the MBE programme, I grant that for those who do not wish to be enhanced, there plausibly would be some disutility generated from an overt MBE programme, which would not be generated from a covert MBE programme. The disutility would of course result from the individuals knowing they are going to be forced to undergo a MBE programme that they wish to avoid. However, this is not the entire picture. Crutchfield has not established that there would be a *net disutility* as a consequence of an overt MBE programme. It may well be the case that the majority of people would experience wellbeing when they come to know of the compulsory MBE. That is, there may be an increase in overall utility in the form of individuals rejoicing at the thought of becoming better (in the moral sense) friends, mothers, fathers, sons, daughters, etc. Moreover, it seems plausible that we would experience a great deal of wellbeing when we come to know that fewer crimes will be committed, that society will become less racist, less sexist, etc., as a result of the MBE programme.

By way of analogy, consider our acceptance of laws, rules and regulations in the workplace/academia, and general social norms, many of which are aimed at achieving morally improved ends. Ends such as less discrimination and harassment, an increase in equality of opportunity, and the limitation of the freedom of those who harm others. Given our efforts to morally improve society with the implementation of such rules and regulations, it may well be the case that this sentiment of endorsing moral progress

⁷ See: <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/numberofindividualswhohaverefuseddeclinedacovid19vaccine>

carries over to compulsory MBE in the following sense: Given that a compulsory MBE would achieve our desired ends – a morally improved society—many of us would likely be in favour of the programme assuming it were safe and effective.

In the case of a covert MBE programme, no one would know about the MBE, and therefore would not experience the wellbeing that comes from knowing MBE is taking place. They would of course experience the wellbeing of being in a less racist and less sexist society, and of being better friends, etc. What I am drawing attention to is the utility that could result from *knowing* that a MBE programme will take place. Ultimately, it has yet to be established, by means of empirical research, whether there would in fact be a net disutility that would result from people knowing about the compulsory MBE programme. We cannot just assume that there would be a net disutility.

Finally, a covert MBE programme carries with it a risk of significant disutility in the following sense. If the covert MBE were to be discovered, this could result in the destabilization of society, as the general public would no longer trust the authorities who implemented the MBE to be truthful. This could result in, among other things, protests, riots, and acts of terrorism. This would, in turn, result in significant disutility. Thus, as far as utility is concerned, we have reason to doubt that it is in fact better preserved if the MBE programme is covert.

Liberty

With respect to liberty, Crutchfield offers a few considerations in support of covert MBE being preferable. Firstly, the consequences of UH, such as harsh conditions and death, would significantly reduce liberties, as ‘dead people have no liberties’ ([4], p5). Since UH is more likely to occur in an overt MBE programme, for the reasons stated above, an overt MBE programme increases the chances of a significant reduction of liberty. Secondly, Crutchfield argues that liberty would be restricted for those who dissent to a mandatory overt MBE programme, as they would be imprisoned, taxed, or isolated. This is a unique problem for an overt MBE programme, as people would not be able to dissent if the MBE were covert, as they would be unaware of the enhancement. Thus, they would not have their liberty restricted in this way [4]. Moreover, in Crutchfield’s [11] book ‘*Moral*

Enhancement and the Public Good’, he later argues that a covert MBE programme is more compatible with libertarianism than an overt one.

With respect to the first liberty consideration, I offer the following response. I concede that there would be a reduction in liberty if UH occurred, due to the dire circumstances we would be in. Moreover, I have granted that there would be a greater risk of UH in the case the MBE were overt. However, it seems that the degree to which the risk is increased is once again a crucial factor, for it may be the case that the increase in risk is small enough that it does not warrant undermining transparency and social trust. Put differently, if left with a choice between either: undermining transparency and social trust, or preventing the liberty reducing consequences of UH, it may well be that choosing the latter is a better means of balancing the PHVs. However, choosing between *increasing the risk* of a substantial reduction in liberty occurring, or undermining transparency and social trust, is a different matter. It seems that in order to know which of the latter two choices achieves a better balancing of the PHVs, we need to know, at least roughly, how much more at risk we would be of UH in the case the MBE were overt. Thus, with respect to the first liberty consideration, until we establish how much more at risk of UH we are in the event of an overt MBE programme, it is not clear that a covert MBE is preferable.

The second consideration Crutchfield offers, namely that those who dissent to a mandatory overt MBE would be imprisoned, taxed, or isolated, and that this would result in a reduction of liberty, seems correct. However, it seems that although a covert MBE programme would preserve liberty in this way, it would also undermine autonomy. More precisely, covert MBE would remove people’s ability to self-govern and dissent to the MBE programme, if they so desire.⁸ On the other hand, if the MBE were overt, individuals would be in a position to self-govern and act in accord with their desires, choosing to either undergo the MBE, or to suffer the liberty restricting consequences of dissenting. Thus, although covert MBE preserves liberty insofar as there would be no dissenters, and therefore fewer people being taxed or

⁸ I will discuss this autonomy-based consideration further in the autonomy section.

imprisoned, we have reason to doubt that such liberty considerations count in favour of an MBE programme being covert, all things considered. It is worth noting that I am using a notion of autonomy that Crutchfield seems to accept, i.e., ‘to be in a position to self-govern in accordance with one’s desires, values, or other attitudes’ ([4], p. 9).

In addition to the above considerations, Crutchfield [11] provides various arguments in support of the claim that a covert MBE is more compatible with libertarianism.⁹ I will not be addressing all such arguments here, as the primary point I want to make is that even if it is true that covert MBE is more compatible with libertarianism, it is nevertheless the case that covert MBE either merely *decreases the risk* of liberty being undermined (to an unknown degree) and/or it preserves liberty in such a way that also undermines autonomy, social trust, and transparency. Such liberty considerations do not seem to render covert MBE preferable all things considered. That being said, given that I find Crutchfield’s reasoning dubious, I do want to draw attention to one of Crutchfield’s arguments that a covert MBE is more compatible with libertarianism.

Very broadly, Crutchfield argues that it is preferable that MBE be covert because it is less of an intrusion on peoples’ liberty than if the programme were overt. That is, having one’s liberty intruded upon (by the MBE programme) and knowing about it, is worse than being intruded upon and not knowing about it. The knowledge of the intrusion makes the overt MBE programme more of an intrusion ([11], p 147). In response, I see no reason to accept this claim. Having your liberty intruded upon seems orthogonal to knowing whether your liberty is being intruded upon. Insofar as liberty denotes the absence of obstacles which prevent you from doing what you desire, or the presence of control on the part of the agent,¹⁰ it seems that not knowing about the MBE merely deludes people into thinking they have more liberty. Not knowing about the intrusion does not increase the public’s

control, nor does it remove obstacles which prevent people from doing what they desire.

Finally, as is the case with utility, the risk of people becoming aware of a covert MBE programme carries with it the risk of a significant reduction in liberty. As discussed above, if the general public were to become aware of the covert MBE, rioting, protests, and terrorism may well ensue. The consequences of such riots, etc., would be the reduction in liberty in the form of the imprisonment, taxation, etc., of those who rioted. Thus, as far as liberty is concerned, we have some reason to doubt that it is in fact better preserved or promoted by a covert MBE programme, or insofar as it is better preserved, it seems that covert MBE merely reduces the risk of undermining liberty at the expense of transparency, social trust, and as I will later argue, autonomy.

Social Trust and Transparency

Whilst Crutchfield [4] concedes that an overt MBE programme is likely preferable when it comes to promoting social trust and transparency, he later argues that transparency is not necessary for public health interventions to be permissible [11], p.111–131). In large part, Crutchfield does this by way of response to Childress et al., who believes that it is necessary.¹¹ The aim here, seems to be to mitigate the unpalatability of covert MBE undermining transparency. To this end, Churchfield provides various arguments in support of his claim, some of which address whether transparency is required to have respect for persons, others of which address whether transparency engenders trust.¹² I will not be providing an exposition of all of Crutchfield’s reasoning here. Instead, I will be considering what I take to be his strongest arguments to the conclusion that transparency is not required for a public health intervention to be permissible.

Broadly speaking, Crutchfield argues that we should be sceptical of the notion that being transparent to the public is more valuable than promoting the values which are brought about by a given public

⁹ See: Crutchfield, P. (2021) Moral enhancement and the public good. New York, NY: Routledge, p 131 – 154.

¹⁰ See: Carter, Ian, "Positive and Negative Liberty", The Stanford Encyclopedia of Philosophy (Spring 2022 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative/>> [12].

¹¹ See: Childress, J., Faden, R. R., Gaare, R. D., Gostin, L. O., Kahn, J., Bonnie, R. J., ... Nieburg, P. (2002). Public health ethics: mapping the terrain. The Journal of Law, Medicine & Ethics, 30(2).

¹² See: Crutchfield, P. (2021) Moral enhancement and the public good. New York, NY: Routledge, p 111 – 130.

health intervention – in this case a covert MBE programme. More precisely, since values such as utility and liberty are ostensibly better preserved in a covert MBE programme, then the morally correct action is to implement covert MBE, even though such a programme would undermine transparency, as it is more important to promote the former values ([11], p.113). In response, whilst I am generally on board that transparency is not a necessary condition for the permissibility of a given public health intervention, as it may be outweighed by other competing values, we do not have sufficient reason to accept that this is the case with covert MBE. This is because a covert MBE programme *merely decreases the risk* of the other values being undermined. As argued above, this decrease in risk could be negligible and thus not warrant undermining transparency, much like in the covert Covid-19 vaccination analogy. Empirical research is required to establish whether the relative difference in risk between two kinds of MBE programmes warrants undermining transparency (social trust and autonomy), in addition to accepting the risk of the consequences of undermining such values, such as those that would ensue from the public finding out about covert MBE.

Crutchfield [11] offers an additional two considerations which support the claim that transparency is not necessary nor important for the permissibility of public health interventions. The first is that it is not necessary for social trust. Crutchfield provides some empirical evidence which suggests that an increase in transparency, with respect to governmental decision-making processes, for example, decreased trust in the governments' competence and benevolence.¹³ Crutchfield ([11], p.117) argues that at best, transparency does not seem to increase social trust, and at worst, that it may weaken it. Secondly, in an attempt to show that not only is transparency unnecessary, but that it is also not particularly valuable all things considered, Crutchfield cites further empirical evidence. The evidence he cites purports to show that people generally desire their policies to be made in the dark, so long as they have the opportunity to shed light upon them if they so desire. Crutchfield stipulates: *'The ideal situation for many people is one in which*

they are not involved in the decision-making process, a process that doesn't even occur out in the open.' (2021, p.123). He continues by arguing that although we would be unable to shed light on the covert MBE, as we would not know about it, this research shows that transparency is not as valuable as we might have thought ([11], p. 123).

In response, whilst it is an empirical question whether covert MBE could be successful, in that it could occur completely unbeknownst to us, it nevertheless seems plausible that someone would discover the covert MBE programme. This is because implementing such a programme would presumably be a complicated logistical task involving the cooperation of many individuals and organisations. It seems implausible to assume that all involved would successfully keep the MBE programme a secret. Moreover, if the MBE programme were discovered, the public would become aware that they are being deceived (by omission), in that their values and desires are being altered unbeknownst to them. It seems reasonable to assume that this level of deception would significantly undermine social trust.¹⁴ That is, it seems plausible to assume that the public would fear that many other covert programmes are being implemented unbeknownst to them or that other mandatory programmes, such as vaccinations, are not what they seem.

Moreover, we have reason to be sceptical that the evidence Crutchfield provides, which purportedly shows that we want policies to be made in the dark so long as we have the ability to shine a light on them, provides us with any information regarding whether the public would be in favour of covert MBE. Firstly, it seems misguided to extrapolate from the answers to the questions the participants were asked, to their support of opaque policy making generally.¹⁵

¹⁴ A somewhat analogous example is 'fake news'. After finding out about fake news, 58% of individuals reported less trust in social media news stories. See: Reuters. 2017. Editorial. Oct. 31. "Fake News Hurts Trust in Media, Mainstream Outlets Fare Better". New York: Thomson Reuters.

¹⁵ Participants were asked the following 3 questions according to Crutchfield: '(1) elected officials would help the country more if they would stop talking and just take action on important problems; (2) what people call compromise in politics is really just selling out on one's principles; (3) our government would run better if decisions were left up to nonelected, independent experts rather than politicians or the people, or that our government would run better if decisions were left up to successful business people.' See Crutchfield [11]

¹³ See: Crutchfield, P. (2021) Moral enhancement and the public good. New York, NY: Routledge, p 117.

Secondly, and more importantly, this evidence does not provide support for people being in favour of covert MBE. This is because, as Crutchfield acknowledges, the public would not be able to shed light on the MBE programme if they so desired, as they would be unaware of it occurring. That the public would be in favour of their policies being made in the dark *without* the ability to shed light on them, is something Crutchfield has not provided support for. Hence, it is not something we should grant. Ultimately, if the MBE programme were covert, billions of people would not be informed that their desires and values are being altered by the enhancement. The scale of the covert MBE programme would constitute an egregious undermining of transparency.

If I am correct, then many of Crutchfield's considerations in favour of a covert MBE programme being preferable to an overt one, seem implausible. Thus, as things currently stand, we have good reason to doubt that covert MBE fares better.

Autonomy

In this section, I will provide a novel autonomy-based reason that counts in favour of the MBE programme being overt, rather than covert, which I believe fares better than the considerations on offer from Crutchfield and Zambrano. Beforehand, I will layout an exposition of the current disagreement between Crutchfield and Zambrano regarding autonomy and compulsory MBE and argue why I do not think either or their accounts works.

In addition to the broad notion of autonomy above, i.e., 'to be in a position to self-govern in accordance with one's desires, values, or other attitudes', Crutchfield [4] takes the following two conditions to be constitutive of autonomy. Firstly, there is the 'independence condition', which stipulates that one must have the ability to reason and choose independently of external manipulation. The second condition is the 'authenticity condition', which stipulates that one's desires, values, attitudes, etc., must be authentic ([4], p 9.).

With respect to the independence condition, Crutchfield [4] argues that if mandatory MBE does undermine autonomy in this way, covert and overt MBE do so equally, as it is the mandatory nature of MBE that results in external manipulation, not the

covertness ([4], p 9). As a result, there is no autonomy-based reason (in the independence from external manipulation sense), to prefer one kind of MBE programme over another. Zambrano [5] disagrees. They take it to be the case that overt MBE better preserves autonomy, as it undermines self-governance and independence from external manipulation less than covert MBE. More specifically, Zambrano argues that in an overt MBE programme alone, people would be in an *epistemic position* such that they could decide to either *reject or embrace* the new values and attitudes brought about by MBE ([5], p 4.). This epistemic position promotes autonomy as people can decide whether to accept the values brought about by the MBE programme. Hence, they can decide to reject the external influence of MBE if they so desire. If the MBE were covert, on the other hand, people would be unaware that they are being manipulated and thus could not decide to reject the new desires brought about by MBE, rather they would accept the new desires and values as their own [5]. Crutchfield has since responded to Zambrano, arguing that such a strong epistemic position, which enables people to reject the values and desires brought about by the MBE, would render the MBE ineffective. This is because people could simply decide not to embrace any of the values and desires brought about by the MBE. Consequently, if the compulsory MBE is ineffective, then it would not be permissible [6].

In response to this back and forth, although I agree with Zambrano's conclusion, namely that an overt MBE better preserves autonomy insofar as it enables independence from external manipulation, I think their reasoning is problematic. More specifically, I take it to be unclear that we would in fact be in such an epistemic position, such that we could reject the values and desires brought about by the MBE. In other words, it seems dubious that we would be able to 'reject' the desires and values brought about by the MBE, whatever that would mean. Presumably 'rejecting' a desire denotes no longer having the desire and/or not acting in accord with it. I do not think it is at all clear that we would be able to do this. To illustrate why, take, for example, the widely shared desire to help loved ones. Assume that we discover that such a desire is in some meaningful way the result of external manipulation. It is not clear to me that being aware that our desire to help loved ones was brought about by the external manipulation of others would

enable us to reject such a desire. Alternatively, take the valuing of wellbeing. If I were to become aware that others had manipulated me in such a way to make me value wellbeing, it does not follow that I would be able to reject this value. Broadly speaking, it may be the case that we are no more able to reject our deeply held desires and values, than we are able to 'reject' our eye colour. Our deeply held desires may be such that we are unable to simply will them away and no longer act in accord with them. At minimum, it seems that more has to be done to establish that we could in fact reject the desires and values brought about by the MBE. Moreover, it is worth mentioning that the kinds of desires that would result from a MBE programme would not be frivolous or superficial desires; they would be our deeply held moral commitments.

Although I disagree with Zambrano's reasoning, I believe that there is nevertheless an autonomy-based reason to prefer an overt MBE programme to a covert one. More precisely, if the MBE is overt, we gain the ability to dissent. In being able to dissent to the MBE programme, we are able to self-govern and avoid the external manipulation of our desires and values. This constitutes a preservation of autonomy that is only possible if the MBE programme is overt, as individuals must first be aware of the MBE programme in order to then be able to dissent to it. It is worth mentioning that Crutchfield could object that although an overt MBE programme would preserve autonomy by enabling us to dissent, it would do so at the expense of undermining other competing values and is thus impermissible. More precisely, since overt MBE preserves autonomy in virtue of allowing for our dissent to MBE, it thereby also increases the risk of UH occurring (at the hands of the dissenters), which in turn increases the risk of values such as utility and liberty being undermined (for reasons discussed in Sect. 3).

In response to this objection, there is indeed tension between autonomy and liberty in that by preserving our autonomy by dissenting, we would reduce our liberty. However, by making the programme overt, we can decide for ourselves whether the trade-off is worth it. If the programme is covert, the decision has been made for us. Moreover, whilst it *could* be the case that the autonomy-based reason is outweighed by the risk of the competing values being undermined, we do not know the relative difference in risk. Thus, we do not currently have

sufficient reason to undermine autonomy (transparency and social trust) by making the programme covert. Importantly, we have an autonomy-based reason to prefer an overt MBE programme. Whether such a reason will be overridden by competing reasons is something that will have to be established by empirical means. Until such research has been carried out and demonstrates that the likelihood of undermining values such as utility and liberty is high enough to warrant undermining autonomy (social trust, and transparency), Crutchfield's possible objection is not successful.

Having responded to the 'independence condition', we can now move on to the 'authenticity condition'. Crutchfield argues that insofar as authenticity is concerned, a covert MBE programme is preferable to an overt one. Put briefly, this is because in a covert MBE programme, the enhanced individuals would have an easier time accepting the new values and desires (brought about by the MBE), as their own. The enhanced individuals would have an easier time embracing the new values and desires as their own because they would be unaware of the MBE programme. The changes in their desires and values would appear to have occurred naturally and hence appear to be their own. Conversely, in the case of an overt MBE programme, it would be apparent, to the enhanced individuals, that their new desires and values are not their own, but rather the result of external manipulation. This would make it more difficult for the enhanced individuals to accept the new values and desires as their own [4]. In other words, Crutchfield [4] takes it to be the case that the authenticity condition is satisfied if the enhanced individuals are able to embrace the new values as their own. Thus, in a covert MBE programme, the new values and desires would be authentic, whereas in an overt MBE programme, they plausibly would not be ([4], p9.).

Zambrano [5] has since objected, taking issue with Crutchfield's notion of authenticity. That is, Zambrano argues that it is not sufficient for authenticity, that the desires brought about by the MBE programme be easier for the enhanced individuals to accept as their own. Rather, the desires and values brought about by the MBE must actually *be one's own* [5]. Moreover, in an attempt to illustrate why Crutchfield's notion of authenticity seems implausible, Zambrano asks us to consider something like the following case:

Suppose an evil genius manipulated us such that we developed ‘immoral desires’. Moreover, suppose that this was done covertly and in such a way that we accepted the new immoral desires as our own. Zambrano argues that on Crutchfield’s view, the desires would be authentic, as we would accept them as our own. However, this seems extremely implausible as the desires are the result of external manipulation. Given that the new desires are the result of external manipulation, they are not our own, and hence far from authentic ([5], p 4.).

Crutchfield [6] has since responded to Zambrano, biting the bullet and granting that the new immoral desires brought about in the evil genius case would in fact be authentic, even though they *merely appear* to be one’s own [6]. Crutchfield then proceeds to argue that Zambrano’s understanding of authenticity, i.e., that the new values and attitudes brought about by the MBE must *actually be one’s own*, is surely mistaken.¹⁶ However, even if we assume that Crutchfield’s objection is plausible, and hence that Zambrano’s conception of authenticity is mistaken, it does not follow that we should adopt Crutchfield’s notion of authenticity. In other words, in drawing our attention to the possibility that Zambrano’s conception of authenticity is problematic, nothing has been done to address Zambrano’s objection that desires that merely appear to be one’s own, is not sufficient for said desires to be authentic. Crutchfield [6] just seems to bite the bullet and agree that in cases such as that of the evil genius, our new desires would in fact be authentic, which seems extremely implausible.

In an attempt to add to Zambrano’s objection, it is worth pointing out that Crutchfield’s notion of authenticity seems to also entail that desires and beliefs brought about by brainwashing can be authentic. That is, so long as the brainwashed individual accepts the new beliefs and desires as their own, then they are authentic beliefs and desires. It seems to me that entailments such as those highlighted by the evil genius case and cases of brainwashing, give us good reason to reject the authenticity condition proposed by Crutchfield, and as a corollary, suspend judgment regarding whether a covert MBE does a better job than an overt one, at preserving autonomy (as far as the authenticity condition is concerned).

In summary, I have argued that with respect to the independence condition of autonomy, there is a reason to prefer an overt MBE to a covert one. Moreover, with respect to the authenticity condition, I have argued, contrary to Crutchfield, that we should not accept that covert MBE better preserves authenticity given that Crutchfield’s understanding of authenticity entails that beliefs and desires brought about by brainwashing can be authentic so long as they are accepted as one’s own beliefs and desires.

Recommendations

Before concluding, I want to draw attention to the upshot of my critique of Crutchfield, by providing a few recommendations as to how we might proceed in establishing which kind of MBE programme is preferable. Firstly, we require empirical research that gauges the public’s response to a compulsory overt MBE programme. This research would inform the public of the goals achievable with the implementation of MBE – a morally improved society conducive to less prejudice, and more kindness and compassion, etc. This research will provide us with the information necessary to determine whether people knowing about the overt MBE is likely to result in a net disutility. If people are in favour of the programme, then we have reason to believe that compulsory overt MBE would generate utility, not disutility, as the public would be aware that their desires to live in a morally improved society are going to be satisfied. A worry about this research, could be that it assumes that people have reliable moral attitudes, when in fact they do not.

In response to such a worry, I am not claiming that if the public were to endorse compulsory overt MBE, that this entails that overt MBE is morally better than covert MBE. After all, it could be the case that the public is mistaken (or morally unreliable) regarding which kind of MBE ought to be implemented. My critique is directed at Crutchfield’s view that being aware of the MBE programme would cause disutility because people’s desires to remain unenhanced would be frustrated. I have argued that it may well be the case that many (perhaps the majority of) people would in fact support MBE. If this is true, then we cannot grant (as Crutchfield does) that peoples’ awareness of the compulsory MBE programme would

¹⁶ See: [6] p.4 for the objection.

result in disutility, as their desires would not be frustrated. To establish how the public would respond, this research is required.

Secondly, we require research which investigates how much more at risk we would be of UH occurring in the event of an overt MBE programme, relative to a covert one. Such research would, at minimum, attempt to estimate what percentage of the public would dissent to the MBE, in addition to the likelihood of the dissenters causing UH. Whilst the details of this empirical work are out with the scope of this paper, the research may involve determining the prevalence of traits that correlate with causing UH, such as the ability to cause UH, sociopathy, low self-control, political ideology, etc. Importantly, as things currently stand, we are unaware of how greater the risk of UH would be if the programme were overt, and hence we do not know if the increase in risk is sufficient to warrant undermining the competing values.

Finally, we require a plausible account of authenticity to act as a criterion by which to adjudicate which kind of MBE is preferable insofar as the authenticity condition of autonomy is concerned. In allowing for the beliefs and desires of those who are brainwashed to be authentic, Crutchfield's definition of authenticity seems implausible.

The above recommendations point to a broader problem with much of the human enhancement literature. Firstly, authenticity has been defined in a myriad of different ways. To illustrate, some have argued that an individual's authenticity can be undermined if characteristics that are fundamental to their identity are changed [13]. Others have argued that we are most authentic when we are unaltered and "natural" [14]. Others again, have argued that the following conditions must be met for a belief to be authentic: responsiveness, non-exclusivity, proportionality, and lack of alienation [15]. Given the ambiguity regarding what is being referred to when we talk about authenticity, the conceptual work I recommend seems required to settle much of the debate between bio conservatives and transhumanists who disagree on the bases of authenticity concerns.

Secondly, it seems that the need for additional empirical work pertains to other areas of the moral enhancement literature, not merely to establish which kind of MBE programme is preferable. For example, such research is also plausibly required to establish whether we can in fact reject desires brought about by

MBE—something Zambrano assumes we could do (2018, p 4.). Moreover, research that establishes the likelihood of UH resulting at the hands of the unenhanced could help settle the debate regarding whether a compulsory MBE ought to be implemented in the first place. This is something Persson and Savulescu [2, 3] argue we have an imperative to do, but something that may not be supported by the evidence.

Conclusion

In this paper, I have addressed the recent dialectic between Crutchfield and Zambrano regarding whether a mandatory MBE programme ought to be covert or overt. After granting that the permissibility of MBE ought to be evaluated with respect to the PHVs, I have provided novel reasons to doubt that a covert MBE programme is preferable to an overt one. Moreover, I have offered a novel autonomy-based reason which counts in favour of a compulsory MBE programme being overt. Subsequently, I offer three recommendations as to how we might proceed to establish which kind of MBE programme is preferable. Firstly, we must carry out more empirical research to gauge the public's response to compulsory MBE. Secondly, we must establish, at least roughly, how much more at risk of UH we would be if the MBE were overt. Thirdly, we must use a more plausible notion of authenticity to act as a criterion by which we can adjudicate which kind of MBE fares better.

Declarations

Competing Interests The author(s) declare none.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Raus, K. et al. 2014. "On defining moral enhancement: A Clarificatory taxonomy," *Neuroethics* 7(3):263–273. <https://doi.org/10.1007/s12152-014-9205-4>.
2. Persson, I., and J. Savulescu. 2012. *Unfit for the future: The need for moral enhancement*. Oxford, UK: Oxford University Press.
3. Persson, I., and J. 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.
4. Crutchfield, P. 2018. Compulsory moral bioenhancement should be covert. *Bioethics* 33 (1): 112–121.
5. Zambrano, A. 2018. Covert moral bioenhancement, public health, and autonomy. *Bioethics* 33 (6): 725–728.
6. Crutchfield, P. 2019. It is better to be ignorant of our moral enhancement: A reply to Zambrano. *Bioethics* 34 (2): 190–194.
7. Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights, United Nations. Doc E/CN.4/1985/4, Annex (1985).
8. Childress, J., Faden, R. R., Gaare, R. D., Gostin, L. O., Kahn, J., Bonnie, R. J., ... Nieburg, P. (2002). Public health ethics: mapping the terrain. *The Journal of Law, Medicine & Ethics*, 30(2): 170–178.
9. Grill, K., and A. Dawson. 2017. Ethical frameworks in public health decision-making: Defending a value-based and pluralist approach. *Health Care Analysis* 25: 291–307.
10. Selgelid, M.J. 2009. A moderate pluralist approach to public health policy and ethics. *Public Health Ethics* 2 (2): 195–205.
11. Crutchfield, P. 2021. *Moral enhancement and the public good*. New York, NY: Routledge.
12. Carter, Ian. 2022. "Positive and Negative Liberty", The Stanford Encyclopedia of Philosophy (Spring 2022 Edition). Edited by Edward N. Zalta. URL = <<https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative/>>. Accessed 10 June 2023
13. Elliott, C. 1998. "The Tyranny of Happiness: Ethics and Cosmetic Psychopharmacology," In *Enhancing Human Traits: Ethical and Social Implications*, ed. E. Parens. Washington D.C.: Georgetown University Press.
14. Maslen, H., N. Faulmüller, and J. Savulescu. 2014. Pharmacological cognitive enhancement-how neuroscientific research could advance ethical debate. *Frontiers in Systems Neuroscience* 8: 107. <https://doi.org/10.3389/fnsys.2014.00107>.
15. Gordon, E.C. 2022. Cognitive enhancement and authenticity: moving beyond the Impasse. *Med Health Care and Philos* 25: 281–288. <https://doi.org/10.1007/s11019-022-10075-2>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.