4

ENHANCING RESPONSIBILITY AND RESPONSIBLE ENHANCEMENT: MORAL BIOENHANCEMENT AND THE ACTUAL-SEQUENCE ACCOUNT OF MORAL RESPONSIBILITY

Niël Conradie

Abstract: In this paper I investigate the relationship between moral bioenhancement and the actual-sequence account of moral responsibility (AS). I first provide definitions of the notions of moral bioenhancement and moral responsibility that I use in this paper, where these notions are based on the capacity definition put forward by DeGrazia (2012) and the guidance theory account of responsibility developed by Fischer and Ravizza (1998) respectively. I then address some shortcomings in these accounts and offer improvements. Thereafter, I explore the relationship between them along two dimensions: backward-looking and forward-looking responsibility. On the backward-looking path I contend that, when assessed through the lens of the AS, moral bioenhancement is permissible only provided that it does not override or degrade the reasons-responsiveness of the enhanced agent. However, if these conditions are met, such enhancement will in many cases be normatively desirable, and

in perhaps some cases –when the foreseen threat of harm resulting from the absence of enhancement is sufficiently great– even obligatory. Along the forward-looking track, I argue that moral bioenhancement is permissible, *ceteris paribus*, and normatively desirable given there is a *pro tanto* reason to pursue promotions of reasons-responsiveness. Finally, I argue that the demands of moral responsibility do not –on their own terms– rule out the possibility that moral bioenhancement can sometimes be obligatory.

Keywords: *Moral Enhancement, Bioethics, Medical Ethics, Moral Responsibility, Reasons-Responsiveness, Actual-Sequence Mechanism.*

INTRODUCTION

One of the many concerns raised by the possibility of moral bioenhancement regards whether or not such enhancement may undermine our usual moral responsibility practices. Is an agent who has been so enhanced still praiseworthy for doing the right thing, or more criticisable for failing to do so than would otherwise be the case? And what are the consequences of this for the permissibility of moral bioenhancement? In this work, I seek to shed some clarity on both these questions. I will do so by examining the relationship between moral bioenhancement and a leading theory of moral responsibility present in the literature – the actual-sequence mechanism account of moral responsibility (AS). Firstly, what is meant by moral bioenhancement is clarified, and secondly the AS is explained. This done, I then explore two dimensions of interaction between these two elements: the backward-looking dimensions and the forward-looking dimension. In the first of these I examine the consequences of moral bioenhancement for the blameworthiness and praiseworthiness of agents. In the second one, I consider what forward-looking responsibilities concerning moral bioenhancement arise from the interaction with moral responsibility.

The aim of this piece is not to provide a final or all-things-considered judgement as to the permissibility or desirability of moral bioenhancement – such a conclusion would require the consideration of many factors that are not addressed here. Rather, my goal is to chart the conceptual space opened up by the interaction of moral bioenhancement and the AS. This charting takes place at a reasonable level of abstraction, as many (if not

all) of the possibilities considered (e.g. promoting reason-responsiveness or hindering it) are multiply realisable through several technological interventions, and the exact shape of these moral bioenhancement technologies remains inconclusive. Thus, the usefulness of this conceptual map is not constrained to current or near-future technologies but can have applicability even when we are considering very different moral bioenhancement technologies in the future.

In conclusion I will contend that, when assessed through the lens of the AS, moral bioenhancement is permissible only provided that it does not override or degrade the reasons-responsiveness of the enhanced agent. However, if these conditions are met, such enhancement will in many cases be normatively desirable, and in perhaps some cases –when the foreseen threat of harm resulting from the absence of enhancement is sufficiently great– even obligatory.

1. MORAL BIOENHANCEMENT AND CAPACITIES

DeGrazia (2012: 361) gives the definition of moral enhancement as follows:

> *moral* enhancement: interventions that are intended to improve our moral capacities such as our capacities for sympathy and fairness

He takes moral *bio*enhancement to be moral enhancement that employs biomedical interventions to achieve its ends. I will be preferring this definition throughout this piece, as I take it to be superior to rival definitions (see Douglas 2008; Crutchfield 2016) in that it places the focus on our *moral capacities* rather than attitudes or motivations.[1] However, I will amend it by removing the qualifier "intended". Interventions that are intended to improve moral capacities but in fact do not do so, are not enhancements – though they may be *attempted* enhancements. On the other hand, an intervention needs not result in behavioural change in order to qualify as an enhancement. To illustrate this, imagine that a

[1] In this I follow a number of other thinkers who have argued for the capacity account (for examples see Jotterand 2011; Simkulet 2012). For an excellent summary of the arguments supporting the focus on capacities over behaviours, attitudes or motivations, see Raus *et al.* 2014: 7-9.

patient undergoes some moral bioenhancement procedure, one that indeed alters her in such a way that there are many possible situations where she would be more capable to act morally than before it. Yet, by a fluke of luck, none of these situations ever arise, and so none of the improved capacities are ever exercised. Yet, she should still qualify as enhanced, and the intervention as an enhancement. This would be so even if she was killed directly after awakening from her procedure by a chance meteorite strike. Simply put: *an improvement to a capacity needs not lead to improved outcomes to qualify as an enhancement.*

There are many possible candidates for moral capacities open to enhancement, with DeGrazia offering sympathy and fairness as examples. Though these two are important and necessary capacities –such that their utter lack would likely leave an agent incapable of moral reasoning– my focus will be on a different capacity, that of *moral reason-responsiveness* (RR). Introduced in its current form by Fischer and Ravizza (1998), RR is the combination of the capacity to identify moral reasons and acknowledge their normative force (called reasons-receptivity) and the ability to adjust behaviour in virtue of these reasons (reasons-reactivity).[2] This is related, but not necessarily identical to, the notion of moral decision-making as it is used by Lara and Deckers (2019). Depending on

[2] The scope of RR is much the same as the set of conditions for moral enhancement put forward by Shook (2012: 6). According to Shook, moral enhancement should do one or more of the following:

1) Enhance a person's sensitivity to moral features of situations – resulting in heightened moral appreciation

2) Enhance a person's thoughtfulness about doing the right thing – resulting in stronger moral decisions

3) Enhance a person's moral judgments that get the right moral answer – resulting in more correct moral judgments

4) Enhance a person's motivated choice to do what moral judgment indicates – resulting in improved moral intention

5) Enhance a person's volitional power to act upon a moral intention – resulting in more will power

All of these five elements are covered by RR, as moral appreciation and thoughtfulness contribute to reasons-receptivity, while improved judgement, intentions, and will power contribute to reasons-reactivity. Where my approach seems to differ from that of Shook is that I contend that enhancement can be felicitously said to have taken place even if the capacities improved are never in fact exercised. So, for example, an improvement in an agent's sensitivity to the moral features of a situation would constitute a moral enhancement *even if* this in fact never results in heightened moral appreciation.

the exact understanding of decision-making, not every exemplification of RR will necessarily be an instance thereof, as at times our responses to reasons can be habitual or reflexive. If decision-making is to be understood as a conscious and deliberative process, then it has a narrower scope than RR.

To be clear, although the content of an agent's motivational set –which may be shaped by an agent's capacity for fairness and sympathy– may well play a role in determining an agent's receptivity and reactivity to moral reasons, I contend that RR is the *operative capacity* for moral bioenhancement. By operative capacity I mean that it is the capacity that *must* be improved for moral bioenhancement to take place, and that an intervention that improves RR via the use of biomedical intervention will *necessarily* be an instance of moral bioenhancement. Improvements to other capacities can constitute moral bioenhancement only insofar as they yield improvements to the capacity of RR. As an analogy, consider a case where we are trying to improve the performance of a car in a drag race. For any intervention we make to constitute an enhancement of the car in this regard, that intervention must improve the car's capacity for speed. This can be achieved through multifarious means –improved aerodynamics, improved engines, lower weight, etc.– but speed is the operative capacity as any intervention is only an enhancement if it improves the capacity for speed. The reason for this contention of the centrality of RR for moral bioenhancement is quite straightforward: only behaviour undertaken in the light of responsiveness to moral reasons can be an expression of moral agency.[3] It is precisely the apparent lack of this capacity in non-human animals, small children, and the sufficiently mentally handicapped that removes them from the set of moral agents.[4]

[3] See Anscombe 1963; Fischer and Ravizza 1998; Arpaly and Schroeder 2013 among many others for arguments to this effect.

[4] Some non-human animals and many (likely most) among the mentally handicapped may well possess some degree of RR, and so are moral agents, with at least some ability to express moral agency. Small children develop their capacity for RR over time, and our readiness to ascribe to moral agency to them seems to follow (at least when we do this felicitously) in step with this development. I take this as indirect evidence for my contention regarding the centrality of RR for questions of moral enhancement.

Having outlined and clarified the understanding of moral bioenhancement that I will be employing, the next step in the path towards mapping the relationship between such enhancement and moral responsibility must be to provide similar clarification for the account of moral responsibility at stake. I turn to this in the following section.

2. THE ACTUAL-SEQUENCE MECHANISM ACCOUNT OF MORAL RESPONSIBILITY

If the relationship between moral bioenhancement and moral responsibility is to be properly explored, it is unavoidably necessary to clarify the account of responsibility that one will employ. There are a great many accounts of moral responsibility, with numerous points of agreement and disagreement (for examples see Strawson 1962; Frankfurt 1988; Kane 1996; Fischer and Ravizza 1998; McKenna 2012; Vargas 2013; Pereboom 2013; Arpaly and Schroeder 2013; Rosen 2014; Scanlon 2015; Shoemaker 2015; Björnsson 2017). For my purposes in this paper I will avail myself of one of the most influential of the current contenders in the field, the *actual-sequence account*. The rest of this section will be devoted to the aim of unpacking the content of this account, as well as discussing some minor improvements to it. Proponents of this account (see Fischer and Ravizza 1998; McKenna 2013; Arpaly and Schroeder 2013; Sartorio 2016) accept the truth of two theses (Sartorio 2016):

> **Reasons-Sensitivity**: Having the kind of freedom or control required for moral responsibility is, at least in part, a matter of being sensitive to reasons in the right kind of way.
> And,
> **Actual Sequences**: The kind of freedom or control required for moral responsibility is exclusively a function of actual sequences (actual causal histories).

Actual-sequence accounts of moral responsibility are compatibilist accounts: holding that moral responsibility is compatible with the truth of determinism. As a result, those who defend such accounts need to formulate conditions for moral responsibility that encompass the appropriate sensitivity to reasons while not depending on the existence of genuine alternate possibilities (GAP).

Perhaps the most well-known attempt to formulate such conditions was made by Fischer and Ravizza (1998). They argued that for an agent to be open to moral responsibility for some outcome X the agent had to have *guidance control* over X, which requires two conditions: the *reason-responsiveness condition* and the *ownership condition*. The first of these conditions builds on the discussion of RR from the previous section. It requires that for an agent to be morally responsible for some outcome, X, that outcome must be the result an *actual-sequence mechanism* that is at least moderately reasons-receptive and at least weakly reasons-reactive. This means that the mechanism must be receptive to a *significant* number of reasons and must be reactive enough to adjust behaviour in the light of, and in accordance with, *at least some* of these reasons (Bratman, 2000: 454). Crucially, given that an actual-sequence account cannot require access to alternative possibilities, this responsiveness is to be understood counterfactually. Fischer and Ravizza argue that reason-responsiveness does not require that the actual-sequence mechanism had to be adjusted in the light of the actual reasons present, only that the mechanism be reactive to reasons to the degree where in a nearby possible world some (relevant) reason –though not the reason that was actually ignored– could have resulted in the behaviour being different from what actually happened. It is worth flagging here that there is an ongoing dispute concerning the exact specification of the scope of receptivity and reactivity, as the qualifications "moderate" and "weak" are very open to interpretation. Though this is undoubtedly an important matter, I will be bracketing this concern for the purposes of this paper.

It is important to note that the reasons under consideration here are *normative* reasons, and that the agent must be able to govern her conduct in reaction to these reasons qua their nature as normative reasons. This last point is well expressed in Levy (2011: 116) when he states that the agent must 'properly appreciate the significance of bringing about that state of affairs, where the significance of a state of affairs consists of the features which provide reasons for bringing it about (often, but not always, moral reasons)'. Importantly, this does not mean that the agent must necessarily have the belief that a given consideration in favour of acting is a normative reason to act, but that the agent recognises the "call to action", or the normative force of the features in question. To illustrate with a well-known example: consider a man standing on the side of a train track who sees another man falling between the tracks with a train oncoming. The first man leaps onto the tracks and pins the second man

down flat so that the oncoming train passes harmlessly overhead. The man who leaps in needs not, at the moment, have a belief with the content, "I have a moral reason to save that man" or even "I should save that man" in order for it to be the case that the outcome was under guidance control. What matters is that the agent was moved by the considerations that constitute the moral reason for action and was receptive and reactive to them.

The second condition is that of ownership. This requires that for an agent to be responsible she must (i) see herself as the source of her behaviour, in the sense that she sees that her motivations, beliefs, and intentions result in changes in the world, (ii) see herself as a legitimate target for moral responses on the basis of how she employs the agency just described, and (iii) the views in (i) and (ii) must be based on evidence. This condition is necessary to deal with so-called *manipulation cases*. In such cases, an agent is manipulated by some force in a way that alters the RR of her actual-sequence mechanism such that she fails to respond to a moral reason that she would have responded to without the manipulation. Subliminal messaging, for example, might subtly influence an agent's motivational set in such a way that her actual-sequence mechanism fails to respond to a moral reason that she would otherwise have responded to. Her conduct remains the result of a reason-responsive mechanism, but it seems wrong to hold her responsible for her failure in this case. The ownership condition seeks to block this challenge by requiring not only that the agent must consider her or himself to be the source of her or his conduct, but that this must be based on *evidence*.

This aspect of Fischer and Ravizza's argument –the ownership requirements for guidance control– is probably the aspect that has met with the most criticism. Alfred Mele (2006) argues that the subjective requirement makes the door for moral responsibility too narrow. It seems unproblematic to assume that in at least some cases an agent could be morally responsible even though she does not see herself as a legitimate target for reactive attitudes. Mele expresses this criticism using an example that runs like this: Imagine Sarah. Sarah is a committed hard incompatibilist, and so does not see herself as a legitimate target for reactive attitudes. Sarah, like most people, sometimes engages in telling lies to improve her own situation. Although it seems clear that we would find Sarah to be morally blameworthy for her lying, since Sarah does not believe herself to be a legitimate target for reactive attitudes, it appears that Fischer would have

to conclude that Sarah is not morally responsible for these lies. Similar examples, where one or the other of the subjective requirements is absent and leading to seemingly counterintuitive results, are not difficult to imagine.[5]

In the light of these shortcomings, the version of the AS account that I will be employing does away with the ownership condition. Instead I endorse the following modification of a principle put forward by Björnsson (2017: 149):

> DEMAND FROM CAPACITY: The degree to which it can be properly demanded that X is responsive to moral reasons depends only on X's capacity for so responding, not on how X came to the degree of reasons-responsiveness she does

This is not a condition for responsibility, but it does serve to answer manipulation cases. In any such case, the agent involved remains responsible provided that their conduct is the result of a sufficiently RR mechanism, regardless of how the responsiveness of the mechanism came about. However, the worry may still persist that this would mean that an agent who has been covertly manipulated to reason increasingly egoistically will be unfairly blamed for her conduct resulting from this change. However, since the *degree* to which an agent can be blamed or praised depends on the capacity for response, manipulations that result in reduced RR –such as in this example– will yield lower degrees of blame.[6] This will hopefully serve to mitigate any discomfort that arises from the verdict that such agents remain responsible.

With this version of the AS account in hand, let us consider what insights it can provide concerning moral bioenhancement.

[5] Pereboom (2007) advances a criticism using an example: Imagine Susan. Susan has been created by neuroscientists such that they are able to modify her reasoning so that she reasons egoistically when they flip a switch. She remains reasons-responsive, and she is unaware of this manipulation and so holds the necessary beliefs for ownership. The manipulation is of such an art that even knowledge of the causal history of her actual-sequence mechanism (which is not itself manipulated by the neuroscientists) would not provide her with evidence to relinquish these beliefs.

[6] This is an example of what McKenna (2013) calls the *hard-line response* to manipulation cases.

3. TWO DIMENSIONS OF INTERACTION

The AS account of responsibility allows us to investigate two different lines of enquiry concerning moral bioenhancement: (1) *Backward-looking responsibility*: we can chart the impact of moral bioenhancement on an enhanced agent's backward-looking responsibility – i.e. blameworthiness or praiseworthiness for her conduct. (2) *Forward-looking responsibility*: we can unpack our duties (or forward-looking responsibilities) as regards moral bioenhancement, whether such enhancement is permissible, whether there is a normative reason to pursue it, and whether it is obligatory. The two lines of investigation do cross paths at times, and where this occurs it will be noted.

BACKWARD-LOOKING RESPONSIBILITY

When moral responsibility is discussed, it is usually concerned with backward-looking responsibility, the attribution of blame or praise to a target in the aftermath of some morally valent event. Following the trend in the literature on such responsibility, I will be assuming that what it is to blame or praise an agent is to have certain justified *reactive attitudes* toward them.[7] To illustrate in the case of blame: these attitudes can be directed by the wronged to a perpetrator, such as in the case of moral anger, by the wronged toward themselves, as in guilt, or by a third party toward the perpetrator, as in indignation.[8] In line with the AS account outlined in the previous section, to be a justified target for reactive attitudes requires that you have guidance control over that on the basis of which you are being blamed or praised. As was mentioned in the introduction, one serious concern raised by moral bioenhancement is that it could

[7] This view was introduced by Strawson in his seminal work, *Freedom and Resentment* (1962). Though there is ongoing dispute about how exactly the reactive attitudes should be understood, Strawson's central thesis –that what it is to be responsible is to be the fitting target for such attitudes– has proven remarkably resilient.

[8] It is an ongoing debate whether or not blame or praise also plays a role in legitimating material sanctions (such as those imposed by the law) and/or revisions of our duties toward the blamed or praised agent, and one I will not be exploring here. Wringe (2012) and Scanlon (2015) are examples of arguments for the first and second positions respectively.

undermine or negate precisely this control link between agent and outcomes, thereby threatening our responsibility practices.

Current evidence indicates that, at least some of the time and in at least some cases, the interventions of moral bioenhancement have the potential to bring about morally better outcomes (Savulescu and Maslen, 2015). That said, we remain far from the point where we can effectively and reliably employ these interventions. Further, there are persistent and compelling concerns about how we are to determine what in fact counts as a "morally better outcome" (Lara and Decker, 2019). However, this is neither here nor there as it pertains to my argument in this piece. I wish to bracket out empirical and normative ethical concerns in order to focus solely on what we can learn about moral bioenhancement from its relationship to our responsibility practices. In the light of this, I will be making the following two assumptions:

> Assumption 1 (AS1): we can reliably identify morally better outcomes.
> Assumption 2 (AS2): our biomedical interventions aimed at the moral enhancement of humans are reliable in achieving their aims.

Having cooked up all these ingredients, what is the implication of blame and praise? Given that only those interventions that promote an agent's RR are in fact examples of moral bioenhancement, the reasons-responsiveness condition of the AS account can, by definition, not be undermined by such intervention. As such, if an agent was a legitimate potential candidate for moral blame or praise before the enhancement, then at least according to this condition they would remain so in the aftermath. What is more slippery is the case where an agent was previously below the necessarily level of RR to be open to blameworthiness or praiseworthiness – as may be the case for some suffering from psychological impairments or cognitive handicaps. In such a case, *it is possible but not necessary* that the intervention may raise the agent's RR across the threshold necessary for responsibility. A possible example of this in practice may be the use of neural feedback to treat psychopathy via assisting in the rapid training of emotions (Sitaram *et al.*, 2007).[9] Critically,

[9] As our emotions play a role in determining our motivational sets, and these in turn influence both which states of affairs we take to count as reasons for action as well as how receptive and reactive we are to these reasons, then the promotion of the right

interventions that serve to undermine or circumvent an agent's RR would risk reducing the agent's openness to blameworthiness or praiseworthiness. However, no such intervention could be plausibly considered an example of *moral* bioenhancement. After all, if our aim in enhancement is to produce morally better agents, then it is necessary that the enhanced remain moral agents. We could, for example, intervene in our pets in order to make them better at maximising utility, yet this would not be a moral bioenhancement of our pets – unless of course the intervention was of such an art so as to provide our pets with moral reasons-responsiveness!

DEMAND FROM CAPACITY will mean that the degree of RR demanded of enhanced agents will change to reflect their new capacities. The effect of this in terms of blame is that individuals are likely to find themselves accountable for a greater number of possible failures, as the scope of moral reasons that they can be expected to respond to will have increased. Of course, since this goes hand in hand with an improved responsiveness, the overall result may still be that the enhanced experiences net less blame. However, since this is a discussion concerning a capacity that may or may not be exercised in any single case, it is not possible to make wholesale judgements on whether overall blame will increase or decrease. Where praise is concerned there will be some instances where an enhanced agent will not receive praise when they would have had she not been enhanced. Praise is merited when an agent responds to moral reasons that she could not have been demanded to respond to, and this can come by degrees. That praise becomes somewhat more difficult to come by might seem initially unfair, but it is exactly how all moral education and improvement works. There is nothing unique or interesting about the case of moral bioenhancement in this respect.

As discussed in Section 2, the version of the AS account that I am employing here excludes Fischer and Ravizza's ownership condition. This being the case, we may worry about examples such as the following one taken from Mele (1995: 145):

> Ann is an autonomous agent and an exceptionally industrious philosopher. She puts in twelve solid hours a day, seven days a week; and she enjoys almost every minute of it. Beth, an equally talented colleague, values a great many things above philosophy, for reasons that she has refined and endorsed on the

emotions (such as empathy in the case of psychopaths for example) will likely serve to improve an agent's RR.

> basis of careful critical reflection over many years. She identifies with and enjoys her own way of life – one which, she is confident, has a breadth, depth, and richness that long days in the office would destroy. Their dean (who will remain nameless) wants Beth to be like Ann. Normal modes of persuasion having failed, he decides to circumvent Beth's agency. Without the knowledge of either philosopher, he hires a team of psychologists to determine what makes Ann tick and a team of new-wave brainwashers to make Beth like Ann. The psychologists decide that Ann's peculiar hierarchy of values accounts for her productivity, and the brainwashers instil the same hierarchy in Beth while eradicating all competing values – via new-wave brainwashing, of course. Beth is now, in the relevant respect, a "psychological twin" of Ann. She is an industrious philosopher who thoroughly enjoys and highly values her philosophical work. Indeed, it turns out – largely as a result of Beth's new hierarchy of values – that whatever upshot Ann's critical reflection about her own values and priorities would have, the same is true of critical reflection by Beth. Her critical reflection, like Ann's, fully supports her new style of life.

Mele uses this case in order to make a point about the relative *autonomy* of Ann and Beth, autonomy here meaning the degree of self-governance of both Ann and Beth. For Mele, Ann is autonomous and self-governing, whereas Beth is not. However, the implications for responsibility seem clear enough: Beth seems to have reduced responsibility on account of the intervention. An ownership condition would be one way of capturing this intuition, and this strategy is part of the broader family of approaches to dealing with manipulation cases like this by taking responsibility to be a historical notion (i.e. *how* an agent comes to have a certain RR matters for whether or not she is responsible). In contrast, the account as I have outlined it instead claims the opposite: how an agent comes by a given RR is irrelevant to responsibility.

Let us consider the case of Ann and Beth in a bit more detail. The first thing to note is that the intervention in this case is certainly not an instance of moral bioenhancement, as though the intervention is biomedical in the wider (and more science fiction-y) sense it does not aim to make Beth a more moral agent. Furthermore, it is difficult to know exactly through what process the manipulators were able to make the changes to Beth's value system. A common philosophical view about values –and one that Mele appears to explicitly endorse in his example– is that they are those motivations toward which we have second order attitudes of affirmation and endorsement – where the process of forming such second order at-

titudes is itself a reasons-responsive process (Watson, 2004) or a matter of critical reflection (Mele, 1995). If this is the case, then manipulating values directly would amount to a circumvention of the agent's RR mechanism. Beth would thus not be responsible for the formation of the given value, as it did not result from her RR mechanism. The effect of this would be to reduce the degree of responsibility Beth would have for a post-intervention outcome in proportion to the contribution of the changed value to the outcome.

This is no different in principle to how we might think of an extended action with multiple subactions. Say I am making a cake and have to prepare and add three ingredients (it's a very simple cake). If during the preparation of one of these ingredients –let us say sugar– you were to grab my arm and force me to add salt, then the saltiness of the final product would not be fairly attributable to me, though I do still bear responsibility for the rest of the taste of the cake. Importantly, this does not apply to things that may be changed that do not normally result from an agent's RR mechanism, such as motivations, emotions, etc. If the change in values is brought about indirectly by changing these unreflective qualities, then Beth remains on the hook.

Even if the intervention is not so direct, the amplification, reordering, and erasing of values undertaken by the employers could quite likely render Beth less RR than she was before – the scope of moral reasons that she is receptive and reactive to has likely diminished. If this is the case, she would have diminished responsibility as her capacity has been reduced, in line with DEMAND FROM CAPACITY. But let us imagine that it was a case where Ann is a very moral person, and Beth's employers undertake the same procedures with an aim to make Beth more like Ann in just this sense. In this case, whether or not Beth will have diminished responsibility hinges on whether the changes made by the employers have left her responsive to a sufficiently wide scope of reasons for moderate reasons-responsiveness. If they have not, then she would again be off the hook. If they have, then Beth would indeed be open to responsibility for her conduct in this case, and rightly so.

Of course, her manipulators would also be responsible for their conduct in intervening, and indirectly and to a degree for her resultant conduct. Whether we think that they would be blameworthy or praiseworthy for intervening will depend on our answer to the question, "when is it permissible to morally bioenhance an agent?" It is precisely to this question that I now turn.

FORWARD-LOOKING RESPONSIBILITY

Forward-looking responsibilities are duties. If I were a surgeon preparing for a surgery, I would have a forward-looking responsibility to read my patients case file in order to ensure I know the facts of the case. Arguably we all, as members of the moral community, have a forward-looking responsibility to extend reasonable care toward all other members. Clearly this can at times be linked to backward-looking responsibility: failures to meet forward-looking responsibilities are often grounds for backward-looking responsibility. Perhaps less obviously, being backward-looking responsible can have implications for your forward-looking duties: if I break a promise made to my partner, I have a duty to make amends by apologising, taking remedial steps, etc. I will not take a stand in this piece on whether all cases of backward-looking responsibility have a forward-looking counterpart.

As regards moral bioenhancement, a crucial facet of the discussion revolves precisely around what our forward-looking responsibilities are. To arrive at an answer to this, let us start not with the assumption that there is a *pro tanto* moral reason for us to promote moral agency where we can. This involves the promotion of RR in ourselves and others, via various means. Probably the most readily apparent example is in the case of children, where we actively attempt to shepherd them into adequately recognising and reacting to moral reasons. But this reason is also at work in our dealings with other adults. It is part of the support for pursuing rehabilitation for moral wrongdoers and undergirds many of our responsibility practices – we blame and praise in part to promote RR. Given its *pro tanto* nature, this reason can certainly be defeated by other considerations at times, perhaps often. Let us call this *pro tanto* reason to promote RR *reason1*. Acting on reason1 can be a morally laudable act, and failing to respond to it appropriately can be a basis for blame. More to the point, reason1 clearly lends support to moral bioenhancement, provided that the latter is in fact moral enhancement according to the requirements previously discussed.

Viewed purely through its relationship to responsibility, provided that AS1 and AS2 are met, and that moral bioenhancement is understood as I have defined it in Section 1, such enhancement is permissible. A consideration of our responsibility practices raises no reason against it. Stronger than that, given that we have a *pro tanto* reason to promote RR in ourselves and others, enhancement is *normatively desirable* – we *pro*

tanto should enhance if the opportunity exists. Again, these are not overriding reasons, and there can be a plethora of considerations that can outweigh them when the choice is before an agent. But all else equal, responsibility is tipping the scales in favour of RR promoting bioenhancement.

So far, I have been considering what reasons we may have to pursue moral bioenhancement voluntarily. However, this is not the only way for which it has been argued, and reason1 is (by an exceedingly long way) not the only reason that has been adduced in support of moral bioenhancement. Persson and Savulescu (2008; 2013) have argued that another reason to pursue enhancement is to stave off the worst possible outcome: making worthwhile life on this planet untenable. Let us call this *reason2*. Of course, reason2 is merely the most extreme version of a more general reason in support of enhancement: to reduce or minimise harm. If there are two possible worlds that we could find ourselves living on, one in which enhancement takes place (E) and one in which it does not (NE), and it follows *because* of this difference that E has less harm than NE, this gives us a plausible reason to enhance. From here, the greater the difference in harm between E and NE (assuming that E always has less), the stronger the reason to enhance. Reason2 is this reason set to maximum. On the basis of this, Persson and Savulescu argue that enhancement can be justifiably seen as *obligatory*.[10] Assuming that AS1 and AS2 hold, this is a compelling argument, and is relevant to what we may want to say about Beth. Let us for a moment take a weaker version of reason2, *reason2**, where E has less harm than NE, but the increase is marginal. Additionally, let us imagine that Beth's case is such that she has been covertly intervened with by her employers with the aim of moral improvement, and that this has been done by promoting, not undermining or circumventing, her RR. Were her employers justified? If reason2* legitimates obligatory enhancement, then the answer would be yes. The reason the answer is no is that the obligation to promote moral improvement is weak and does not outweigh considerations of autonomy. We can infer this from the ease with which it is defeated in everyday life. Once a person reaches adulthood, it usually requires a high degree of harm to be at stake in order to justify the application of mandatory meas-

[10] Others have argued, though for different reasons, that moral bioenhancement should be mandatory treatment for psychopaths (Baccarini and Malatesti, 2017). Though I find their argument problematic, I leave it aside in this work.

ures aimed at moral improvement. The moral value of autonomy is clearly weighty, and usually provides sufficient reason not to intervene. Now, the folk could of course be wrong in their practice, and perhaps we should be quicker to intervene in the lives of others in order to bring about improved RR, but this is then no longer a discussion pertaining especially to moral bioenhancement but it is in fact a general call for increased paternalism. If we turn to reason2 itself, the sheer scale of the possible harm does indeed seem sufficient to justify taking severe measures in order to ensure any survival at all. There is then a sorites situation between reason2 and reason2*, and it may not be possible to provide a clear-cut distinction. At the very least, a careful consideration of moral responsibility does not provide us with a means to do so.

CONCLUDING REMARKS

In this paper I investigated the relationship between moral bioenhancement and the actual-sequence account of moral responsibility. I provided definitions of the notions of moral bioenhancement and moral responsibility that I use in this paper and addressed some of their shortcomings and offered improvements. Thereafter, I explored two dimensions of the relationship between them: backward-looking and forward-looking responsibility. On the backward-looking path I argued that since only bioenhancement that promotes an agent's reasons-responsiveness to moral reasons can be justifiably considered to be an example of moral bioenhancement, an enhanced individual would remain open to responsibility if they were before, may become open to responsibility if they were not before, and the degree of responsiveness that can be demanded of the agent will increase with increases in capacity for such responsiveness. Along the forward-looking track, I argued that moral bioenhancement is permissible, *ceteris paribus*, and normatively desirable given there is a *pro tanto* reason to pursue promotions of reasons-responsiveness. Lastly, it was shown that the demands of moral responsibility do not –on their own terms– rule out the possibility that moral bioenhancement can sometimes be obligatory.

BIBLIOGRAPHY

Anscombe, G. E. M. 1963. *Intention* (second edition). United States of America: Harvard University Press.

Arpaly, N. & Schroeder, T. 2013. *In Praise of Desire*. Oxford Scholarship Online.

Fischer, J. M., & Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Baccarini, E., & Malatesti, L. 2017. 'The moral bioenhancement of psychopaths.' *Journal of Medical Ethics*. 43: 697-701.

Björnsson, G. 2017. 'Explaining (Away) the Epistemic Condition on Moral Responsibility'

in Robichaud, P., & Wieland, J. W. (eds.) 2017. *Responsibility: The Epistemic Condition*. Oxford Scholarship Online.

Bratman, M. 2000. 'Fischer and Ravizza on Moral Responsibility and History.' *Philosophy and Phenomenological Research*. 61(2): 453-458.

Crutchfield, P. 2016. 'The Epistemology of Moral Bioenhancement.' *Bioethics*. 30(6): 389-396.

DeGrazia D. 2014. 'Moral enhancement, freedom, and what we (should) value in moral behaviour.' *Journal of Medical Ethics*. 40: 361-368.

Douglas, T. 2008. 'Moral enhancement.' *Journal of Applied Philosophy*. 25: 228-245.

Frankfurt, H. 1988. *The importance of what we care about*. Cambridge/ New

York/Melbourne: Cambridge University Press.

Jotterand, F. 2011. ' "Virtue engineering" and moral agency: will post-humans still need the virtues?' *AJOB Neuroscience*. 2(4):3-9.

Kane, R. 1996. *The Significance of Free Will*. New York: Oxford University Press.

Lara, F. & Deckers, J. 2019. 'Artificial Intelligence as a Socratic Assistant for Moral Enhancement.' *Neuroethics*. doi: 10.1007/s12152-019-09401-y.

Levy, N. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford Scholarship Online.

McKenna, M. 2012. 'Directed Blame and Conversation' in Coates, J. D. and Tognazzini, N. A. (eds.) 2012. *Blame: Its Nature and Norms*. Oxford Scholarship Online.

McKenna, M. 2013. 'Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin.' *Philosophy and Phenomenological Research*. 89(2): 467-486.

Mele, A. R. 1995. *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.

Mele, A. R. 2006. 'Fischer and Ravizza on Moral Responsibility.' *The Journal of Ethics*. 10(3): 283-294.

Pereboom, D. 2007. 'Hard Incompatibilism', in Fischer, J. M., Kane, R., Pereboom, D. & Vargas, M. 2007. *Four Views on Free Will*. Singapore: Blackwell Publishing.

Pereboom, D. 2013. 'Free Will Skepticism, Blame, and Obligation', in Tognazzini, N. and Coates, D. J. (eds.) 2013. *Blame: Its Nature and Norms*. New York: Oxford University Press.

Persson, I., & Savulescu, J. 2008. 'The perils of cognitive enhancement and the urgent imperative to enhance the moral character of humanity.' *Journal of Applied Ethics*. 2008. 25: 174.

Persson, I., & Savulescu, J. 2013. 'Getting moral enhancement right: the desirability of moral bioenhancement.' *Bioethics*. 27: 124-31.

Raus, K., Focquaert, F., Schermer, M., Specker, J., & Sterckx, S. 'On defining moral enhancement: a clarificatory taxonomy.' *Neuroethics*. doi: 10.1007% 2Fs12152-014 9205-4.

Rosen, G. 2014. 'Culpability, Duress and Excuses.' *Proceedings of the Aristotelian Society*. 88: 69- 90.

Sartorio, C. 2016. 'A Partial Defense of the Actual-Sequence Model of Freedom.' *Ethics*. 20: 107-120.

Savulescu, J., & Maslen, H. 2015. 'Moral enhancement and artificial intelligence: Moral AI?', in Romportl, J., Zackova, E., & Kelemen J., (eds.) 2015. *Beyond artificial intelligence. The disappearing human-machine divide*. Springer.

Scanlon, T. M. 2015. 'Forms and Conditions of Responsibility', in Clarke, R., McKenna, M., & Smith, A. M., (eds.) 2015. *The Nature of Moral Responsibility: New Essays*. Oxford Scholarship Online.

Shoemaker, D. 2015. *Responsibility from the Margins*. Oxford Scholarship Online.

Shook, J. R. 'Neuroethics and the possible types of moral enhancement.' *AJOB Neuroscience*. 3(4): 3-14.

Simkulet, W. 2012. 'On moral enhancement.' *AJOB Neuroscience*. 3(4): 17-18.

Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., Kuebler, A., & Birbaumer, N. 2007. 'fMRI brain-computer interface: A tool for neuroscientific research and treatment.' *Computational Intelligence and Neuroscience*. 1(1):10.

Strawson, H. F. 1962. 'Freedom and Resentment.' *Proceedings of the British Academy*. 48: 187-211.

Vargas, M. 2013. 'How to Solve the Problem of Free Will', in Russel, P., & Deery, O., (eds.) 2013. *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. Oxford University Press.

Watson, G. 2004. *Agency and Answerability*. Oxford: Oxford University Press.

Wringe, B. 2012. 'Pre-Punishment, Communicative Theories of Punishment, and Compatibilism.' *Pacific Philosophical Quarterly*. 93: 125-136.

Niël Conradie
RWTH Aachen University
niel.conradie@humtec.rwth-aachen.de

Submission: September 15th 2019
Acceptance: November 21st 2019