

# Moral Worth, Biomedical Moral Enhancement and Communicative Punishment

ELIZABETH SHAW\*

---

## **Abstract**

*Imagine that a criminal offender were provided with a 'reform pill', which significantly weakened his desire to reoffend. After consuming the pill, he obeys the law. Is the offender's subsequent, apparently 'good' behaviour genuinely good?*

*Various theorists have intuited that biomedical 'moral enhancement' techniques used for the purposes of reducing reoffending could somehow undermine the moral worth of the recipient's future actions. This article draws on the communication theory of punishment in order to shed new light on a potential source of this intuition in relation to the moral bioenhancement of those who have committed serious criminal wrongs. In doing so, it will consider the contention that the ultimate source of this intuition can be attributed to the intrinsically valuable freedom to do wrong before rejecting this account. The article then proceeds to explore the implications of the communication theory of punishment for the question of whether biomedical moral enhancements would undermine the moral worth of offenders' future law-abiding behaviour and highlights the need for the proponents of such interventions to address these issues. The arguments presented in this article have potential implications for biomedical interventions that are currently being used in the criminal justice system.*

## **1 Introduction**

Imagine that a criminal offender were provided with a 'reform pill', which significantly weakened his desire to reoffend. After consuming the pill, he obeys the law. Is the offender's subsequent, apparently 'good' behaviour genuinely good? Various theorists have intuited that such biomedical "moral enhancement" techniques could somehow undermine the moral worth of the recipient's future actions. I will draw on the communication theory of punishment in order to shed new light on a potential source of this intuition, in relation to the moral bioenhancement of those who have committed serious criminal wrongs. The arguments presented in this article have potential implications for biomedical interventions that are currently being used in the criminal justice system, eg anti-libidinal pharmaceuticals and interventions to

---

\* Dr. Elizabeth Shaw is a lecturer at the University of Aberdeen School of Law. [eshaw@abdn.ac.uk](mailto:eshaw@abdn.ac.uk). This article reflects the law and technological environment as at the date of approval for online publication on 14 November 2017.

reduce substance abuse, as well as to interventions that may be used in the future, such as pharmaceuticals to reduce aggression.

Before setting out the moral worth argument, I will consider the objection that the ultimate source of the intuitive suspicion of moral bioenhancement is its potential to compromise freedom, rather than moral worth. Specifically, it might be thought that the freedom to do wrong is valuable in itself and that biomedical interventions might interfere with this freedom. I will agree with John Danaher's rejection of the idea that the freedom to do wrong is intrinsically valuable, but will add that it might be instrumentally valuable, insofar as it may be one of the preconditions for an action to have moral worth. I will suggest that the capacity to perform actions with moral worth is something that we should value and preserve. In section 2, I will provide some reasons for thinking that the state does not have the right to deprive offenders of the capacity to perform actions with moral worth, as this would constitute degrading treatment. In section 3, I will provide reasons stemming from the communication theory of punishment for caring about whether biomedical moral enhancement of offenders would undermine the praiseworthiness of offenders' law-abiding behaviour post-enhancement. In the light of the communication theory, I will consider the idea that the effort required for actions to have moral worth could be undermined by biomedical interventions. I will suggest the possibility that effort may be a precondition for moral worth in certain contexts — specifically, in cases where the agent has previously committed a seriously immoral and criminal action. The communication theory arguably implies that in such cases, the reformed offender's subsequent avoidance of committing similar crimes will only have moral worth if his improved behaviour has a certain causal history — it should emerge from an effortful process of repentance of the original crime. Thus my account of the relationship between effort and moral worth is historical and context-dependent. This way of explaining the connection between effort and moral worth, in the context of biomedical interventions, has not yet received detailed consideration. In the remainder of the article I will distinguish between three different types of biomedical intervention, which I will call 'pure capacity enhancements', 'inclination-altering interventions', and 'non-reasons-responsive interventions'. I will consider whether each type of intervention would undermine the moral worth of offenders' actions on the communication theory. I will conclude that a communication theorist could accept the consensual use of such interventions as a) a means of facilitating repentance through moral communication,<sup>1</sup> and b) a means of helping to bring about law-abiding

---

<sup>1</sup> Elizabeth Shaw, 'Direct Brain Interventions and Responsibility Enhancement' (2014) 8(1) *Criminal Law and Philosophy* 1 discusses the possibility that biomedical interventions might be given to offenders to facilitate the kind of moral dialogue and reform required by Duff's communication theory. William Bulow, "It Will Help You Repent": Why the Communicative Theory of Punishment Requires the Provision of Medications to Offenders with ADHD' in Nicole A Vincent (ed), *Neuro-Interventions and The Law: Regulating Human Mental Capacity* (Oxford University Press, 2020) has also argued that, on Duff's communication theory, offenders with ADHD should be offered

behaviour, provided that the offender has already undergone communicative punishment and that he is motivated to accept the intervention by his repentant desire to refrain from crime. On the communication theory of punishment, it would never be acceptable to use biomedical interventions to replace moral communication aimed at inducing repentance.

This conclusion has important practical implications, because it has the potential seriously to weaken the motivation for using such interventions within the criminal justice system. It has been claimed that biomedical interventions are likely to be 'politically attractive', if they can '(partially) replace' very financially costly forms of traditional punishment.<sup>2</sup> However, communication theorists would oppose using biomedical interventions as a way of (even partially) replacing forms of punishment that involve moral dialogue. Communication theorists would insist that, if the state were to employ biomedical interventions to rehabilitate offenders, it must bear the financial cost of these interventions in addition to the financial cost of communicative punishment. It has also been suggested that biomedical interventions could be more humane — involving less suffering and allowing the offender to return to his normal life more quickly — than traditional forms of punishment. While communication theorists, such as Antony Duff, criticise our current penal practices for inflicting much unnecessary suffering (including excessively lengthy prison terms), nevertheless such theorists insist that punishment of serious crimes must involve suffering, as repentance requires much time, effort and emotional pain. If, as communication theory implies, biomedical interventions should only be used, if at all, alongside communicative punishment, then such interventions must not be employed in order to 'speed up' the process or to bypass the suffering, which according to this theory, is a necessary part of punishment.

My intention is not to defend the role of repentance in communication theories of punishment. Rather, my more modest aim is to explore the implications of this highly influential view for the question of whether biomedical moral enhancement would undermine the moral worth of offenders' future law-abiding behaviour and to argue that proponents of such interventions should address this issue.

---

pharmaceutical treatment for this disorder in order to facilitate repentance and reform. Bulow focuses specifically on the treatment of ADHD with methylphenidate. However, the current article focuses on moral enhancement of offenders who would not necessarily be diagnosed with any disorder, and discusses a wider range of possible interventions, including interventions that would render the recipient not responsive to reasons. Furthermore, neither of the above-mentioned articles explicitly discussed the argument that certain uses of biomedical interventions would intuitively undermine the moral worth of offenders' future conduct, because they would allow offenders to avoid repentance for their earlier crimes.

<sup>2</sup> Thomas Douglas and David Birks, 'Introduction' in David Birks and Thomas Douglas (eds), *Treatment for Crime* (Oxford University Press, 2018).

In this article, I will draw on Tom Douglas's distinction between 'moral conformity' and 'moral worth'.<sup>3</sup> An agent acts in a way that conforms to morality, insofar as her overt behaviour is supported by moral reasons (ie her overt behaviour is consistent with moral norms) even if she is not actually motivated by those moral reasons. 'An agent fully conforms to morality on a given occasion when she performs an act that is at least as well supported by moral reasons as any alternative act.'<sup>4</sup> An action might fully conform to morality, but lack any moral worth, eg because it is not motivated by the right reasons. 'Morally worthy conduct is conduct that merits praise.'<sup>5</sup>

## 2 Freedom and Moral Worth

Michael Hauskeller is one of the most prominent defenders of the idea that the 'freedom to do evil might be worth protecting from the grasp of moral bioenhancement'. He uses a striking fictional example of a violent criminal, called 'Little Alex', from Antony Burgess's novel, *A Clockwork Orange*.<sup>6</sup> While in prison, Alex receives an intervention called 'Ludovico's technique', which involves Pavlovian conditioning. The intervention prevents Alex from reoffending by causing him to become nauseous whenever he thinks of violence. Hauskeller invites us to share the intuition (voiced by the prison chaplain in the novel) that it is wrong to try to bring about moral conformity, by depriving someone of the freedom to do evil. This intuition, according to Hauskeller, should lead us to accept the paradoxical conclusion that sometimes 'the bad is better than the good',<sup>7</sup> or, in the words of the prison chaplain 'a man who chooses the bad [is] perhaps in some way better than a man who has the good imposed upon him'.<sup>8</sup> Hauskeller claims that a world with moral freedom (ie free will, which necessarily includes the freedom to do evil) and with some evil deeds in it, is better than a world with more good outcomes but no moral freedom.

---

<sup>3</sup> Thomas Douglas, 'Enhancing Moral Conformity and Enhancing Moral Worth' (2014) 7 *Neuroethics* 75.

<sup>4</sup> Ibid 75.

<sup>5</sup> Ibid 81.

<sup>6</sup> Antony Burgess, *A Clockwork Orange* (Penguin, 2007).

<sup>7</sup> Michael Hauskeller, 'Is It Desirable to Be Able to Do the Undesirable? Moral Bioenhancement and the Little Alex Problem' (2017) 26 *Cambridge Quarterly of Healthcare Ethics* 365, 368.

<sup>8</sup> Burgess (n 6) 81.

John Danaher has argued that Hauskeller's argument presupposes that moral freedom is an intrinsic good.<sup>9</sup> According to Danaher, if you are going to consistently favour possible worlds that contain both evil and moral freedom over possible worlds with many morally good outcomes, but no moral freedom, you have to believe that there is some 'intrinsic moral magic' to moral freedom. However, Danaher persuasively argues that moral freedom is not intrinsically good. Rather, freedom is an 'axiological catalyst'. It is something that makes good deeds better and bad deeds worse. To support this claim he compares the following examples: a murderer, A, freely murders ten people and an unfree individual, B, (whose freedom is undermined by a brain tumour or by duress) kills ten people. According to Danaher, A is 'obviously' much worse, because there is no excusing factor in that case. This suggests that, 'moral freedom makes a bad deed much worse.'<sup>10</sup> To illustrate the corresponding claim that freedom makes good outcomes better, Danaher imagines a millionaire freely donating money to charity, which seems preferable to a millionaire whose accountant accidentally transfers the money to charity.

Danaher claims that, if freedom lacks intrinsic value, this significantly undercuts Hauskeller's objection against biomedical moral enhancement — the objection that it is wrong to deprive individuals of the freedom to do evil. Danaher writes,

If moral enhancement techniques work as advertised, they would simply be preventing moral freedom from being exercised in wicked ways. They would, consequently, be making the world a better place, without taking away something that is intrinsically good.<sup>11</sup>

If freedom lacks intrinsic value, then this seems to undermine *some* of Hauskeller's claims. In particular, it can be pointed out that if Danaher is right that B, the unfree harm-doer, is better than A, the free harm-doer, then it cannot be true that 'a man who [freely] chooses the bad is...better than a man who has the good imposed upon him'.<sup>12</sup> This can be demonstrated by comparing C, the unfree person who produces good outcomes, because he has had the good imposed upon him, with B, the unfree harm-doer. Surely C must be better than B, since both of them lack freedom and the only difference between them is that C produces good outcomes and B produces bad outcomes. If C is better than B and B is better than A — the free harm-doer — then A cannot be better than C.

Danaher's argument that freedom is not an intrinsic good is very persuasive. However, even if we accept this conclusion, we could still agree with Hauskeller's main claim that, 'the freedom to do evil might be worth protecting

---

<sup>9</sup> John Danaher, 'Moral Enhancement and Moral Freedom: A Critique of the Little Alex Problem' in Michael Hauskeller and Lewis Coyne (eds), *Moral Enhancement: Critical Perspectives* (Cambridge University Press, forthcoming) 6.

<sup>10</sup> Ibid 11.

<sup>11</sup> Ibid.

<sup>12</sup> Burgess (n 6) 81.

from the grasp of moral bioenhancement'. This is because the freedom to do evil might be necessary for truly praiseworthy actions to be possible, ie perhaps an agent's conformity to moral norms only has moral worth if the agent was capable of acting badly. If we think that praiseworthy actions are highly valuable, then freedom to do evil might be of great indirect, instrumental value and might therefore worth be preserving. Hauskeller could make the claim that the *capacity* to do bad, is instrumentally valuable (because it is necessary for moral worth), without making the implausible claim that *actually* freely performing bad acts is intrinsically valuable and intrinsically better than unfree good deeds.

If my argument is correct, then those opponents of biomedical moral enhancement who currently focus on the value of moral freedom *in itself*, should instead shift their attention to considering the following questions:

- Does the capacity to perform truly praiseworthy actions have so much intrinsic value that biomedical moral enhancements that undermine this capacity should be ruled out, even if these enhancements would lead to much greater moral conformity?
- Is moral freedom really necessary for moral worth and would it be undermined by biomedical moral enhancements?
- What other pre-requisites for moral worth might be undermined by biomedical moral enhancements?

Regarding the first question: in section 2, I will briefly provide some reasons for thinking that the capacity to perform praiseworthy acts is valuable. In section 3, I will argue that, from the perspective of the communication theory of punishment, the state does not have the right to use biomedical interventions that would undermine offenders' capacity to perform acts with moral worth once their punishment is complete.<sup>13</sup> In Section 3 and in the subsequent sections, I will focus mainly on the third question – the other pre-requisites for virtue, apart from freedom. I will focus in particular on the importance of effort for moral worth. This question has received attention from various writers. However, these writers have discussed this question at a general level, and have not, I will suggest, paid sufficient attention to various contextual factors. It is worth exploring the possibility that effort is *sometimes* required for the moral worth of actions, depending on the context. I will suggest that pre-conditions for a good, law-abiding action to have moral worth may differ depending on whether or not the actor has previously committed a seriously immoral (and criminal) act. In

---

<sup>13</sup> I am not saying that ex-offenders' capacity to perform praiseworthy actions is so valuable that a world where they have this capacity, but frequently fail to exercise it, is necessarily better than a world where they lack this capacity but conform to moral norms. I am just arguing that the state does not have *the right* to deprive them of the capacity to perform actions with moral worth, and should in fact seek to promote this capacity.

making this argument, I will draw on the communication theory of punishment. Theorists are just beginning to apply penal theoretic considerations to the topic of the moral bioenhancement of offenders. However, so far, they have not applied theories of punishment to the topic of whether biomedical interventions could (objectionably) allow offenders to avoid the effort required for their future actions to have moral worth.

### 3 *The Value of the Capacity to Perform Praiseworthy Actions*

In this section I will provide some reason for thinking 1) that it is objectionable for the state to deprive an offender of a valuable capacity and 2) that the capacity to perform actions that deserve moral praise is valuable.

First, consider the example of impairing someone's intellectual capacities. It seems intuitively objectionable for the state to deliberately impair an offender's cognitive capacities significantly as a form of 'rehabilitation'. For instance, imagine that impairing a fraudster's cognitive abilities would prevent him from carrying out complex financial frauds in future. This intervention seems ethically unacceptable. Intelligence is a valuable capacity (even though it can be misused). Furthermore, an individual's cognitive abilities might be regarded as central to who she is. An attack on something that is so important to a person's identity should be prohibited.

It might be objected that whether such an intervention is unacceptable is a matter of degree. If an offender has an above-average IQ would it still be impermissible to knock off just a few points in order to facilitate rehabilitation? In response, it is important to note that impairing intelligence could only have a significant effect on the offender's propensity to commit fraud if it deprived him of an ability to exercise certain *intellectual skills*. This is likely to have more wide-ranging effects than simply preventing the offender from committing fraud. It will also prevent him from engaging in innocent activities that require a similar degree of cognitive ability. It will prevent the offender from having access to various domains of knowledge.

Similar things could be said about the value of moral praiseworthiness and the wrongfulness of impairing the capacity to merit praise. Indeed, studies suggest that people view moral capacities as being more central to personal identity than intellectual capacities. For instance, Nina Strohminger and Shaun Nichols write that, '[m]oral traits ... are considered the most essential part of identity, the self, and the soul.'<sup>14</sup>

---

<sup>14</sup> Nina Strohminger and Shaun Nichols, 'The Essential Moral Self' (2014) 131(1) *Cognition* 159.

Even if it were accepted that the capacity to perform praiseworthy acts is valuable, it might be argued that the badness of undermining this capacity could be outweighed by competing considerations.

In response, it could be argued that destroying or significantly impairing valuable capacities, such as intellectual ability or the capacity to perform praiseworthy actions would be classed as inhuman and degrading treatment, which is prohibited under article 3 of the *European Convention On Human Rights*. This is usually regarded as an absolute prohibition, which cannot be overridden. Elaine Webster has argued that restricting autonomy is capable of constituting degrading treatment.<sup>15</sup> She refers to the case of *Ireland v UK*, in which the European Commission of Human Rights explicitly referred to the value of 'independence of the will' when considering whether ill-treatment of internees by UK security forces was degrading. The Commission stated that, '[t]he will to resist or to give in cannot, under such conditions, be formed with any degree of independence. Those most firmly resistant might give in at an early stage when subjected to this sophisticated method to break or even eliminate the will.'<sup>16</sup> The European Court of Human Rights in the same case also stated that the treatment in question was degrading since it was, among other things, capable of breaking the victim's 'physical or moral resistance.'<sup>17</sup> Autonomy and the will are closely connected to the capacity to perform actions with moral worth. They are necessary conditions for that capacity. So it seems plausible to conclude that significantly impairing the capacity to perform praiseworthy actions would constitute degrading treatment for similar reasons.

Much more could be said about this topic. However, for the purpose of this article, I only aim to motivate the idea that there are grounds for objecting to moral enhancements that interfere with the capacity to perform acts with moral worth.

#### 4 Moral Worth and Communicative Punishment

There are reasons stemming from penal theory in particular for caring about whether biomedical moral enhancement of offenders would undermine the moral worth of offenders' actions post-enhancement. I will focus on the communicative theory of criminal punishment defended by Antony Duff – one of the most influential and persuasive accounts of the nature and justification of punishment.<sup>18</sup> This theory presupposes the existence of a political community, in

---

<sup>15</sup> Elaine Webster, 'Degradation: A Human Rights Perspective' in Paulus Kaufmann, Hannes Kuch, Christian Neuhauser and Elaine Webster (eds), *Humiliation, Degradation, Dehumanization: Human Dignity Violated* (Springer, 2011).

<sup>16</sup> *Ireland v United Kingdom* (1976) 31 Eur Comm HR 402.

<sup>17</sup> *Ireland v United Kingdom* (1978) 25 Eur Court HR (ser A) [167].

<sup>18</sup> RA Duff, *Punishment, Communication and Community* (Oxford University Press, 2001).



which members are expected show a mutual concern for each other. That mutual concern involves 'a readiness to assist one another in pursuing and preserving the community's distinctive goods' and precludes, for example, 'simply exploiting others for one's own ends'.<sup>19</sup> Members of the community also owe each other 'civic trust' that their fellow citizens will not seriously breach their obligation to treat them with concern and respect.<sup>20</sup> Civic trust is not a not a 'naïve state of blinkered optimism', but a recognition of others as fellow members of the community, rather than 'enemies' against whom one must guard oneself.<sup>21</sup> According to Duff, the institution of punishment must 'foster... and respect ... the community's defining values, and must treat and address citizens in ways that embody those values.'<sup>22</sup> One way in which the criminal justice system expresses the value of civic trust is through the presumption of innocence. Furthermore, even after it has been proved beyond reasonable doubt that an offender is not innocent — that he committed a criminal wrong and deserves punishment — the legal system must presuppose that the offender will once again be worthy of civic trust after he has been punished. One of the key aims of punishing the offender is ultimately to restore the offender to the community and to reconcile him to his fellow citizens. This emphasis on reconciliation implies that once the punishment is complete, the offender is once again entitled to civic trust. He must again be treated as someone who is capable of acting in accordance with the standards expected of a member of the community, capable of showing an appropriate level of respect for others. We must not give up on offenders by treating them as if they are beyond redemption. This means that we owe it to the offender to 'to treat him as someone who can, and to whom we owe it to hope that he will, refrain from crime in the future.'<sup>23</sup>

If, after receiving biomedical moral enhancements, offenders' actions were morally worthless (or were rendered significantly deficient in moral worth) then these interventions could not properly restore offenders to the community or reconcile them to their fellow citizens. These 'enhanced' offenders would have been deprived of the opportunity to redeem themselves. Their law-abiding actions would be more like mere conformity, rather than a true expression of civic virtue. To accept the suggestion that the need for public protection could outweigh the value of preserving the moral worth of offenders' future actions would therefore go against the aims of communicative punishment.<sup>24</sup> The only

---

<sup>19</sup> Ibid 48.

<sup>20</sup> RA Duff, 'Pre-Trial Detention and the Presumption of Innocence' in Andrew Ashworth, Lucia Zedner and Patrick Tomlin (eds), *Prevention and the Limits of the Criminal Law* (Oxford University Press, 2013) 115–33.

<sup>21</sup> Ibid 21.

<sup>22</sup> Duff, *Punishment, Communication and Community* (n 18) 48.

<sup>23</sup> Ibid 124.

<sup>24</sup> Douglas suggests that there are '... cases of ... moral enhancement where the disvalue of any loss in freedom to be immoral is outweighed by the value of the reduction in immoral behaviour or motivation': Thomas Douglas, 'Moral Enhancement via Direct

possible exception, on Duff's view, would be incorrigibly persistent, serious, violent, offenders. Duff reluctantly concedes that citizens may not have an 'unconditional' right to be restored to the community.<sup>25</sup> If an offender has waged what could be viewed as a 'continuing campaign of attacks on the community's members and its central values', then it might be unrealistic to think that civic trust could ever be restored.<sup>26</sup> Duff tentatively accepts that life imprisonment might be appropriate for such offenders. Even if biomedical moral enhancement would deprive offenders' actions of moral worth, so that they could not be fully restored to the community, nevertheless employing such techniques so that they could safely be released into the community would arguably be more humane and closer to restoration than life imprisonment.<sup>27</sup> However, for all but this most serious category of offenders, it would be disproportionate to deprive offenders of the capacity to perform actions with moral worth (or to severely impair that capacity), as this would constitute a kind of 'life sentence' that would not be warranted.

So far, in this section, I have argued that biomedical moral enhancement of offenders would be impermissible, on a communicative theory of punishment, if it undermined the moral worth of offenders' future conduct. I will now turn to another aspect of Duff's theory that allows us to identify a way in which biomedical moral enhancements might diminish the moral worth of offenders' future behaviour. In subsequent sections I will discuss whether three different

---

Emotion Modulation: A Reply to John Harris' (2013) 27(3) *Bioethics* 160, 166. Even if the freedom to be immoral is necessary for good conduct to have moral worth, Douglas would presumably maintain that the disvalue of depriving an offender's future conduct of moral worth could, in some cases, be outweighed by the value of the reduction in immoral behaviour.

<sup>25</sup> Duff, *Punishment, Communication and Community* (n 18) 173.

<sup>26</sup> Ibid 172.

<sup>27</sup> According to Duff, even for the most serious category of offenders, life imprisonment should only be 'presumptively' permanent, not 'irreversibly permanent': Duff, *Punishment, Communication and Community* (n 18) 172. The offender should eventually be released if he convincingly demonstrates that he has in fact reformed. Analogously, if this type of serious persistent offender somehow demonstrates that he can be trusted to abide by the law without the aid of a biomedical intervention, then the intervention should be withdrawn. It is not clear, however, how the offender would be able to demonstrate his trustworthiness in practice, in either case. Perhaps the offender could be allowed periods of somewhat greater freedom (eg periods of time under a less restrictive prison regime, or supervised periods of time without the biomedical intervention) and his freedom could gradually be increased if he showed that he could be trusted. There may be other reasons, apart from the 'moral worth' consideration, why it might be permissible to impose a sentence of life imprisonment on some offenders, but impermissible to biomedically morally enhance them. For instance, it may be easier to justify interfering with offenders' freedom of movement than it is to justify interference with their bodily and mental integrity. See, eg, Elizabeth Shaw, 'Against the Mandatory Use of Neurointerventions in Criminal Sentencing' in David Birks and Thomas Douglas (eds), *Treatment for Crime* (Oxford University Press, 2018).

types of moral enhancement could compromise the moral worth of offenders' law-abiding behaviour in this way.

Arguably, for an action to have moral worth, it is not sufficient for the agent to have certain rational capacities at the time of acting and to be motivated by the right reasons. It is often argued that these capacities and motivations must have a certain causal history.<sup>28</sup> Theorists have observed that, if an intervener imposes a biomedical intervention on the recipient, then the capacities and motivations arising from this non-consensual intervention might not have a worth-conferring causal history.<sup>29</sup> This article suggests a further possibility, stemming from the communicative approach to punishment: the historical pre-conditions for a law-abiding action to have moral worth may differ depending on whether or not the actor has previously committed a seriously immoral (and criminal) act.

The law-abiding behaviour of a non-offender may have a high degree of moral worth, even if that non-offender is "naturally good" and finds moral conformity easy. Douglas provides the following example of an agent whose actions intuitively 'can possess a very high degree of moral worth, even if they are relatively effortless':<sup>30</sup>

*David*

Compared to his peers, David conducts himself in a way that accords well with the moral reasons that apply to him. Indeed, he finds it easy to morally conform ... It is not that he automatically does what morality requires; he frequently has to deliberate about what to do. But his deliberation is seldom biased or disrupted by powerful impulses ... and sound deliberation is facilitated by the ease with which he is able to imagine the consequences of his actions and empathise with those he affects.<sup>31</sup>

In contrast, it may be that if someone who has committed a serious criminal and moral wrong in the past fails to exert effort and experience emotional distress, their future conduct will be deficient in moral worth. Imagine that a husband, one night, flies into a temper and beats his wife. He acknowledges, intellectually, that his assault on his wife was seriously wrong, but he does not go through any effortful, emotionally painful process of repentance. However, his conduct towards his wife after that conforms to his moral obligations and overtly appears affectionate and considerate. Furthermore, this future conduct is motivated partly by his recognition of his moral obligations. His motivations at the time of

---

<sup>28</sup> For example, John Fischer and Mark Ravizza argue that it would be conceptually impossible to create a 'virtue pill', as the resulting virtues would not have been acquired in the right manner: John Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge University Press, 1998) 182.

<sup>29</sup> Douglas, 'Enhancing Moral Conformity and Enhancing Moral Worth' (n 3) 83.

<sup>30</sup> Ibid 85.

<sup>31</sup> Ibid 84.

his apparently considerate conduct towards his wife, when viewed as a time-slice, may be indistinguishable from the motivations of a genuinely good husband. However, it might seem that his apparently considerate conduct towards his wife after the beating is markedly deficient in moral worth, as it is tainted by his previous wrongdoing. It is not only that repentance is desirable in itself. It might be thought that his future conduct towards his wife (assuming that she is willing to continue their relationship) will be tainted by his past wrongdoing, unless it emerges from and is carried out in the light of his sincere repentance. This repentance would have to involve taking the time and effort to confront the moral flaws that lay behind his wrongdoing, to appreciate the full implications of his wrongdoing and to work out the steps he should take if their relationship can be healed.

On the communication view of punishment, an offender's sentence should aim to bring about moral reform, via an effortful process of repentance that involves emotionally painful reflection on the crime the offender committed. Duff argues that in order to be restored to the community, the offender must receive a burdensome punishment that should both aim to facilitate his reform and function as a symbolic apology. If a biomedical moral enhancement brings about moral conformity in a way that is 'too easy' or is disconnected from the offender's deliberations about the wrongfulness of her original offence, then the offender's subsequent law-abiding behaviour may altogether lack moral worth, or its moral worth may be significantly diminished.<sup>32</sup> This point can be illustrated with the following examples:

#### *Reform Through Moral Communication*

X used to be callous and hot-tempered with violent tendencies, particularly when drunk. He was convicted and punished for committing an alcohol-fuelled assault that left his victim with serious injuries. As part of his punishment, he participated in a victim-offender mediation programme, which forced him to confront the effects of his crime on the victim. The probation officer, as mediator, made clear to the offender that the censure communicated through this process came not only from the victim, but also from the community as a whole. As a result of this process, X agreed to undertake reparation, which included participating in a rehabilitation programme aimed at addressing his violent behaviour. This rehabilitation included 'confrontational group work involving re-enactment and discussion'.<sup>33</sup> The rehabilitation programme was not merely therapeutic. Rather it encouraged X to take responsibility for his crime, to fully recognise its seriousness and to face up to the need to change. Through this

---

<sup>32</sup> Although Duff does not explicitly discuss the pre-conditions for the moral worth of offenders' actions after punishment, it is plausible that his theory has these implications for moral worth.

<sup>33</sup> Duff, *Punishment, Communication and Community* (n 18) 103.

programme and with the help of a probation officer, the offender also identified strategies for developing self-restraint that allow him to curb his violent temper and address his misuse of alcohol. He gradually developed greater compassion and insight into the effect of his actions on others, as a result of the human interactions he experienced as part of his punishment — interactions with his victim, with the people who ran the rehabilitation programme, with the other offenders on this programme and with the probation officer. Both the victim mediation programme and the rehabilitation programme were intentionally burdensome. They were designed to ‘forcefully and uncomfortably’ confront X with the nature of his crime and to induce a ‘necessarily painful repentant recognition’ of his wrongdoing. His participation in these programmes required considerable effort. In particular he had to exert effort in order to gain mastery over his violent impulses, to identify the factors that led him to commit the crime and to challenge rationalisations that he had previously used in order to minimise his wrongdoing. Now that he has been released into the community, X manages to conform to his moral and legal obligations by continuing to exert effort, eg by reminding himself of his commitment not to repeat his original crime, by continuing to subdue his anti-social urges, and by forcing himself to consider the impact of his behaviour on others. Over time he finds it easier to avoid breaking the law, but the process of gaining a repentant recognition of his earlier wrongdoing has a ‘sustained effect on how [X] thinks about [his] future actions in the shadow of [his] past wrong.’<sup>34</sup>

### *Biomedical “Reform”*

Y used to be callous and hot-tempered with violent tendencies, particularly when drunk. He was convicted of committing an alcohol-fuelled assault that left his victim with serious injuries. The state provided him with a biomedical intervention that altered his neurochemistry in a way that greatly reduced his violent impulses and increased his capacity for empathy. These psychological changes occurred without effort on his part. It is now as easy for Y to conform to morality as it is for a naturally virtuous individual, like the earlier example of David. As a result of the biomedical intervention, Y does conform.

There are a number of different reasons that might be given for preferring *Reform Through Moral Communication* to *Biomedical Reform*. This article focuses on just one — the intuition that the law-abiding actions of offenders like X, after they have been reformed through moral communication, have significantly greater moral worth than the law-abiding behaviour of offenders like Y, after they have received the biomedical intervention. This intuition is distinct from the

---

<sup>34</sup> Jonathan Pugh and Hannah Maslen, ““Drugs That Make You Feel Bad”? Remorse-Based Mitigation and Neurointerventions’ (2015) 11(3) *Criminal Law and Philosophy* 499.

consideration that repentance has intrinsic value.<sup>35</sup> This article focuses on the importance of repentance (and the painful feelings and effort that it involves) for enabling the offender's future conduct to have moral worth.

I think some would intuit that Y's behaviour has significantly less moral worth than X's behaviour. This intuition may be shared by theorists who agree with the doubts expressed by the prison chaplain from *A Clockwork Orange* about whether Ludovico's technique could 'really make a man good'.<sup>36</sup> As mentioned above, one reason that has been suggested for this intuition about Ludovico's technique is that it deprives the offender of the ability to *choose* between good and bad behaviour, because it makes bad behaviour *impossible*<sup>37</sup> (I will briefly return to this consideration in the next section). However, this consideration has limited application to biomedical interventions that might realistically be given to offenders. Most biomedical interventions that are currently available (eg methadone for drug-addicted offenders) do not make reoffending strictly impossible and it seems relatively unlikely that interventions with such an extreme effect will be used in the near future. Even surgical castration, which does make one specific way of offending impossible, cannot prevent the offender from committing crimes (including sexual crimes) in other ways. Rather than claiming that biomedical interventions could make bad behaviour impossible, it would be more plausible to suggest that they might strongly incline the offender towards law-abiding behaviour and away from criminal behaviour. The option of bad behaviour would still be there, but it would be much easier for the offender to choose to be law-abiding. Why might one still have doubts about whether such interventions could 'make [an offender] good'?<sup>38</sup> The explanation I have

---

<sup>35</sup> It can also be distinguished, in theory, from the idea that the suffering of the guilty is intrinsically good – an idea defended by the retributivist Michael Moore in *Placing Blame* (Oxford University Press, 1997). However, it might be argued that the plausibility of the idea that the offender's reform must be effortful simply stems from this retributive desire to make the offender suffer. In order to rebut this claim it would need to be established that effort really is necessary for true reform to take place.

<sup>36</sup> Burgess (n 6) 72.

<sup>37</sup> Hauskeller (n 7). Hauskeller ultimately concludes that biomedical moral enhancement is problematic, not because the choice to do good can only be free if bad behaviour is possible, but rather because it undermines *relational* freedom if *other people* make it impossible for an individual to behave badly. I agree that such manipulation would be *wrong*, because of the unequal power-relationships involved (among other reasons). However, I am not convinced that the simple fact of manipulation by another *per se* can explain the intuition that biomedical interventions *cannot make an offender good*. For an argument that manipulation *per se* need not undermine moral responsibility for one's actions, see, eg, G Harrison, 'A Challenge for Soft-Line Replies to Manipulation Cases' (2010) 38 *Philosophia* 555. The argument in the current article is that manipulation might undermine the moral worth of offenders' actions if it allowed the offender to avoid effortful repentance.

<sup>38</sup> Burgess (n 6) 72.

suggested in this article is that the offender's law-abiding behaviour would not have been the result of an effortful process of repentance.

It might be objected that, even if the moral worth of Y's good actions were somewhat diminished by the fact that he took the 'easy route', nevertheless his actions could still have a high degree of moral worth. If, after the intervention, he is reasons-responsive and treats others with respect, because he recognises the moral reasons for doing so, surely his actions have considerable moral worth, despite the fact that his good behaviour has been made much easier by his biomedically increased empathy and reduced aggression. Furthermore, even if avoiding effortful repentance for his crime (assault) undermined the moral worth of his refraining from repeating the same type of crime, surely it would not undermine the moral worth of good deeds he performs that are unrelated to his original crime (eg his honest business dealings, or his good parenting of his children). Even if the effortful route of repentance were the only way that all of the offender's future actions could have the highest degree of moral worth, why should he be forced to pursue this route? If, after receiving the biomedical intervention, he were still capable of performing a wide range of actions with a high degree of moral worth, how would the biomedical intervention prevent him from being genuinely reformed and reconciled to his fellow citizens?

In response, a communication theorist might acknowledge that Y's moral conformity after the intervention would have some moral worth, but insist that his actions would be severely deficient in one particular kind of moral worth; and if his actions cannot have this kind of moral worth, reconciliation and reform cannot take place. In order to be a reformed member of the community and to be reconciled to his fellow citizens, his future actions must be capable of having moral worth as the actions of a good member of the community, committed to the community's values. By perpetrating his crime, he failed to live up to the standards expected of a member of the community. His law-abiding actions after punishment can only express a renewed commitment to those standards if he has repented his violation of the community's values. Arguably, a failure to repent a serious violation of one of the community's central values (eg a norm prohibiting unjustified violence) would taint all of his future actions, so that even his conformity to norms that he had not previously violated would fail to have moral worth as the actions of a good member of the community. Even if this idea were not accepted, his failure to properly repent his crime might still seriously undermine the moral worth – qua the actions of a good citizen – of his refraining from repeating the same type of crime and thus prevent him from being fully reconciled to the other members of the community (the requirement for effortful repentance seems more plausible in relation to the most serious crimes, eg crimes involving violence or sexual assaults, rather than, for instance, certain property crimes or regulatory offences. However, this requirement has particular relevance in the present context, as proponents of the biomedical moral enhancement of offenders cite as one of the main advantages of such interventions their potential to prevent serious violent and sexual offences).

The communication theorist's argument might be criticised for relying on the idea that the biomedical moral enhancement would necessarily allow the offender to avoid repenting his crime. What if the offender is already repentant by the time he is convicted? Perhaps he does not require the state's help to repent, but just requires some help putting his repentant resolution not to reoffend into action. In this case, why not just offer him a biomedical intervention that makes it easier to conform to his resolution, rather than putting him through a laborious process of moral communication, designed to induce the repentance that he already experiences? Alternatively, if he is not yet repentant, perhaps the biomedical intervention could instil repentance, without the need for the offender to undergo the laborious process described in the example of Reform Through Moral Communication.

A reply to this criticism is suggested by Duff's claim that, conceptually, the process of repentance requires considerable time and effort. He draws an analogy with grief.<sup>39</sup> It is not conceptually possible for a person to complete the grieving process for someone who was very important to her in a single day. A 'grief pill' could not make this possible. Similarly, repentance is not something that the offender can quickly undergo by himself straight after committing his crime. Nor could immediate repentance be instilled by a biomedical intervention. Duff writes:

Repentance is not something that can be achieved and completed in a moment: At least with serious wrongs, it requires time and effort. The initial horror or distress at what I have done, which might indeed strike me in a moment, must be deepened and strengthened into an understanding of my action as a wrong – an understanding that will stay with me.<sup>40</sup>

... It must go deep with the wrongdoer and must therefore occupy his attention, his thoughts, his emotions, for some considerable time.<sup>41</sup>

However, this does leave open the possibility that a biomedical moral enhancement that *facilitates repentance*, when *combined* with moral communication might be acceptable to the communication theorist. This possibility will be discussed in subsequent sections.

Before considering which specific biomedical interventions might be acceptable to the communication theorist, it is necessary to make a further clarification to my argument. I am not suggesting that if an offender's sentence, in fact, brings about mere conformity, rather than reform, that sentence is necessarily unjustified. Indeed, for some offenders, traditional punishment simply operates as a deterrent, causing them merely to conform, for prudential reasons, rather than to undergo genuine moral reform. According to Duff, this does not

---

<sup>39</sup> Duff, *Punishment, Communication and Community* (n 18) 108.

<sup>40</sup> Ibid 119.

<sup>41</sup> Ibid 118.



undermine the communicative justification for punishing them. I argue that it would be problematic, on a communicative theory of punishment, if a) biomedical moral enhancements deprived the offender of the *capacity* to perform actions which have genuine moral worth, or b) the use of biomedical interventions allowed the state to avoid making a serious attempt to persuade the offender to take the effortful, repentant route to moral worth.

In the next three sections of this article I will distinguish between three different types of biomedical moral enhancement – pure moral capacity enhancements, inclination-altering enhancements, and non-reasons-responsive enhancements. I will consider each type of intervention from the perspective of the communicative theory of punishment and will consider whether each intervention would undermine the moral worth of the offender's future actions.

## 5 Pure Moral Capacity Enhancements

I will use the term 'pure moral capacity enhancement' to refer to biomedical interventions that enhance one's ability to engage in rational deliberation about one's moral obligations and to translate one's decisions into action. The aim of such enhancements is not to ensure (or make it more likely) that the agent will act one way rather than another. Rather, these enhancements increase the agent's ability to make up her own mind about how to act.<sup>42</sup> If the state were to give an offender a pure moral capacity enhancement, this may or may not result in the offender conforming to the state's conception of morality. Furthermore, this type of enhancement could only cause the person to conform to morality, *via* enhancing her deliberative capacities. It could not bring about moral conformity independently of the agent's deliberation. In contrast, inclination-altering enhancements (which will be discussed in the next section) make it more likely that the person conforms to morality, even without engaging in deliberation about what duty requires. Although inclination-altering enhancements might also enhance a person's capacity to engage in moral deliberation, they are not 'pure' moral capacity enhancers, because enhancing capacities is not their only effect – they also increase the likelihood of conformity.

The category of pure capacity enhancements has some similarities to Schaefer and Savulescu's concept of 'procedural enhancements', which aim to enhance an agent's ability reliably to make correct moral judgements.<sup>43</sup> Procedural enhancements include cognitive enhancements such as increasing logical competence, conceptual understanding and empirical competence, reducing bias

---

<sup>42</sup> Shaw (n 1) states that this type of intervention enhances moral responsibility in the sense of enhancing capacities required for responsibility. G Owen Schaefer refers to them as indirect moral enhancements: G Owen Schaefer, 'Direct vs Indirect Moral Enhancement' (2015) 25(3) *Kennedy Institute of Ethics Journal* 261.

<sup>43</sup> G Owen Schaefer and Julian Savulescu, 'Procedural Moral Enhancement' (2016) *Neuroethics* (advance) <<https://doi.org/10.1007/s12152-016-9258-7>>.

and increasing one's openness to revising pre-existing views in the light of reasons for doing so. Pure capacity enhancements would include all of those capacities. In addition, they would also include increasing the ability to focus and to ignore distractions. Cognitive enhancement might be achieved through pharmaceuticals, such as methylphenidate.<sup>44</sup> Another relevant capacity, is the cognitive and emotional capacity for 'mental time-travel' – the ability to identify with one's past and future self. Also, memory enhancements could help to facilitate repentance, by restoring or sharpening the offender's memory of his crime.<sup>45</sup> Potential memory enhancements include pharmacological agents,<sup>46</sup> transcranial direct current stimulation (which involves placing a device outside the skull, which transmits an electrical current to brain areas underneath),<sup>47</sup> and direct brain stimulation (which involves implanting a device inside the brain).<sup>48</sup>

Pure capacity enhancements could also include the strengthening of will power. Savulescu and Schaefer did not class this as a 'procedural enhancement' because it did not fit with their focus on judgements, although they said it may be useful in conjunction with procedural enhancements.<sup>49</sup> I include this in the category of pure capacity enhancements, because it would increase the individual's ability to engage in repentance and reform, but would not by itself bring about moral conformity. Will power could equally be used to resist the temptation to offend or to overcome the fear of being caught when committing a crime. A technique, which might enhance the capacity to exercise will power is neurofeedback. This involves watching real-time images of one's brain functioning and thereby, over time, learning to alter one's own brain activity in response to this visual feedback. Studies have shown behavioural improvements in juvenile offenders who were

---

<sup>44</sup> Methylphenidate can also reduce impulsivity and so might be classed as an inclination-altering enhancement. This drug is sometimes given to prisoners with ADHD, but many prisoners with ADHD do not receive treatment. Ylva Ginsberg et al, 'Long-Term Functional Outcome in Adult Prison Inmates with ADHD Receiving OROS-Methylphenidate' (2012) 262(8) *European Archives of Psychiatry and Clinical Neuroscience* 705.

<sup>45</sup> Pugh and Maslen (n 34).

<sup>46</sup> Eg acetylcholine: Paula Croxson et al, 'Acetylcholine Facilitates Recovery of Episodic Memory after Brain Damage' (2012) 32(40) *The Journal of Neuroscience* 13787.

<sup>47</sup> Rosa Manenti et al, 'Enhancing Verbal Episodic Memory in Older and Young Subjects after Non-Invasive Brain Stimulation' (2013) 49 *Frontiers in Aging Neuroscience* 1.

<sup>48</sup> Rollin Hu, Emad Eskandar and Ziv Williams, 'Role of Deep Brain Stimulation in Modulating Memory Formation and Recall' (2009) 27(1) *Neurosurgical Focus*. The risky and invasive nature of neurosurgery gives rise to distinct ethical issues, which will not be discussed here.

<sup>49</sup> Schaefer and Savulescu (n 43).

able to re-train their brain-wave patterns and thereby gain more control over their impulses.<sup>50</sup>

Another type of enhancement, which Savulescu and Schaefer do not mention, is 'emotional distancing' – the ability to mentally 'stand back' from one's emotions and to make a decision, without being overpowered by one's emotions. This is distinct from simply reducing strength of the emotion. Again, I class this as a pure capacity enhancement, as the intervention by itself would not make moral conformity more likely. The criminal might use this capacity in order to harden himself to his victim's distress.

In addition, another relevant capacity is the body's ability to physically 'remember' emotionally aversive experiences, by generating a subconscious, autonomic response when faced with a similar situation. This somatic 'marking' of aversive experiences can allow the individual to avoid repeating past mistakes, and may play a role in learning to avoid immoral behaviour.<sup>51</sup> It is debatable whether this could count as a pure capacity enhancement, as it might have a direct effect on the offender's motivations. However, it would not lead to a reduction in the likelihood of immoral behaviour, without further deliberation, if the individual did not already find immoral behaviour aversive.

Schaefer and Savulescu include empathy enhancements in the category of 'procedural enhancement'. However, I think it is likely that increasing the capacity for empathy would somewhat increase one's motivation to benefit others or refrain from harming others, even without further moral deliberation about one's moral obligations. Therefore, I have classed empathy enhancement as an inclination-altering enhancement.

In principle, pure capacity enhancements seem relatively easy to reconcile with the communicative theory of punishment, provided that the offender gives valid consent. The main hurdle, as I will discuss briefly below, is ensuring that offenders can give valid consent within the coercive context of the criminal justice system. If consent is given, the effects of pure capacity enhancements seem largely unproblematic for the following reasons.

First, given the above-mentioned definition of these enhancements, it would be difficult for a communication theorist to argue that these interventions would undermine the *capacity* to perform actions with moral worth. Even if, as some theorists suggest, the freedom to do bad acts is necessary for good actions to have moral worth, this freedom could not be threatened by pure capacity

---

<sup>50</sup> P Smith and S Marvin, 'Neurofeedback with Juvenile Offenders: A Pilot Study in the Use of QEEG-Based and Analog-Based Remedial Neurofeedback Training' (2006) 9 *Journal of Neurotherapy* 87.

<sup>51</sup> W Schmitt, C Brinkley and J Newman, 'Testing Damasio's Somatic Marker Hypothesis with Psychopathic Individuals: Risk Takers or Risk Averse?' (1999) 108 *Journal of Abnormal Psychology* 538.

enhancements.<sup>52</sup> Nor would these interventions interfere with moral worth, on a Kantian conception of moral worth, as they would not make it any less likely that the agent would act from the motive of duty.<sup>53</sup> It seems that the effortful repentance, which communication theorists consider necessary for true reform, would not be undermined by (most) pure capacity enhancements, but repentance would in fact be facilitated by such interventions. However, it might be argued some of these enhancements are problematic on a communication view of punishment, particularly the idea of increasing will power through biomedical means. It might be thought that the offender should learn, through the process of repentance, to control her own impulses.

Secondly, it is unlikely that employing these enhancements would allow the state to avoid making a serious attempt to reform the offender through moral communication. This is because these enhancements could not, by definition, increase the likelihood of moral conformity by themselves, without the recipient engaging in further deliberation. Therefore, there is little danger that they would be used as a substitute for moral communication. Furthermore, since there is a risk that the offender might even use her enhanced capacities to learn to become a *more effective criminal*, the state would have an added incentive to try to engage the offender in moral dialogue aimed at persuading her to use her enhanced capacities in order to reform.

If these interventions could increase the offender's capacity for repentance and reform, why not impose them on a mandatory basis? I believe communication theorists would insist that such interventions should only be used consensually, if at all. This is because communication theory stresses the importance of moral dialogue. This presupposes the existence of *two separate parties*. If the state were to force the offender to accept biomedical means of changing her communication style and mental capacities, this would unacceptably blur the boundary between the two parties. The state would not only be presenting the offender with reasons to change, but would also be manipulating the way in which the offender mentally responds to that information (albeit in a way that does not ensure that the offender is persuaded). In order to respect the offender as a separate party to the dialogue, the state must acknowledge the offender's right to respond using her pre-existing mental capacities.

This gives rise to the problem of whether the offender can give valid consent in the coercive context of the criminal justice system. I have argued elsewhere that just because the offender faces a hard choice, eg between accepting the enhancement or potentially facing a longer period of time in prison, this does not necessarily undermine the validity of her consent, provided that the initial prison sentence was legitimate.<sup>54</sup> After all, many individual's make hard choices

---

<sup>52</sup> See the discussion of freedom in Section 2, above.

<sup>53</sup> Douglas, 'Enhancing Moral Conformity and Enhancing Moral Worth' (n 3).

<sup>54</sup> Shaw (n 1).

between two undesirable outcomes (e.g. in the context of medical decisions about whether to undergo risky surgery or continue to live with painful symptoms) and yet they are still able to give valid consent.

John McMillan has argued that the state should not link the length of an offender's prison sentence with the offer of a biomedical intervention, ie the offender should not be given the impression that his acceptance or refusal of the intervention will have any effect on his sentence.<sup>55</sup> Although, I am not convinced that this requirement is strictly necessary in order for consent to be valid, it is probably a requirement that would be in accordance with communication theory. This is because according to that theory the length of the sentence should be proportionate to the crime and express its level of seriousness. To release the offender early, because she had accepted a biomedical intervention that facilitated her reform, might send out the message that the offence was less serious than it actually was.

## 6 *Inclination-Altering Enhancements*

I will use the term 'inclination-altering enhancements' to refer to direct brain interventions that affect emotions or impulses that incline the individual toward or away from moral conformity. The enhancement might 1) instil a positive desire to behave in ways that conform to morality or 2) dampen the individual's desire to engage in immoral behaviour. Tom Douglas's example of increasing sympathy through directly altering subconscious processes falls into the first category.<sup>56</sup> In addition, increasing sympathy might counterbalance immoral desires, eg an urge to resort to physical violence when frustrated. Oxytocin seems to play a role in empathy and might be the target of inclination-altering interventions.<sup>57</sup> Drugs that decrease aggressive impulses fall into the second category.<sup>58</sup> There is some evidence that selective serotonin reuptake inhibitors ('SSRI's') may reduce aggression.<sup>59</sup> I will use the term 'inclination-altering enhancements' to cover only interventions that leave the agent's ability to respond to reasons intact. Many

---

<sup>55</sup> John McMillan, 'The Kindest Cut? Surgical Castration, Sex Offenders and Coercive Offers' (2014) 40(9) *Journal of Medical Ethics* 583.

<sup>56</sup> Douglas, 'Enhancing Moral Conformity and Enhancing Moral Worth' (n 3).

<sup>57</sup> Mark Dadds et al, 'Methylation of the Oxytocin Receptor Gene and Oxytocin Blood Levels in the Development of Psychopathy' (2014) 26(1) *Development and Psychopathology* 33.

<sup>58</sup> Ibid. Decreasing racial or sexist bias could also be classed as an inclination-altering enhancement, because removing such biases would reduce immoral behaviour. However, a reduction in these attitudes might also be achieved through cognitive enhancement (discussed in the previous section). Savulescu and Schaefer categorise bias reduction as a 'procedural enhancement'.

<sup>59</sup> P Ferari et al, 'Escalated Aggressive Behaviour: Dopamine, Serotonin and GABA' (2005) 526 *European Journal of Pharmacology* 51.

theorists endorse Fischer and Ravizza's view that, for an agent to be deserving of praise and blame, it is necessary for the agent to be reasons-responsive.<sup>60</sup> Reasons-responsiveness requires that the agent is receptive to a range of relevant reasons for doing otherwise and would react to at least one of these reasons if it were present.<sup>61</sup> I will consider enhancements that undermine reasons-responsiveness in the next section.

Unlike pure capacity enhancements, inclination-altering enhancements have the potential to bring about moral conformity on their own without the recipient engaging in further deliberation. For example, an offender who receives aggression-reducing drugs might refrain from assaulting people simply because she can no longer be bothered to get into fights – it no longer seems exciting. If the state gave this kind of intervention to offenders, without any attempt to get offenders to pursue the route of effortful repentance, would that deprive offenders of the *capacity* to perform acts with moral worth? Douglas thinks that the aggression-lowering drug would not deprive the recipient of this capacity (on a Kantian understanding of moral worth).<sup>62</sup> The drug would not *prevent* the recipient from engaging in moral deliberation and from acting out of the motive of duty. However, if the argument outlined in section 3 is correct, being motivated by duty is not enough for the offender's future actions to have a sufficient degree of worth, qua the actions of a good citizen (at least in cases where the offender has previously committed one of the most serious crimes, e.g. involving violence or sexual assaults). To have this kind of worth, the offender's future law-abiding behaviour must emerge from a process of effortful repentance for his previous wrongdoing. It might be objected, though, that even if the offender's sentence consisted solely in the administration of an aggression-lowering drug, this still would not prevent him from taking a worth-conferring route to law-abiding behaviour on his own, by privately repenting, after receiving this drug. Duff might reply that to the extent that 'private repentance' for, eg a serious assault, is possible, it cannot reconcile the offender with the community (although it might reconcile him with God or with his conscience).<sup>63</sup> If the offender does not undergo effortful, communicative punishment, he has missed his only chance to be reconciled with the community. It might be thought, therefore, that if the state were simply to give the offender the aggression-lowering drug, without requiring him to undergo communicative punishment, he would lack the capacity to perform actions that have sufficient worth qua the actions of a good member of the community.

Even if an offender's *capacity* to perform actions with moral worth were in no way threatened by a sentence consisting solely in an aggression-lowering drug, communication theorists would still object to this type of sentence. According to

---

<sup>60</sup> Fischer and Ravizza (n 28).

<sup>61</sup> Ibid.

<sup>62</sup> Douglas, 'Enhancing Moral Conformity and Enhancing Moral Worth' (n 3).

<sup>63</sup> Duff, *Punishment, Communication and Community* (n 18) 119.

communication theorists, the state has a duty to make a serious attempt to bring about genuine reform through punishment. The sentence should aim to get the offender to take the effortful, repentant route to moral worth. Giving the offender an aggression-lowering drug, by itself, would not constitute a serious attempt at trying to reform him.

This raises the question of whether a communication theorist could endorse a sentence that combined an inclination-altering intervention with an attempt to reform the offender through moral dialogue. The answer to this would seem to depend partly on the state's intentions when offering the drug. Consider the following three possibilities:

1. The state engages the offender in moral communication in the hope that he will reform as a result, but the state employs an inclination-altering intervention as a back-up, intending that, if communication fails to induce repentance, the biomedical intervention will bring about moral conformity.

This possibility would be unacceptable according to Duff's communication theory. According to Duff the state must always intend to bring about reform via moral communication. To rely on a route to moral conformity which bypasses moral communication (even as a 'back-up') would 'be inconsistent with a proper regard for actual and potential offenders as members of the normative political community'.<sup>64</sup> He makes this point when rejecting the idea that the state could *rely* on the prudential deterrent effect of punishment, as a 'supplement' to the moral reasons for refraining from crime.<sup>65</sup> This also suggests that the communication theorist could not endorse *partially* replacing communicative punishment with the biomedical intervention (eg by combining an aggression-lowering drug with a scaled-down attempt at moral communication – a sort of 'communication light'). Such partial reliance on the drug would not constitute a sufficiently serious attempt to bring about reform through effortful repentance.

2. The state offers an inclination-altering biomedical intervention with the intention of using it in order to facilitate repentance through moral communication. For example, an aggression-lowering drug might enable the offender to engage in moral communication more effectively, without being distracted by violent impulses. The state *intends* that reform will arise from the drug-facilitated communication, rather than from the drug directly. However, the state can *foresee* that, in addition to facilitating moral communication, the drug may have a direct impact on the offender's violent behaviour, so that the drug-induced reduction in aggressive impulses may be (part of) the reason why he refrains from violent crime.

---

<sup>64</sup> Ibid 124.

<sup>65</sup> Ibid.

This second possibility might be acceptable on Duff's view. As argued in the previous section, the consensual use of a drug to facilitate repentance seems consistent with communication theory. Nor are the foreseen side-effects of this drug on the offender's moral conformity necessarily problematic. This is indicated by Duff's comments about the deterrent effects of punishment. According to Duff, merely *foreseeing* that punishment will have a deterrent effect does not undermine the communicative justification of punishment.<sup>66</sup> If the state intends to bring about reform through the right means, this shows sufficient respect for offenders' agency and membership of the moral community, even if, in fact, offenders conform for the wrong reasons. However, if the state is only justified in using the drug to facilitate communication, then, if it is practically possible, the state should try to ensure that the drug does not have long-term effects that extend beyond the period of communicative punishment.

3. The state engages the offender in moral communication in order to bring about repentance. The offender shows signs of remorse, but it is doubtful whether the offender will be able to stick to his resolution to refrain from crime. The state offers the offender an inclination-altering biomedical intervention (eg an aggression-lowering drug) in the hope that the offender will accept the drug as the result of his repentant desire to do everything he can to prevent himself from reoffending.

This third possibility is less than ideal from the perspective of communication theory, as the state would have to acknowledge that moral communication is insufficient, by itself, to persuade the offender to refrain from crime. The moral worth of the offenders' law-abiding conduct is arguably not as great as it would have been if he refrained from wrongdoing without the aid of the drug (which perhaps he might have been able to do if he were more deeply repentant). However, in this scenario, even though the drug increases the chance that he will refrain from crime, this increase in the likelihood of moral conformity has still come about *via* repentance, albeit indirectly. Furthermore, the moral worth of his conformity does not stem *entirely* from his earlier, repentance-driven decision to take the drug. If the intervention does not undermine the offender's reasons-responsiveness it is still possible that, *at the time he conforms* to morality, he does so partly because he is responding to moral considerations (including his remorse for his previous crime). Using a biomedical intervention in this way might therefore be acceptable to the communication theorist provided that, first, the state has already made a serious attempt to reform the offender through moral communication and, secondly, there is evidence that the offender's repentance is genuinely what motivates him to accept the biomedical intervention. It might be problematic in practice to ensure the first requirement is satisfied, if effective biomedical means of rehabilitation become available. The temptation simply to rely on these biomedical techniques might weaken the state's motivation to make a sufficiently serious attempt at reforming him through moral communication.

---

<sup>66</sup> Ibid.



The second requirement also might be hard to fulfil, as there are clearly empirical problems establishing what the offender's true motivations are.<sup>67</sup> If these practical difficulties can be surmounted, however, communication theorists might accept using biomedical interventions to enable offenders to act in accordance with their repentant desire to refrain from crime.

## 7 *Non-Reasons-Responsive Interventions*

In this section, I will discuss interventions that increase conformity with moral norms, but do so in a way that undermines the agent's reasons-responsiveness in relation to that behaviour (I will refer to these techniques as 'non-reasons-responsive interventions'). Imagine an intervention that produces an irresistible inclination to conform to morality, or an *irresistible* disinclination to behave immorally. For example, if the recipient experiences extremely powerful feelings of sympathy it might be impossible to focus on the motive of duty when faced with someone in distress. These sympathetic feelings might drive the individual, compulsively, to help people in distress. Or a strong aversion to violence (similar to the results of the Ludovico method in *A Clockwork Orange*) might make it impossible for her to show aggression.<sup>68</sup> A similar real-world example is disulfiram, a treatment for alcoholism that causes vomiting, headaches and severe flu symptoms when alcohol is consumed. A person might find these symptoms so distressing that she becomes literally incapable of choosing to become intoxicated — after taking this intervention she is not reasons-responsive in relation to her abstinence. Disulfiram may be given to offenders whose alcohol consumption plays a role in their criminal behaviour. Henry Greely observes that the voluntary acceptance of this treatment 'in order to avoid harsher criminal sanctions seems relatively uncontroversial'.<sup>69</sup> Wayne Hall comments that disulfiram 'is a potentially cost-effective alternative to imprisonment for repeat [drink-driving] offenders'.<sup>70</sup> An intervention might also bring about moral conformity by making certain types of offending behaviour physically very difficult or impossible. For example, as noted above, surgical and chemical castration (while certainly not preventing all forms of sexual offending) can preclude certain ways of committing sexual offences. After receiving the intervention, the agent ceases to engage in these specific prohibited acts due to physical incapacity, rather than refraining from these actions as the result of

---

<sup>67</sup> For further discussion of the problem of ascertaining the offender's true motivations, see McMillan (n 55).

<sup>68</sup> Burgess (n 6).

<sup>69</sup> Henry Greely, 'Neuroscience and Criminal Justice: Not Responsibility but Treatment' (2006) 56(5) *Kansas Law Review* 1103, 1137.

<sup>70</sup> Wayne Hall, *The Role of Legal Coercion in the Treatment of Offenders with Alcohol and Heroin Problems* (National Drug and Alcohol Research Centre Technical Report No 44, 2007) 8.

psychological processes that are responsive to reasons.<sup>71</sup> Chemical castration is currently used in the rehabilitation of sex offenders in several US states (California, Florida, Louisiana, Montana, Iowa, Oregon, and Texas) as well as various European countries, including Sweden, Denmark, and the United Kingdom. Surgical castration is used in Germany and the Czech Republic and in the US states of California, Florida, Iowa, Louisiana, and Texas.

Of the interventions discussed in this article, non-reasons-responsive interventions are the hardest to justify on the communicative view of punishment. Such interventions could only (if ever) be justifiable after the state had made a serious attempt to reform the offender through moral communication. It would also be crucial that the offender was motivated to accept the intervention for the right reasons, ie due to a repentant recognition of the moral reasons for refraining from crime. This latter requirement would be particularly important, because the moral worth of the seemingly 'good' behaviour produced by the intervention would depend *entirely* on the 'tracing' principle – the idea that sometimes the praiseworthy/blameworthy character of behaviour at a later time, depends on tracing it back to an earlier decision for which the agent was praiseworthy/blameworthy. It is therefore necessary to consider the tracing principle in more detail.

An analogy can be drawn between tracing one's responsibility for praiseworthy conduct (eg refraining from harm and fulfilling one's responsibilities to help others) back to an earlier voluntary decision and 'prior fault' in the context of blameworthy conduct. Both the law<sup>72</sup> and many philosophers<sup>73</sup> acknowledge that a person can be held morally/criminally responsible for a *wrongful* action due to 'tracing'. Imagine a person who is reasons-responsive at time A and knowingly takes a pill that causes him to commit a crime at time B while not reasons-responsive. Intuitively, he is blameworthy for the bad outcome he produced at time B and he would be held criminally responsible. Similarly, it seems plausible, that a person who is reasons-responsive at time A and knowingly takes a pill that causes him to perform good actions at time B while not reasons-responsive is praiseworthy for the good actions produced at time B.

It might be objected, however, that there are certain asymmetries between praise and blame that make this analogy unreliable. For instance, 'it is generally thought that whether a good act is virtuous depends upon the actor's actually being motivated by the good, whereas a bad act could be vicious even if the actor wasn't

---

<sup>71</sup> In addition, these interventions have psychological effects (reducing sexual thoughts) that may reduce the likelihood of committing other kinds of sexual offences, while preserving reasons-responsiveness in relation to those acts.

<sup>72</sup> *R v Bailey* [1983] Crim LR 353.

<sup>73</sup> John Fischer states that tracing 'seems both highly plausible and theoretically indispensable' and observes that philosophers representing a diverse range of theories have embraced the idea. See John Fischer and Neal A Tognazzini, 'The Truth about Tracing' (2009) 43(2) *Nous* 531, 553.

motivated by the bad.’<sup>74</sup> One can be blameworthy/criminally responsible for merely being insufficiently motivated by the good. Many criminal offences only require, as the *mens rea*, recklessness as to the consequences of one’s actions.

In response, this asymmetry between praise and blame does not provide sufficient reason to reject the idea that ‘tracing’ can apply to praiseworthy behaviour. Rather, it suggests that the way tracing operates may differ depending on whether the behaviour is blameworthy or praiseworthy. To be responsible for the effects of a pill that induces bad behaviour, it is sufficient that one was reckless as to the consequences at the time one took the pill. In contrast, to be praiseworthy for the consequences of taking the moral enhancement pill, one must take the pill for the right reasons – one must be motivated by the good.

If this is correct, then it provides grounds for opposing mandatory non-reasons-responsive interventions. If such an intervention is forced on an unwilling offender then she will lack the capacity to be praiseworthy for the moral conformity the intervention produces. She cannot be praiseworthy for the pill’s effects in virtue of a prior voluntary decision to take the pill, because she made no such decision. Nor can she be praiseworthy for the moral conformity the pill produces even if she *subsequently* endorses its effects. This is because the effects are produced via a mechanism that is not responsive to reasons. Although she might come to recognise the reasons for behaving in the way the pill makes her behave, her recognition of those reasons is not the thing that brings about her behaviour. Therefore, if, as argued earlier, depriving someone of the capacity to be praiseworthy is ethically problematic, then this provides grounds for concern about administering non-reasons-responsive interventions on a mandatory basis. In terms of undermining praiseworthiness, mandatory non-reasons-responsive interventions are more problematic than inclination-altering interventions (that leave reasons-responsiveness intact). This is because inclination-altering enhancements do not guarantee that the recipients will behave in one particular way. These enhancements leave open the possibility that the recipients will conform out of the motive of duty. Recipients of inclination-altering enhancement therefore retain the capacity to be praiseworthy.

This line of reasoning also has implications for the permissibility of voluntary non-reasons-responsive interventions. To be praiseworthy for the consequences of taking the moral enhancement pill, one must take the pill for the right reasons – one must be motivated by the good. This implies that an offender who consents to receive a moral enhancement intervention simply in order to be released early from prison (without caring about the value of behaving more morally) lacks the capacity to be praiseworthy for non-reasons-responsive conduct that flows from this intervention. If, as argued earlier, it is unethical to deprive someone of the capacity to be praiseworthy, then it may be unethical to administer a non-reasons-responsive intervention to an offender, if one knows the offender is consenting for the wrong reasons. For instance, the practice of administering

<sup>74</sup> Mitchell Berman, ‘Rehabilitating Retributivism’ (2013) 32 *Law and Philosophy* 83, 101.

disulfiram to prisoners who consent in order to reduce their jail time may be more problematic than Greely acknowledges in the quote cited above.

Another asymmetry between praise and blame arises in relation to an agent who endorses the intervention to begin with, but then comes to regret taking the pill. If an agent recklessly takes a bad-behaviour pill and immediately regrets this decision, she is intuitively still blameworthy for the consequences and would still be held criminally responsible for any offences she commits while under the pill's influence. This is at least true in the short-term. Intuitively, if the pill makes it impossible in the long-term to conform to morality, despite the agent's best efforts, then at some point the agent ceases to be blameworthy. Similarly, although an agent who recklessly becomes drunk is responsible for the effects of acute intoxication, she is not criminally responsible for the effects of delirium tremens induced by long-term alcohol consumption.

In contrast, imagine that an agent takes a moral enhancement pill initially for the right reasons, but immediately regrets this decision. It is not merely that she wishes to give in to temptation due to akrasia. She reflectively rejects the reasons that moved her to take the pill in the first place and no longer endorses her decision to take it. She now desires to behave immorally and endorses that desire, but the pill prevents her from acting on that desire despite her strongest efforts. It seems plausible that, given this change of heart, she ceases to be praiseworthy even for the short-term effects of the pill. Therefore, if we believe it is important to preserve the capacity to be praiseworthy, non-reasons-responsive interventions should ideally be reversible and should be periodically reviewed.

Finally, it should be noted that there are other objections (separate from the idea of moral worth) against certain interventions that undermine reasons-responsiveness. For instance it might be objected that interventions which closely resemble the fictional Ludovico method, inflict intolerable levels of suffering on the recipient. The Ludovico method also created such a strong aversion to violence that recipients were unable to defend themselves when necessary, which exposed them to the risk of serious harm. Another objection is that certain moral enhancements create the wrong kind of relationship between the state and offenders. If these enhancements give the state too much control over the offender's inner life then they give rise to unacceptable inequality.<sup>75</sup>

## 8 Conclusion

I have focused on the widely discussed intuition that biomedical moral enhancement might somehow undermine the moral worth of the recipient's future actions. I have drawn on the communication theory of punishment in order to shed new light on a potential source of this intuition, in relation to the moral bioenhancement of those who have committed serious criminal wrongs.

---

<sup>75</sup> Shaw (n 1).

The communication theory arguably implies a reformed offender's subsequent avoidance of committing similar crimes will only have moral worth if his improved behaviour has a certain causal history – it should emerge from an effortful process of repentance of the original crime. Biomedical interventions would be unacceptable on the communication theory if they undermined the offender's capacity to perform actions with moral worth. Such interventions would also be unacceptable if they allowed the state to avoid making a serious attempt to persuade the offender to take the effortful, repentant route to moral worth.

I concluded that a communication theorist could accept the consensual use of biomedical interventions as a) a means of facilitating repentance through moral communication and b) a means of helping to bring about law-abiding behaviour, provided that the offender has already undergone communicative punishment and that he is motivated to accept the intervention by his repentant desire to refrain from crime. On the communication theory of punishment, it would never be acceptable to use biomedical interventions to *replace* moral communication aimed at inducing repentance.

This conclusion has important practical implications, because it has the potential seriously to weaken the *motivation* for using such interventions within the criminal justice system. It has been claimed that biomedical interventions are likely to be 'politically attractive', if they can '(partially) replace' very financially costly forms of traditional punishment. However, communication theorists would oppose using biomedical interventions as a way of (even partially) replacing forms of punishment that involve moral dialogue. Communication theorists would insist that, if the state were to employ biomedical interventions to rehabilitate offenders, it must bear the financial cost of these interventions *in addition* to the financial cost of communicative punishment. It has also been suggested that biomedical interventions could be more humane – involving less suffering and allowing the offender to return to his normal life more quickly – than traditional forms of punishment.<sup>76</sup> While communication theorists, such as Antony Duff, criticise our current penal practices for inflicting much unnecessary suffering (including excessively lengthy prison terms), nevertheless such theorists insist that punishment of serious crimes must involve suffering, as repentance requires much time, effort, and emotional pain. If, as communication theory implies, biomedical interventions should only be used, if at all, *alongside* communicative punishment, then such interventions must not be employed in order to 'speed up' the process or to bypass the suffering, which according to this theory, is a necessary part of punishment.

---

<sup>76</sup> Douglas and Birks (n 2).

My aim in this article was not to defend the role of repentance in communication theories of punishment. Rather, my more modest aim was to explore the implications of this highly influential view for the question of whether biomedical moral enhancement would undermine the moral worth of offenders' future law-abiding behaviour and to argue that proponents of such interventions should address this issue.