

Chapter 19

Can Neurointerventions Communicate Censure? (And so what if they can't?)*

David Birks

[This is a draft prepared for David Birks & Thomas Douglas (eds.)
*Treatment for Crime: Philosophical Essays on Neurointerventions in
Criminal Justice*, Oxford University Press, 2018]

1. Introduction

According to some philosophers, a necessary condition of morally permissible punishment is that it communicates deserved censure for the offender's wrongdoing.¹ Let's call this the Communicative Condition of punishment. In this chapter, I consider whether the use of mandatory neurointerventions — the method of exerting a non-

* I am grateful to two anonymous reviewers for their perceptive objections and suggestions for the chapter. I would also like to thank Ian Carroll, Tom Douglas, Clare Heyward, Tippy Jackson, Julia Kulyomina, and Frej Klem Thomsen for helpful discussions on the topic. Finally, I would like to thank the participants at the Oxford-Denmark Workshop, at the University of Oxford, October 2016. I am particularly grateful to Katrien Devolder and Rune Klingenberg for providing several challenging objections as respondents to the chapter. It was in part written while supported by the Wellcome Trust grant number 100705/Z/12/Z, and the German Ministry of Education and Research grant number 01GP1311.

¹ R. A. Duff, *Punishment, Communication, and Community* (Oxford: Oxford University Press, 2001); John Tasioulas, 'Punishment and Repentance', *Philosophy* 81 (2006), pp. 279-322. See also Christopher Bennett, *The Apology Ritual* (Cambridge: Cambridge University Press, 2008); Andrew von Hirsch, *Censure and Sanctions* (Oxford: Clarendon Press, 1993). Throughout the chapter I treat wrongdoing and offending as synonymous, despite the fact that they are often distinct.

consensual physical, chemical or biological effect on the brain in order to diminish the likelihood of some forms of criminal offending — is compatible with the Communicative Condition.² I argue that it is not. If we accept the Communicative Condition, it follows that it is impermissible to administer mandatory neurointerventions on offenders as punishment.

It might be thought that such a conclusion has little import. It is frequently said that communicative theories of punishment have been persuasively shown to be inadequate justifications for punishment.³ Consequently, the fact that the use of neurointerventions as punishment fails to meet the Communicative Condition might be a concern for only a minority of legal philosophers. Nevertheless, I hope to show that the central thought behind communicative theories could have broader appeal than generally supposed. If we understand the Communicative Condition as a constraint on the type of punishment we administer, rather than a reason to punish, it has further significant implications for the permissibility of administering neurointerventions as punishment.

² For this definition of a neurointervention see: Thomas Douglas and David Birks, 'Introduction' in D. Birks and T. Douglas (eds.), *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice* (Oxford: Oxford University Press, 2017). Given the chapter is a part of this collection, I will not spend time elaborating on the various types of neurointerventions. However, I will stipulate that neurointerventions make the offender less likely to offend by creating or altering his desires through means that bypass his rational capacities. Compare this to educating an offender about the wrongness of an offence by providing an explanation of the reasons it is wrong, which has the result that he no longer desires to commit the offence. Although the education alters the offender's desires, it does so by engaging with his rational capacities. For a discussion on how neurointerventions operate see: Tom Douglas, 'Neural and Environmental Modulation of Motivation: What's the Moral Difference?', in D. Birks and T. Douglas, (eds.), *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice* (Oxford: Oxford University Press, 2017), Chapter 11, esp. Section 3.

³ For a selection of objections to communicative theories of punishment see: Russ Shafer-Landau, 'Can Punishment Morally Educate?', *Law and Philosophy* 10 (1991), pp. 189-219. See also: Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2012), Chapter 5.

My argument proceeds as follows. I begin in Section 2 by setting out two fundamental questions in the ethics of punishment: why should we punish? And how should we punish? I note that traditionally communicative theories are thought to provide an answer to the former question, but I try to show that they can also provide a partial answer to the latter. Following this, in Section 3, I provide four necessary and jointly sufficient conditions for punishment to meet the Communicative Condition. I consider in Section 4 whether administering mandatory neurointerventions as punishment could meet these four conditions. I propose that mandatory neurointerventions as punishment struggles to meet two of the conditions, namely, it fails to serve as a vehicle by which the offender can experience penance, and it fails to treat the offender as a rational agent. As a result, it is a form of punishment that fails to meet the Communicative Condition, and so, if one accepts this condition, it is impermissible. Finally, in Section 5 I consider whether it is permissible to offer an offender a neurointervention as a replacement for incarceration or in exchange for a shorter sentence. I note that this could meet the Communicative Condition although, somewhat oddly, only in cases where the neurointerventions have harmful effects. When we administer a harmless neurointervention following the offender accepting its offer, this would not meet the Communicative Condition.

Before I proceed to my argument, it will be helpful to delineate the scope of the chapter. It could be asked to what extent does it matter whether neurointerventions are an impermissible form of punishment. It might be thought that even if I showed that mandatory neurointerventions were impermissible as punishment, this does not say

anything about their permissibility for other non-punitive aspects of an offender's criminal sentence. Indeed, much of the debate on neurointerventions focuses on a different question, namely, the permissibility of administering mandatory neurointerventions as rehabilitation, and this is thought to be distinct from punishment.⁴ I suspect that if one accepts the Communicative Condition, and if the arguments in this chapter are successful, neurointerventions would also be impermissible when deployed as any mandatory part of the offender's criminal sentence. However, I am unable to explore adequately this issue here.

2. Communicative theories of punishment

Consider the following two central questions in the ethics of punishment: Why should we punish criminal offenders, and how should we punish them? Communicative theories of punishment can provide answers to both of these questions, a complete answer to the former, and a partial answer to the latter. The discussion in this section will clarify the role communicative theories can play in addressing both of these questions. This will then enable us to proceed to a clear discussion on the permissibility of neurointerventions as punishment.

Let's begin by addressing the first question, why should we punish criminal offenders? It might be easier to start by asking, why we should we not? An essential feature of punishment is that it involves hard treatment, understood as intentional harm inflicted on

⁴ For example, see: Tom Douglas, 'Criminal Rehabilitation through Medical Intervention: Moral Liability and the Right to Bodily Integrity', *The Journal of Ethics* 18 (2014), pp. 101-22.

an offender as a response to his wrongdoing.⁵ Accordingly, one reason not to punish is the common sense view that there is a reason not to harm persons.⁶ We do something at least *pro tanto* wrong when we inflict a harm.⁷ Now, we might think that once a person commits a criminal offence, new reasons are generated that defeat this original reason not to harm.⁸ Deterrence theorists hold that the reason not to harm is defeated by a reason to deter future offending.⁹ Retributivists of a certain stripe believe that by committing an offence, there is a reason to treat offenders as they deserve, and that they deserve to suffer.¹⁰ This desert reason defeats the reason not to harm. On communicative theories of punishment, offenders deserve to have censure communicated to them as a response to their wrongdoing.¹¹ The only way we can adequately communicate censure for

⁵ This is a standard definition of punishment. See: David Boonin, *The Problem of Punishment* (Cambridge: Cambridge University Press, 2008), Chapter 1. Other accounts avoid using harm as a part of their definition, and instead stipulate that the hard treatment requirement to simply involves being unpleasant, burdensome, or involve pain (See for instance, H. L. A. Hart, 'Prolegomenon to the Principles of Punishment', in *Punishment and Responsibility* (Oxford: Clarendon Press, 1968)). This would be unsatisfactory. As Boonin rightly notes, we could painlessly kill someone as a response to her wrongdoing, and it would be curious to deny this to be punishment because of its painless nature. Rather, a better definition of punishment involves harm, understood as a diminution of wellbeing. As even if we painlessly kill someone, it results in a diminution of wellbeing, as it deprives her of any future benefit she would have experienced had we not painlessly killed her.

⁶ There is a multitude of definitions of harm. For a good overview, see: Molly Gardner, 'On the Strength of the Reason against Harming', *Journal of Moral Philosophy* 14 (2015), pp. 73-87.

⁷ This reason might seem to be implausible. We harm people all the time. When Mallory breaks up with Paul, she inflicts a far greater harm on Paul than when a stranger punches Paul in the face. It is plausible to think the latter is morally objectionable, but the former is not. But this does not mean that the latter is less *pro tanto* wrong than the former. It might be that Mallory breaking up with Paul is a far more *pro tanto* wrong than the stranger's punch. Rather we can think that it is permissible for Mallory to break up with Paul this by appealing to an *all things considered* judgement about the value of Mallory's liberty concerning her personal relationships.

⁸ I should note that not all people will agree with this analysis. For instance, some consequentialists might think that we ought to punish people who have not yet committed a crime.

⁹ Anthony Ellis, 'A Deterrence Theory of Punishment', *The Philosophical Quarterly* 53 (2003), pp. 337-51.

¹⁰ Michael S. Moore, 'Justifying Retributivism', *Israel Law Review* 27 (1993), pp.15-49.

¹¹ Duff writes "To take wrongs seriously as wrongs involves responding to them with criticism and censure", Duff, *Punishment, Communication, and Community*, p. 81.

significant wrongdoing is through punishment.¹² We cannot sufficiently communicate censure for the wrongness of the offender's conduct without inflicting hard treatment. More will be said below about what it means to communicate deserved censure, but this should suffice for now.

It might be thought, however, that regardless of our answer to the first question, why should we punish offenders, our answer to the second, how should we punish, should be constrained by the central idea of communicative theory. We might think that even if the reason to punish is something quite different to the reason to communicate censure, we ought to punish offenders in a way that communicates censure. Communicating censure is a deontic constraint on the type of punishment it is permissible to administer.¹³ So for example, we might think that we ought to punish offenders in order to deter offending, but our method of deterrence ought to be constrained by a requirement that the method of deterrence communicates deserved censure. Our method of punishment is impermissible if we do not communicate deserved censure.¹⁴

Why might this be a deontic constraint? The central thought behind communicative theories is the necessity of the state treating its citizens as rational agents, inclusively as

¹² It might be the case that for minor wrongs, we can adequately communicate censure without inflicting harm – and therefore without punishment – but let's set aside this issue for present purposes.

¹³ Hart argues that our justifying aim of punishment should be constrained by side-constraints. Hart, 'Prolegomenon to the Principles of Punishment'.

¹⁴ Earlier I wrote that it is only a partial answer to the second question. This is because there may be other constraints on permissible punishment, besides the Communicative Condition. For example, there might also be a requirement that punishment is not cruel and unusual.

members of the normative community.¹⁵ Punishment that failed to do this is impermissible.¹⁶ This idea has an appeal that goes far beyond those who propose communicative theories of punishment. Indeed, it is plausibly compatible with, and justified by several liberal theories of legitimacy. For example, it seems consistent with the central Rawlsian idea that coercive power of the state is only permissible when it treats its reasonable citizens as free and equal.¹⁷ This echoes the requirement of the communicative theory that we treat offenders in ways that are consistent with fundamental liberal values.¹⁸ While much more needs to be said here, the Communicative Condition could be a way of capturing this central liberal thought.

Moreover, we might reject the communicative theories of punishment for reasons that do not apply once we understand them as providing a deontic constraint. For example, Victor Tadros rejects communicative theories of punishment on the basis that communicating deserved censure is not sufficiently important to justify the significant harm caused by punishment and criminal justice institutions.¹⁹ This objection to

¹⁵ Duff, *Punishment, Communication, and Community*, pp. 72, 144.

¹⁶ Peter Vallentyne argues against the existence of non-harm based deontic constraints on punishment in 'Neurointerventions, Self-Ownership, and Enforcement Rights', in D. Birks and T. Douglas, (eds.), *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice* (Oxford: Oxford University Press, 2017), Chapter 6. Although I think this is mistaken, I am unable to discuss it here.

¹⁷ John Rawls, *Political Liberalism* (New York: Columbia University Press, 2005), p. 134. For a Rawlsian discussion on how we ought to treat unreasonable citizens see: Jonathan Quong, *Liberalism without Perfection* (Oxford: Oxford University Press, 2011), Chapter 10.

¹⁸ See for instance the discussion in Duff, *Punishment, Communication, and Community*, pp. 76-77. This strongly echoes the Rawlsian requirement that political principles must be acceptable to reasonable persons who accept, inter alia, that political principles to treat citizens as free and equal. Rawls, *Political Liberalism*, p. 137. A neat summary of this requirement is provided by Quong, *Liberalism without Perfection*, pp. 37-8. However, it has been questioned whether communicative theories are compatible with liberalism see: Andrew von Hirsch and Andrew Ashworth, *Proportionate Sentencing* (New York: Oxford University Press, 2005), p. 93.

¹⁹ Tadros, *The Ends of Harm*, Chapter 5.

communicative theories does not touch them being deployed as a deontic constraint on punishment. For even if they are insufficiently strong to permit punishing offenders, one could hold that they are sufficient to provide constraints on other non-communicative reasons to administer punishment.

So we can arrive at the Communicative Condition from two directions.²⁰ We might think there is a reason to communicate deserved censure for wrongdoing, and that this is a reason to punish. This reason to punish defeats the reason not to harm persons, whereas other proposed reasons to punish do not defeat this reason not to harm.²¹ Hence why it would be a necessary condition of morally permissible punishment. Alternatively, we might think that there are other distinct, non-communicative reasons to punish, but our method of punishing in order to act in accordance with these reasons ought to communicate deserved censure. If it does not, it is impermissible punishment.

I will not attempt to provide a defence of the Communicative Condition here. Rather, my aim is more modest. By setting out how communicative theories can provide answers to these two distinct questions in the ethics of punishment, I show that the stakes are higher as to whether a method of punishment could meet the Communicative Condition. Even if we reject communicative theories in terms of providing reasons why we ought to punish offenders, we might accept them as a constraint on how we ought to punish.

²⁰ It is of course, possible to hold both views, that our reason to punish is to communicate deserved censure, and that our harming is ought to be constrained by communication.

²¹ Hence why it is a necessary condition of justified punishment. Some might think that it is not a sufficient reason, and it deterrence is also additionally required. See, von Hirsch, *Censure and Sanctions*.

3. Four conditions of communicative punishment

There are four necessary and jointly sufficient conditions for punishment to meet the Communicative Condition:

- (a) The Hard Treatment Condition
- (b) The Desert Condition
- (c) The Penance Condition
- (d) The Rationality Condition

Although proponents of communicative theories of punishment do not set out their view this way, I think it is accurate and helpful to frame it as such.²² In this section, I proceed to set out each condition, and then in the following section I consider whether administering a mandatory neurointervention as punishment could meet them.

Let's begin with the Hard Treatment Condition. Recall earlier I wrote that in order for something to be punishment, it must inflict hard treatment — a harm — on an offender as a response to his wrongdoing. For communicative theories, the hard treatment must involve an experiential harm, by which I mean that the offender has to be aware that the harm is taking place as a response to his wrongdoing. For example, we could imagine administering a painless poison to a person without his knowledge, and this reduced the length of his life. This would still plausibly be hard treatment, as it would be a significant

²² Some proponents of communicative theory may reject one or more of these conditions. These are based on what is in my view the most fully developed and persuasive account of communicative theory, namely, that set out in Duff, *Punishment, Communication, and Community*.

harm, and could be punishment, as it is a response to his wrongdoing, but it would not be the appropriate type of hard treatment that could communicate censure.²³

It might be questioned why we could not communicate deserved censure without hard treatment? One reason is that the censure communicated without hard treatment would be insufficiently sincere.²⁴ To see this, imagine that rather than punishing a person for several violent murders, we instead publically declared that she had acted significantly wrong, with no further sanction. It seems that this would not adequately communicate the extent of the wrongness of the offences committed. In order for our censure to be understood as sincere, it is necessary for the offender to experience hard treatment.²⁵ As Tasioulas writes “punishment enables us properly to evince the sincerity of our condemnation”.²⁶

²³ It could also be questioned whether an offender would have to experience the hard treatment, in addition to knowing that the hard treatment has taken place. Take the poison example, does the Hard Treatment Condition require that the offender needs to experience the pain of the poison in addition to knowing that the poison has been administered as a response to his wrongdoing? I suspect knowledge that the harm has taken place is insufficient when we consider how the Hard Treatment Condition supports the Penance Condition. While knowledge that one has been harmed might be sufficient for the censure to be understood as sincere, it is the experiential element of the hard treatment that focuses the offender’s attention on his wrongdoing.

²⁴ Duff, *Punishment, Communication, and Community*, p. 28; Tasioulas, 'Punishment and Repentance', p. 296; Igor Primoratz, 'Punishment as Language', *Philosophy* 64 (1989), p. 200.

²⁵ It is sometimes objected that communicative theories should be able ‘say it with flowers’, and that hard treatment is not required to communicate deserved censure. See Scanlon. 1988: 214. But as Matravers has noted, this is an inadequate (and somewhat) odd reply. After all, the meaning of giving a person some flowers is not one of censure. Rather, we understand it a pleasant thing to do for someone. Admittedly, we could imagine a possible society where giving someone flowers is a grave insult. The fact remains though, in this society, giving someone flowers does not communicate censure, whereas inflicting harm plausibly does. Matt Matravers, *Justice and Punishment: The Rationale of Coercion* (Oxford: Oxford University Press, 2000), p. 257.

²⁶ Tasioulas, 'Punishment and Repentance', p. 296.

Let's now turn to the Desert Condition. The details of this condition need not concern us too much here, but I should note that in order for the punishment to communicate *deserved* censure for the offender's wrongdoing, at least two factors need to hold.²⁷ First, in order for a person to deserve censure communicated to him, he must have committed an offence. Non-offenders do not deserve censure, and so the punishment of innocents is impermissible. The second factor is that the extent of the censure needs to be proportionate to the offence. Although much has been written about proportionality and punishment, it is not necessary for me to engage deeply with this requirement here.²⁸ However, I will remark that proportionality is often thought to have at least two domains: quantitative and qualitative. The former is a requirement that the severity of censure tracks the gravity of the offence. For example, a murderer deserves greater censure than a shoplifter.²⁹ The latter domain is that the type of punishment needs to be appropriate for the offence.³⁰ Many philosophers of punishment believe that there needs to be some

²⁷ One respect punishment communicates deserved censure is through its symbolic meaning. For instance, the diminution of liberty through incarceration symbolises condemnation of an offender's behaviour in modern Western societies because of the value placed on liberty. For instance, see: Dan M. Kahan, 'What's Really Wrong with Shaming Sanctions', *Texas Law Review* 84 (2006), p. 2077; Dan M. Kahan, 'What Do Alternative Sanctions Mean?', *The University of Chicago Law Review* 63 (1996), pp. 591-653. In this chapter, I will not discuss whether mandatory neurointerventions could come to symbolise condemnation. For even if they could, they would not meet the Communicative Condition for the reasons discussed below.

²⁸ For Duff's discussion on proportionality and communicative punishment see: Duff, *Punishment, Communication, and Community*, pp. 141-43. See also Andrew von Hirsch, 'Proportionality in the Philosophy of Punishment', *Crime and Justice* 16 (1992), pp. 55-98.

²⁹ Importantly, Duff notes "Rather than saying...that a communicatively appropriate punishment must, above all, be proportionate to the crime, we should therefore say that communicatively appropriate punishment must, among other things, not be disproportionate to the crime", Duff, *Punishment, Communication, and Community*, p. 143.

³⁰ Duff notes that punishment "must address only those aspects of her conduct and life that are directly implicated in her crime". Duff, *Punishment, Communication, and Community*, p. 148.

qualitative connection between the punishment and the offence, and this could be understood as the requirement that the punishment is qualitatively deserved.³¹

The third condition is the Penance Condition, namely that punishment must serve as a vehicle by which the offender can experience penance.³² This might sound a little mysterious. The central idea here is that punishment should provide the opportunity for the offender to think about his wrongdoing. Duff writes that punishment “provides a structure within which, we hope, he will be able think about the nature and implications of his crime, face up to it more adequately than he might otherwise (being human) do, and so arrive at a more authentic repentance”.³³ The punishment “is a way of trying to focus [an offender’s] attention on [his] crime”. And enables the offender “To recognize and repent the wrong [he has] done”.³⁴ By punishing the offender, we provide circumstances in which he can undergo penance.³⁵

The final condition is the Rationality Condition. In order to meet this condition, the punishment needs to engage the offender as a responsible agent with moral rational persuasion, with the hope that the offender will accept that his behaviour was wrong.³⁶

³¹ For a discussion and defence of such a requirement see: Jeremy Waldron, 'Lex Talionis', *Arizona Law Review* 34 (1992), pp. 25-51.

³² The phrasing of describing punishment as a *vehicle* is taken from Matravers, *Justice and Punishment: The Rationale of Coercion*, p. 257.

³³ Duff, *Punishment, Communication, and Community*, p. 108.

³⁴ Both references at Duff, *Punishment, Communication, and Community*, p. 108.

³⁵ It might be the case that in order for the Penance Condition to be met, the Hard Treatment Condition also needs to be met. For unless the punishment meets the Hard Treatment Condition, the punishment would be unable to focus the offender’s attention on his crime, and thus provide circumstances he can undergo penance. This point requires the hard treatment to be of the type that the offender is both aware has taken place and also experiences.

³⁶ For a neat and clear statement of this requirement, see Antony Duff, ‘Legal Punishment’, *Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), Edward N. Zalta (ed.), URL =

The detail of this condition can be fleshed out by comparing the difference between communicative and expressivist theories, namely, the latter does not require the punishment to treat the offender as a rational agent.³⁷ For example, we could express censure by administering a beating as punishment. When administered as a response to wrongdoing this would be clear that we disapprove of the wrongdoing, but while it would express censure, it would not *communicate* it. An expressive punishment may have an intended effect on the offender, namely it may reduce recidivism, but crucially the effect “could be entirely nonrational—it need not be mediated by the recipient’s reason or understanding”.³⁸ Rather, in order for punishment to *communicate* censure and meet the Rationality Condition, it must aim through a process of rational moral persuasion that the offender “does what is right because she sees it to be right...that it can be achieved only by a process of rational moral persuasion”.³⁹

How punishment could do this might still seem vague, and further details will be provided in my discussion on whether mandatory neurointerventions meet the Rationality Condition. But in any case, it has been objected that the Penance and Rationality Conditions are too similar to require two separate conditions. The Penance Condition provides a structure for the offender to reflect and focus on his wrongdoing and to repent,

<http://plato.stanford.edu/archives/sum2013/entries/legal-punishment/>, Section 6. See also, Duff, *Punishment, Communication, and Community*, p. 108.

³⁷ Duff, *Punishment, Communication, and Community*, p.144.

³⁸ Duff, *Punishment, Communication, and Community*, pp. 79-80. See also, Duff, *Punishment, Communication, and Community*, p. 81.

³⁹ Duff, *Punishment, Communication, and Community*, p. 81. The offender could also be provided with the *prudential* reasons not to offend, such as the reason to avoid hard treatment, if one holds that providing the offender with both moral and prudential reasons meets the Rationality Condition. For this view see: von Hirsch and Ashworth, *Proportionate Sentencing*, p. 25. I do not need to take a stand on this here.

whereas the Rationality Condition requires that the punishment rationally persuades the offender concerning the wrongdoing of his behaviour. These are two distinct requirements of the Communicative Condition. Indeed, it is possible for a punishment to meet only one of these two conditions. For example, suppose as a response to wrongdoing we detain an offender, with no explanation regarding the reasons why the detention is taking place. This form of punishment would meet the Penance Condition, but it would not meet the Rationality Condition. It would enable the offender to reflect on his wrongdoing, but it would not involve the process rational persuasion necessary to meet the Rationality Condition.

4. Neurointerventions and the communicative condition

In the previous section, I set out four necessary and jointly sufficient conditions for punishment to meet the Communicative Condition. In this section, I consider whether administering a mandatory neurointervention as punishment could meet these conditions. While it is plausible to think that the first two conditions can be met when we administer mandatory neurointerventions as punishment, I argue that doing so would not meet the Penance and Rationality Conditions, and so according to the Communicative Condition, it is impermissible to do so.

Let's consider each condition in turn, beginning with the Hard Treatment Condition. One objection is that when we administer neurointerventions to offenders as a response to their wrongdoing, we do not meet the Hard Treatment Condition because

neurointerventions are *too good* for the offender.⁴⁰ Indeed, given that Hard Treatment is a necessary condition of punishment, it could be questioned whether a neurointervention could be punishment at all. This is because by diminishing the likelihood that the offender acts wrongly, the neurointervention confers a significant benefit on the offender. It is a plausible view of what makes a life go well that performing disvaluable activities is bad for the person who performs them.⁴¹ Doesn't this mean a neurointervention is too good for the offender to count as punishment?

This objection is unsuccessful because it requires us to hold an implausible version of Hard Treatment as a condition of punishment. Namely it needs hard treatment to inflict an *all things considered* harm, rather than merely a pro tanto harm on an offender.⁴² To see why this is implausible, imagine a person, Victoria, who spends her life committing petty crime, such as stealing, littering, and vandalism.⁴³ Eventually Victoria gets caught, and is sent to prison. There, she reflects on her past crimes, and she becomes motivated to change her life in pursuit of valuable activities. She then has an enriching life helping others. As a consequence of the incarceration then, Victoria's life has gone significantly better than it otherwise would have gone. If we hold that Hard Treatment needs to cause an *all things considered* harm for her, then it follows that Victoria was not punished. It

⁴⁰ Jeff McMahan plausibly notes this. See Jeff McMahan, 'Moral Liability to 'Crime-Preventing Neurointervention'', in D. Birks and T. Douglas, (eds.), *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice* (Oxford: Oxford University Press, 2017), Chapter 5.

⁴¹ For example, see: Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), Joseph Raz, 'The Role of Well-Being', *Philosophical Perspectives* 18 (2004), pp. 269-94. It is also likely that on an informed desire account of wellbeing, the desires to offend are not autonomous. I will say more about the non-autonomous element of offending below.

⁴² A similar point is also made by Nathan Hanna, 'Facing the Consequences', *Criminal Law and Philosophy* 8 (2014), p. 593.

⁴³ I make the same point with this example in 'Benefitting Offenders' (unpublished manuscript).

would also mean that successfully rehabilitated offenders would not have been punished if the rehabilitation resulted in their lives going better. But this would be an unusual account of punishment. However, if the Hard Treatment Condition only needs to be a *pro tanto* harm, we can claim that Victoria was punished by incarceration. For example, one *pro tanto* harm inflicted on Victoria is the diminution of her liberty.

What then are the *pro tanto* harms of administering a mandatory neurointervention as a response to offending? First, there is a plausible harm due to the infringement of the offender's bodily integrity.⁴⁴ That is, it is harmful in one respect for a person to have something done to his body, or with his body, without his consent.⁴⁵ It is also likely that the neurointervention will have harmful side effects. For example, chemical castration often results in an increase of fat tissue, and reduces all sexual desires, not only the disvaluable ones.⁴⁶ Given these side effects, it seems plausible to hold that a mandatory neurointervention as punishment could be genuinely harmful.⁴⁷ Moreover, these harms are experiential, in that the offender could be aware of them, and so meet the Hard Treatment Condition.⁴⁸

⁴⁴ I follow Judith Jarvis Thomson by using *infringe* to mean a justified behaviour contrary to a right. See: Judith Jarvis Thomson, *The Realm of Rights* (London: Harvard University Press, 1990).

⁴⁵ This definition is adapted from Cécile Fabre, *Whose Body Is It Anyway? Justice and the Integrity of the Person* (Oxford: Clarendon Press, 2006), p. 110.

⁴⁶ Peer Briken, Martin P. Kafka, 'Pharmacological Treatments for Paraphilic Patients and Sexual Offenders', *Current Opinion in Psychiatry* 20 (2007), pp. 609–13.

⁴⁷ A further purported harm of neurointerventions is that they infringe the offender's mental integrity. See: Jan Christoph Bublitz and Reinhard Merkel, 'Crimes against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination', *Criminal Law and Philosophy* 8 (2014), pp. 51-77.

⁴⁸ Of course, we can infringe the offender's bodily integrity without meeting the experiential requirement by administering the neurointervention when he would be unaware that it was taking place, for instance, while he is asleep.

I will now turn to the Desert Condition. It is relatively straightforward to see that a mandatory neurointervention could meet the quantitative dimension of this condition, and as such, I will not dwell on it here.⁴⁹ But it is worth noting one point about the qualitative dimension. It might be thought that an advantage of using mandatory neurointerventions as punishment is that they are qualitatively similar to the offence. As such, they are likely to meet a qualitative proportionality requirement of desert. By administering a mandatory neurointervention, a sexual offender could be punished in a way that diminishes his sexual offending by receiving chemical castration.⁵⁰ A drug offender is punished in a way that eliminates his desire to take drugs.⁵¹ This might suggest there is a far tighter link between punishment and offence than many standard forms of punishments. Nevertheless, it is important to note that this is often not the case. As stated earlier, the hard treatment of administering a mandatory neurointervention include its side-effects, and the infringement of the offender's bodily integrity. However, consider the case of administering a mandatory neurointervention on an offender as a response to a drug offence. Here the hard treatment is quite different to the offence. The hard treatment involves infringing the bodily integrity of the offender, and the harmful side-effects,

⁴⁹ I will not defend here that the pro tanto harmfulness of mandatory neurointerventions can be used to communicate quantitatively proportionate censure for at least some offences. For a persuasive defence that mandatory neurointerventions can inflict proportionate deserved harm, see Jesper Ryberg, 'Neuroscientific Treatment of Criminals and Penal Theory', in D. Birks and T. Douglas, (eds.), *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice* (Oxford: Oxford University Press, 2017), Chapter 9. If the neurointerventions can inflict proportionate deserved harm, and in part it is the harm of the punishment that determines the severity of the censure, then this is a reasonable assumption to hold.

⁵⁰ One common drug used for this purpose is a luteinizing hormone-releasing hormone agonist. See: Peer Briken, Andreas Hill, Wolfgang Berner, 'Pharmacotherapy of Paraphilias with Long-Acting Agonists of Luteinizing Hormone-Releasing Hormone: A Systematic Review', *The Journal of Clinical Psychiatry* 64 (2003), pp. 890-7.

⁵¹ For example, Vivitrol could be administered which significantly reduces the enjoyment of heroin and thus reduce the desire to use the drug. See: <https://www.vivitrol.com/About/HowVIVWorksOPD>.

whereas the wrongness of the offence does not involve these harms.⁵² It is unlike the qualitatively similar punishment of the death penalty for murderers, where the hard treatment involves death, something the offender inflicted on another person. It might be easier to defend the qualitative similarity when considering administering mandatory neurointerventions as a response to some sexual offences, which also are in part wrong due to violations of bodily integrity. However, the point remains that the qualitative proportionality of neurointerventions is not as neat in many cases as it might first seem.

Let's now consider the Penance Condition.⁵³ How could a neurointervention serve as a vehicle by which the offender can experience penance? For those who committed sexual offences who are unable to focus due to being overcome by sexual thoughts, the

⁵² The drug might have harmful side-effects, but certainly it is not clear that it is the side-effects that account for the fact they are offences.

⁵³ An anonymous reviewer objected that the Penance Condition is incompatible with my statement of the Hard Treatment Condition. In particular, the Hard Treatment Condition was thought to be problematic because it requires the intentional infliction of harm, rather than a burden. However, once the punishment is accepted as a penance, the objection followed, it remains a burden but it no longer inflicts even a *pro tanto* harm on the offender. Rather, it becomes something good for the offender. This objection can be circumvented by clarifying the nature of a *pro tanto* harm. Even if something is voluntarily accepted, it does not eliminate its *pro tanto* harmfulness (though as discussed in Section 5, in one respect its harmfulness will be diminished). Likewise, the fact that something is good for me does not eliminate its *pro tanto* harmfulness. To illuminate these points, imagine that I give my informed consent to receive a lifesaving, but painful operation. It would be beneficial for me to have the operation performed. However, the operation would still be *pro tanto* harmful, regardless of it being *all things considered* good for me by saving my life. It is *pro tanto* harmful because it makes me worse off in one regard, namely, its painfulness, despite its overall benefit. Compare the painful operation to an equally successful painless operation. If we didn't hold that the pain of the operation was *pro tanto* harmful, then it is not clear on what grounds we would think it better to perform the painless operation. Whereas the fact that one of the operations makes the patient worse off in one regard (namely by experiencing pain) is clear grounds for this. Similarly, even if a punishment is accepted as a penance, it remains *pro tanto* harmful, regardless of whether the punishment then becomes *all things considered* good for the offender. Hence why the Hard Treatment Condition as stated is compatible with the Penance Condition.

neurointervention might increase the offender's ability to reflect on their wrongdoing.⁵⁴ This might be necessary but insufficient by itself in order for such offenders to experience penance. However, for the many other types of offenders, it is not obvious that neurointerventions could play any role in enabling the offender to experience penance. In order to meet the Penance Condition, the punishment needs to bring the offender to think about his wrongdoing. It seems that administering a mandatory neurointervention does not do that – it simply makes the offender less likely to offend. So, for example, a person who commits a sexual offence may or may not think that sexual offences are wrong, but administering a testosterone-lowering neurointervention does not address this at all. Rather it simply makes the sexual offender less likely to commit sexual offences. As such we should doubt that it could meet the Penance Condition.

Similar difficulties are faced when we consider whether mandatory neurointervention can meet the Rationality Condition.⁵⁵ After all, when we administer a neurointervention, we are not providing the offender reasons not to offend, but rather we make the offender less likely to offend through means that create or alter his desires by bypassing his rational capacities. This occurs regardless of the offender's understanding and acceptance of the reasons concerning the offence. The punishment is thus non-rational. It is not mediated

⁵⁴ I thank Katrien Devolder for helpfully suggesting this respect in which neurointerventions could enable penance. For a selection of real cases, see Malcolm Alexander et al., 'Controversies in Treatment: Should a Sexual Offender Be Allowed Castration?', *BMJ* 307 (1993), pp. 790-3.

⁵⁵ Elizabeth Shaw also makes the similar observation that neurointerventions fail to engage offenders in a rational moral dialogue. See Elizabeth Shaw, 'Direct Brain Interventions and Responsibility Enhancement', *Criminal Law and Philosophy* 8 (2014), p. 13.

by the offenders understanding of the wrongness.⁵⁶ While it could *express* censure, it does not communicate it. It circumvents his rational agency, and as a consequence, does not communicate censure. Let me elaborate on this. The neurointervention does not alter the offender's view on the wrongfulness of the behaviour. It does not attempt to rationally persuade the offender that it is wrong to offend. It simply reduces the likelihood that the offender will offend by creating or altering his desires through means that bypass his rationality. As such, it does not meet the Rationality Condition.

One response would be to argue that a mandatory neurointervention addresses an offender with rational moral persuasion because it *restores* his rational capacities.⁵⁷ After all, it seems reasonable to assume that many offences are not the result of an autonomous choice formed as a result of the offender's rational capacities, but rather, they occur in spite of the offender's autonomous desires. Often violent offenders act impulsively, for example. There is evidence that suggests that this is the cause of much offending.⁵⁸ As a consequence, it could be claimed that by administering the neurointervention, we would be respecting the rational agency of the offender.

While it is true that in some cases a neurointervention may restore the rational capacities of the offender, and enable him to engage in a process of rational moral persuasion, by itself this is insufficient to meet the Rationality Condition. It is not a condition that the

⁵⁶ Duff, *Punishment, Communication, and Community*, p. 79.

⁵⁷ This point has been made by Thomas Douglas, Pieter Bonte, Farah Focquaert, Katrien Devolder, Sigrid Sterckx, 'Coercion, Incarceration, and Chemical Castration: An Argument from Autonomy', *Journal of Bioethical Inquiry* 10 (2013), pp. 393-405.

⁵⁸ See for instance, Alexander et al., 'Controversies in Treatment'.

punishment makes the offender better able to act in accordance with procedural rationality or to enable the offender to act autonomously. Rather the condition is a requirement that the punishment engages the offender in a process of rational moral persuasion, and it is not clear how administering a neurointervention does this, even if in some cases it could restore the rational capacities of the offender.

Now, it may not be clear how other forms of punishment typically considered to be permissible could meet these conditions, but this will be dealt with below. Before I do that, let's first consider a possible solution that enables mandatory neurointerventions to meet the Penance and Rationality Conditions. It might be proposed that we could change the *process* of administering the mandatory neurointerventions so that it meets these conditions. Suppose that the standard process to receive a mandatory neurointervention would be for the offender to be required to attend a clinic on a certain date and time, and receive an injection from a medical professional. This method of administering a mandatory neurointervention does not bring the offender to think about his wrongdoing. Nor does it rationally persuade the offender that it is wrong to offend.

Now suppose that we supplemented the process of administering the mandatory neurointervention in order to meet the Penance and Rationality Conditions. We could imagine an alternative process involving a ceremony where it is explained to the offender why the neurointervention is being imposed. The ceremony could explicitly state that the purpose of the neurointerventions is to hope that the offender will reflect on his wrongdoing. The hope is that while the offender is being punished by receiving the

neurointervention, he will reflect on his wrongdoing, and that eventually he will repent for his wrongdoing. Following the offender being administered the mandatory neurointervention, we might also require him to spend time in a quiet area without distraction in order to reflect on his wrongdoing. It might be suggested that we could meet the Rationality Condition by supplementing the process of administering the mandatory neurointervention with some form of discussion, where the offender is provided the reasons for the wrongness of his offence. Let's call these additions to a method of punishment, *communicative supplements*, for short.

The problem with claiming that communicative supplements enable a punishment to meet the Communicative Condition is the troublesome implication that *any* type of punishment could meet the condition, no matter how cruel or degrading.⁵⁹ Imagine the punishment of public flogging, with communicative supplements, such as a ceremony explaining how the flogging is a way of trying to focus the offender's attention on his crime, and a discussion concerning the wrongness of his offending. If a punishment can meet the Penance and Rationality Condition by communicative supplements then it seems the punishments such as flogging would be permitted in principle by the Communicative Condition.⁶⁰ This permissive view of punishment is objectionable, and it certainly is contrary to the prevailing view of those who hold the Communicative Condition. Duff is

⁵⁹ This might only be worrying if one holds the Communicative Condition providing reasons to punish rather than as a constraint on punishment. For if it is merely the latter there could be other conditions that rule out degrading punishment.

⁶⁰ Even Duff, who opposes much of our current practices of incarceration accepts that imprisonment could be justified if it reconciles the offender with the community by serving as a penance. Duff, *Punishment, Communication, and Community*, pp. 148-52.

clear that not all punishment can communicate deserved censure, even in principle.⁶¹ Is Duff justified in restricting the types of possible punishments that can in principle be permitted? If so, on what basis?

I can think of two possible justifications. One would be to claim that some forms of hard treatment degrade and humiliate the offender, and this undermines any communicative supplement that attempts rational persuasion. It also prevents the offender from reflecting on his wrongdoing, and results in blocking out the reasons that we aim to illuminate with punishment.⁶² For example, imagine that we torture an offender as a response to his wrongdoing. Even if the torture is administered with a communicative supplement in an attempt to rationally persuade the offender concerning his wrongdoing, the nature of the hard treatment means that the offender would not be rationally persuaded not to offend.⁶³ Even if post-punishment he does not reoffend, it would be due to a non-rational response, rather than rational persuasion.⁶⁴ The trauma of undergoing the torture could also prevent the offender from reflecting on his wrongdoing. Admittedly this relies on empirical claims concerning how offenders respond to punishments of a certain type, but it at least seems plausible basis for restricting some forms of punishment.

⁶¹Duff, *Punishment, Communication, and Community*, p. 144. Duff approvingly cites Jeffrey Murphy for an argument that torture cannot be in principle be a method of communicating censure. See: Jeffrie Murphy, 'Cruel and Unusual Punishments', in *Retribution, Justice, and Therapy: Essays in the Philosophy of Law* (London: Pallas Paperbacks, 1979), pp. 233-7.

⁶² I thank Julia Kulyomina for suggesting this possible response. Duff writes that punishment that involves "trauma, indignity and degradation" cannot in principle communicate censure. Duff, *Punishment, Communication, and Community*, p. 144.

⁶³ Though to be clear, the Penance and Rationality Conditions do not require that the offender actually thinks about his wrongdoing, or is actually rationally persuaded, merely that it has to provide the opportunity to do so.

⁶⁴ It would be similar to the type of punishment Jean Hampton refers to when she writes of punishment that "affect us in the way that electrical fences affect animals - they would deter us, nothing more". Jean Hampton, 'The Moral Education Theory of Punishment', *Philosophy and Public Affairs* 13 (1984), p. 213.

A second possible justification for restricting the types of punishment that can communicate deserved censure is the view that the hard treatment itself has to play at least some role in rationally persuading the offender. This could be seen by considering Duff's discussion of criminal mediation, which Duff holds to be an ideal example of communicative punishment.⁶⁵ Broadly, this is where the offender is "vividly confronted, through his victim's voice, with his crime. But he will also have a chance to explain himself".⁶⁶ Here the hard treatment would include the unpleasant "process of being confronted with and having to listen to the victim...as is the remorse that that process aims to induce".⁶⁷ But crucially perhaps, the hard treatment itself rationally persuades the offender that his conduct was wrong.⁶⁸ Hence why the criminal mediation form of punishment would meet the Rationality Condition, without need for a communicative supplement.

While this justification has prima facie appeal, it rules out almost all forms of punishment besides criminal mediation. It is difficult to think of a further method of punishing offenders that involve rational persuasion brought on by hard treatment. Yet Duff thinks we should in certain circumstances communicate censure through other forms of

⁶⁵ See: Duff, *Punishment, Communication, and Community*, pp. 92-99.

⁶⁶ Duff, *Punishment, Communication, and Community*, p. 93.

⁶⁷ Duff, *Punishment, Communication, and Community*, p. 97.

⁶⁸ I think we could also doubt whether the trauma of talking with one's victims would rationally persuade an offender about his wrongdoing. It seems just as likely to think that it is a non-rational response brought about by sympathy for the victim, which could be wholly independent of the reasons not to offend. But given the further problems with this solution, this further difficulty does not matter here.

punishment beside criminal mediation.⁶⁹ For example, while incarceration could meet the Hard Treatment, Desert, and Penance Conditions, it is not clear how it could meet the Rationality Condition unless we could also add a communicative supplement that involves rationally persuading the offender that his conduct was wrong.

Let's now consider how these two possible justifications relate to mandatory neurointerventions. Suppose we hold that any form of punishment can communicate censure with the right form of communicative supplements, as long as the punishment does not degrade or humiliate the offender. We would need to show that administering a mandatory neurointervention did not do this. But given the importance people attach to non-interference with their bodies without consent, and the extent people consider it a significant wrong when one is non-consensually interfered with, it is reasonable to think that receiving a mandatory neurointervention could be degrading and humiliating for the offender.⁷⁰

Suppose alternatively we held that the hard treatment itself had to rationally persuade the offender about the wrongness of his behaviour, in the same manner as in criminal mediation cases. This would also rule out administering mandatory neurointerventions as punishment. As noted above, in no sense do they rationally persuade the offender about the wrongness of offending. The neurointervention simply reduces the likelihood that the

⁶⁹ It would be inappropriate for offences where meeting the offender is a traumatic experience for the victim, for example.

⁷⁰ One problem that I am unable to address here is that different people are likely to find different punishments as humiliating, and this would seem to have the implication that the same punishment could be permissible when administered on one person but impermissible on another person. For a helpful discussion on a related matter see: Adam J. Kolber, 'The Subjective Experience of Punishment', *Columbia Law Review* 109 (2009), pp. 182-236.

offender will behave wrongfully, through means that bypass their rational capacities. Administering a mandatory neurointervention as punishment would not meet the Rationality Condition on this justification either.

In summary then, administering mandatory neurointerventions as punishment clearly meets two of the four necessary and jointly sufficient conditions of the Communicative Condition. It could be Hard Treatment as it inflicts a pro tanto experiential harm on the offender as a response to his wrongdoing, through infringing the offender's bodily integrity as well as the harm of experiencing side effects of the neurointervention. It could meet the Desert Condition as it could be administered on those who deserve it, while in a limited number of cases it could also be qualitatively proportionate given the close link between the punishment and the offence. However, things became less clear when we turned to the Penance and Rationality Conditions. Mandatory neurointerventions do not bring the offender to think about his offence and experience penance. They do not attempt to rationally persuade the offender with reasons not to offend, but rather simply makes the offender less likely to offend by means that bypass his rationality. I then considered the solution of employing communicative supplements in order to enable mandatory neurointerventions to meet the Communicative Condition. However, I noted that this would have the unpalatable implication that any type of punishment could be permitted on this Condition, and this is clearly not the intention of its proponents. I then considered two possible justifications for restricting the type of punishment that could meet the Communicative Condition. I showed that neither

justification meant that administering mandatory neurointerventions would meet the Communicative Condition.

5. The offer of a neurointervention

Some people think that there is a crucial difference in terms of permissibility between administering a mandatory neurointervention on an offender as punishment, and offering the offender a choice between incarceration or a neurointervention as punishment. Let's assume that an offender could provide valid consent to such an offer, and we do not do anything wrong by making this offer.⁷¹ It is possible that this punishment could be consistent with the Communicative Condition. When we offer the neurointervention to the offender, we are addressing him as a rational agent. By offering the offender a choice of his punishment, and it is accepted by the offender, it involves the offender with the rational process that seeks a response mediated by the other person's understanding. The offender's agreement to undergo the neurointervention could form part of the offender's response.⁷² By agreeing to receive the neurointervention, it could be the means by which the offender repents his wrongdoing.⁷³

⁷¹ This view has been contested by numerous authors. See: Kari A. Vanderzyl, 'Castration as an Alternative to Incarceration: An Impotent Approach to the Punishment of Sex Offenders', *Northern Illinois University Law Review* 15 (1994), pp. 107-140, William Green, 'Depo-Provera, Castration, and the Probation of Rape Offenders', *University of Dayton Law Review* 12 (1986), pp. 16—17. For a comprehensive overview of this debate see: Jonathan Pugh, 'Coercion and the Neurocorrective Offer', in D. Birks and T. Douglas (eds.), *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice* (Oxford: Oxford University Press, 2017), Chapter 4.

⁷² The reason why the offender chooses the neurointervention is important here. We do not want the offender to choose the neurointervention because it is less harmful, but for penance reasons.

⁷³ Duff is clear that he accepts that the fact an offender agrees to be punished does not mean that it is no longer punishment. Duff, *Punishment, Communication, and Community*, p. 97.

While an offer of a neurointervention could meet the Penance and Rationality Condition, the Hard Treatment Condition is more difficult. The fact that the offender consents to receive the neurointervention eliminates one of the major harms of receiving a neurointervention, namely, it no longer infringes the offender's bodily integrity. Admittedly, our existing neurointerventions have in harmful side-effects, and so for the foreseeable future, the Hard Treatment Condition could be met, even when such treatment would not infringe the offender's bodily integrity. This reduction in harm might be relevant when considering the proportionality of our punishment, but it could still be hard treatment.⁷⁴ However, by depending on the side effects of the neurointervention to provide the hard treatment, it has the unusual implication that we could only offer neurointerventions that had harmful side-effects as permissible punishment. It would be wrong to offer harmless neurointerventions as a proposed punishment.⁷⁵ Suppose developments in neuroscience mean that in the future there are neurointerventions that have the sole effect that the offender becomes less likely to offend, with no harmful side effects. Now, suppose we offered an offender the choice between incarceration or a side-effect free neurointervention, and the offender chose the neurointervention. In this case, because the offender consents to receive it, there is no infringement of bodily integrity, and if the neurointervention had no side effects, then it would not be harmful. Given that harm is essential to communicate deserved censure, it would mean that administering a voluntarily accepted, side-effect free neurointervention as a response to his wrongdoing could not meet the Communicative Condition.

⁷⁴ It would be less harmful, and the offender's consent reduces the penal bite of the punishment.

⁷⁵ I say *proposed* punishment, because it would not be actual punishment as it would not involve hard treatment.