Is it morally permissible to offer neurointerventions that reduce aggressive behaviour to offenders to reduce recidivism?

Introduction

Recent advancements in neurotechnology have led to our ability to measure and modify neurological activity. Neurointerventions are treatments that can alter neural parameters underlying certain behaviours. There is significant discussion surrounding the moral permissibility of offering neurointerventions aimed at reducing aggressive behaviour to criminal offenders with the aim of reducing recidivism.[1] Although neurointerventions are similar to, but not identical to, medical interventions, it is useful to place neurointerventions in the domain of medical ethics as there are important shared normative concerns to consider such as invasiveness and consent.

The idea of offering medical interventions to offenders in the penal system is not a new one, but the idea of offering medical interventions to specifically treat behaviours underlying propensity to reoffend is one that must be met with specific ethical considerations. Some neurointerventions are already used in Europe and the US for 'chemical castration' through the administration of psychotropic drugs aimed at lowering testosterone levels in adult male sex offenders to reduce their risk of reoffending.[2] Similarly, newer neurointerventions can reduce recidivism risk in a population of criminal offenders.[3] These include Deep Brain Stimulation, psychoactive drugs and transcranial stimulation.[4] Ideally, with any medical method, the aim is to minimise the level of invasiveness as this often correlates with the severity of the infringement to the right to bodily integrity. Factors particularly relevant in differentiating between these methods include bodily contact, insertion, and diffuse versus localised alterations.

These potential neurointerventions highlight many normative concerns. One major concern is the fear that neurointerventions induce personality changes which potentially threaten a consistent narrative identity in these individuals.[5] However, in this paper, I argue that not only is the evidence for these claims weak, but that all significant life events change identity[6] and that the intention of these neurointerventions is to positively change identity as a part of the rehabilitation goal of imprisonment.

Another key normative concern regarding offering neurointerventions to criminal offenders is the validity of the consent that can be obtained. Critics argue that the vulnerability of offenders due to the power asymmetry in the penal system undermines the voluntariness and, hence, the validity of their consent.[7] On the other hand, denying criminals the opportunity to make their own choices is infringing on their moral right to mental self-determination, which I argue does exist for offenders.[8]

In this paper, I argue that, assuming they are safe and effective at reducing unprovoked aggressive behaviour, neurointerventions should be offered to criminal offenders. The prevention of crime is an obvious goal for society and the particular nuance of using these newer neurointerventions is an important one to discuss as they have the potential to achieve this goal more efficiently than current methods.

1. Bodily Integrity & Level of Invasiveness

Historically, offenders have been subjected to medical experimentation and treatment, often without informed consent. By the end of the 1960s, it is believed that around 90% of phase 1 trials of drugs in the US were carried out on inmates in American prisons.[9] This was especially the case for neurointerventions aimed at "curing" offenders' propensity for crime. These methods ranged from drug therapy to lobotomy.[10] Rightfully, the implementation of these neurointerventions was found to be an immense infringement on many human rights including the Right to Bodily Integrity. However, recent advancements in neuroscience have led to the development of neurointerventions that are much less invasive and have the potential to be efficacious in influencing behaviour in order to prevent crime.

Given the history of the use of neurointerventions in crime prevention, theorists may feel a knee-jerk reluctance to the modern revival of crime control on this level.[1] One obvious way in which this conversation has changed is in framing the use of neurointerventions in a consensual context. I will delve deeper into this later in this paper. Another way in which the discussion has changed is with the emphasis on reducing the invasion of bodily integrity. For the use of neuroinventions to be justified, it is necessary, but not sufficient, that they are minimally invasive. Other factors that would need to be considered are effectiveness and side effect profile which I will assume, for this discussion, are equal.

The most discussed forms of neurointervention are Deep Brain Stimulation (DBS), Transcranial Magnetic Stimulation (TMS), Transcranial Direct Current Stimulation (tDCS) and psychoactive drugs. It is important to compare many alternative options as the magnitude of invasiveness is thought heavily to depend on the context of what other interventions are available.[11] The way in which level of invasiveness is measured is equivalent to the accounts of infringement to the right to bodily integrity (RBI). Broadly, Ryberg considers 3 types of infringement of RBI: bodily contact, bodily insertion, and bodily alteration (non-physical).[10] There are also different factors to consider when determining the severity of the infringement to each of these accounts including magnitude, intensity, centrality of the affected body part and how functioning is affected.[4]

I would argue that transcranial neurostimulation is the least physically invasive of these listed options. TMS and tDCS both do not involve intensive physical contact: tDCS involves minimal bodily contact

with the electrodes placed on the scalp[12] and TMS is even less invasive as it uses electromagnetic coils placed near the scalp.[13] There are also studies which have shown the strong potential of the efficacy of transcranial methods in altering behaviour.[3] Although minimally invasive on the bodily contact and insertion counts, these methods do involve bodily alterations at the cellular level. The electrical and magnetic stimulation techniques do induce minor neurochemical alterations that are localised to specific brain regions. This is similar to the bodily alteration induced by DBS electrical signals, but transcranial techniques do not involve the very invasive process of brain surgery that DBS requires.[14] Hence, clearly, the brain surgery and implantation of a generator deem DBS as more invasive than TMS and tDCS.

It is less clear, instinctively, how the invasiveness of transcranial stimulation compares with the use of psychoactive drugs. When considering the level of invasiveness of this neurointervention, we must consider the route of administration and the neurochemical effects. These drugs are often administered either orally or via injection. Injection clearly involves bodily insertion, contact and alteration (and often pain). It is less clear whether oral administration involves significant bodily insertion or contact – nonetheless, it does involve at least a small magnitude of these factors. The largest difference between oral administration of psychoactive drugs (currently) and transcranial stimulation is the specificity of the brain regions that are affected. Whilst the neurochemical alterations are localised to brain regions in TMS and tDCS, psychoactive drugs act more diffusely throughout the brain leading to this being more invasive on this count.[4]

Although there are new technologies being developed in the fields of optogenetics and chemogenetics to reduce the non-specific actions of pharmacological therapies, for the purpose of the following discussion, I will be focusing on these newer neurointerventions (tDCS and TMS) as these are the least invasive and often the focus of the normative concerns that arise in this field.

2. Mental Integrity & Identity

As well as the aforementioned physical factors underlying level of invasiveness, interventions can also be invasive in non-physical ways.[15] This is particularly pertinent in the case of neurointerventions due to how neurological functioning underlies behaviour and sense of self. The aim of these neurointerventions in this context is to reduce unprovoked aggressive behaviour. The rationale is to mimic the effects of psychotherapy that is currently used[16] to raise an individual's threshold for switching to a more aggressive state. Ultimately, this should reduce how much external triggers result in unwanted, criminal behaviour.

However, part of the instinctive apprehension towards approving the use of neurointerventions, it might be said that they are therapies that artificially induce personality changes and threaten identity.[17] These

factors are usually grouped as PIAAAS (personality, identity, agency, authenticity, autonomy and self) and changes to PIAAAS as a result of neurointerventions (particularly the newer technologies – DBS and TMS) is an idea that is deeply rooted in neuroethics as an unchallenged view.[18] I would like to argue that the fear that neurointerventions induce changes to PIAAAS is a misplaced one for 3 main reasons:

(1) the supporting evidence is weak;
(2) neurointerventions, much like other significant life events, change one's identity;
(3) and, used in this context, neurointerventions contribute positively to the outcome of reducing recidivism.

There have been reports of personality changes following neurointerventions, especially DBS, that 'represent a threat to personal identity'.[17] One of the most influential quotes supporting this claim is from a DBS patient who declared "I feel like an electric doll"[19] and Witt et al. concluded from their patient cohort that "the risk of becoming another person following DBS surgery is alarming".[20] In order to understand identity from a philosophical perspective to attain whether these concerns are significant, 2 common forms are used: numerical and narrative. Numerical identity refers to the matter of existing as the same biological and psychological substance over time.[6] Narrative identity is a broader idea concerning what makes a person who they are – pertaining to their beliefs, values, desires, and other psychological features.[21] If neurointerventions threaten identity at all, it is thought that they threaten narrative identity more plausibly than numerical. Consistent narrative identity relies on individuals being able to explain the motivations for any changes. The concern here is that neurointerventions are not "natural personal development" and are an artificial change where the motivations for the change cannot be adequately explained.[22]

However, although the claims of PIAAAS changes due to neurointerventions are rife throughout neuroethics, others argue that the conclusions of studies[19,23] investigating the effects of DBS on PIAAAS (where these ideas seem to stem from) do not match these claims.[18] Supporting (1), Huoeto et al. argue that the 8 out of 24 patients who experienced changes related to PIAAAS following DBS implantation had changes that were more likely explained by the DBS reducing primary disease (in this case Parkinson's Disease) symptoms which unmasked previously unnoticed symptoms of psychiatric comorbidities.[24] This could also be related to the fact that these changes could be due to underlying disease progression (the so-called "burden of normality")[25] rather than the DBS alone.

As well as the negative changes to PIAAAS that are reported following neurointerventions, others also report positive changes in their sense of self due to being freer of their symptoms – evidence that the neurointerventions are having the desired therapeutic effects.[26] Hariz et al. reported that patients who underwent DBS for dystonia "still [were] the same person inside, but with new abilities and another physical appearance, [which] was difficult to comprehend and come to terms with".[27] Therefore, one

argument is that these effects alter an individual's identity, which is constantly changing (2), and that this is the primary intention of the treatment (3). This is especially pertinent to the use of neurointerventions with the intention of reducing aggression. Assuming that a non-invasive neurointervention achieves this, should we not actually be striving for a PIAAAS change, and that preserving previous identity is futile and counter-intuitive? In the case of psychotherapy (if successful) very few would object that psychotherapy is morally problematic because it ended up changing the personal identity. Thus, there are 2 potential claims here:

- psychotherapy is not problematic because it does not change narrative identity at all;
- or psychotherapy is not problematic because the very fact that it does change narrative identity is not morally problematic (or at least that we have intuitions that it is not).

The difficulty, it seems, with reconciling this idea with newer neurointerventions is the fact that psychotherapy seems a more "natural personal development" consistent with a narrative than neurointerventions. I'd argue that this difference should not be relevant and that, if anything, neurointerventions such as DBS and TMS are more efficient but have the same desired outcomes as more traditional methods such as psychotherapy. I understand that there are more concerns about the artificiality of these treatments being almost an "easy way out" in the context of offenders wanting to change their behaviour. This depends on the validity of consent that can be obtained from offenders which I will elaborate on in the next section.

Overall, regarding the moral permissibility of using newer neurointerventions, I would argue that it is an acceptable therapy and not a "threat" to narrative identity but a positive contributor to changing behaviour and, with it, identity. The next concern is the validity of the consent that can be obtained for such a therapy in the criminal justice system.

3. Consent & Vulnerability

Intuitively, if neurointerventions are completely safe, effective, and affordable, they are expected to have beneficial effects on increasing public safety and reducing taxpayer money spent on re-incarceration.[28] Assuming these to be true, these additional considerations can be used to strengthen my argument. However, if carried out without consent, the implementation of neurointerventions in this context is likely to infringe on many legal and moral rights (including RBI and right to mental integrity). Therefore, neurointervention therapy can only be morally permissible with the valid consent of the offender. However, offers of treatments to offenders in the criminal justice system raise normative concerns.[8] Some argue that the vulnerability of offenders might put into question whether they can

voluntarily choose when, for example, implicitly faced with the alternative of serving a longer prison sentence.

Valid informed consent involves disclosure of appropriate information to a competent person who is permitted to make a voluntary choice.[29] One major concern is that offering neurointerventions in order to reduce aggressive behaviour might be 'an offer you cannot refuse'. When making offers, vulnerability and power asymmetry have the potential to undermine the voluntariness of the choice and, thus, undermining the validity of the informed consent. Vulnerability is a broad term with various facets, but situational and pathogenic vulnerabilities are particularly pertinent here. Situational vulnerability is context-specific, and, in the case of offenders, the context of detention and general deprivation of liberty renders people more vulnerable than others.[8] Also, pathogenic vulnerability refers the cognitive psychological problems which are disproportionately present in imprisoned offenders compared to the rest of the population. This is found due to the fact that offenders, before incarceration, are more likely to have psychiatric disorders, and periods of imprisonment can lead to or worsen already present pathogenic vulnerability.[30] With these definitions of situational and pathogenic vulnerability, critics argue that the consent to neurointerventions is invalidated due to the inherent power asymmetry and dependency that detention leads to.[31]

On the contrary, whilst one's situational vulnerability is typically static until imprisonment ends, an offender's pathogenic vulnerability has the potential to improve during imprisonment.[8] Individuals with full cognitive capacity to consent can still be vulnerable because they are subjected to authority (situational vulnerability). The European Court of Human Rights (ECHR) argues that people who are deprived of their liberty are in a vulnerable position and that the authorities have a duty to protect them. This links to one of the primary goals of imprisonment being rehabilitation. It gives offenders the chance to reintegrate back into society as productive citizens and should reduce recidivism rates. I would argue that offering neurointerventions would be a great step towards providing rehabilitation opportunity to offenders. It more addresses the root causes of criminal behaviour – it is not perfect and does not remove external causes but does, theoretically, improve one's response to such triggers. Therefore, categorically denying offenders the benefit of safe and effective neurointerventions would likely increase the pathogenic vulnerability of an already vulnerable group by restricting the options of successful rehabilitation and achieving a crime-free life.[32] Prisoners are vulnerable specifically due to the nature of the penal system and the deprivation of liberty which only increases the risk of developing or exacerbating psychiatric disorders. Therefore, vulnerability should not be an argument to deny convicted offenders the opportunity of using novel medical interventions to voluntarily modify their minds for rehabilitation, where affordable and effective.[33]

This links to the wider discussion of whether convicted offenders have a moral right to neurorehabilitation. It can be argued that offenders do have a non-absolute moral right to the offer of safe and affordable neurointerventions when these would be part of the most effective package for facilitating their rehabilitation.[8] This right is not absolute in that, according to Article 3 of the ECHR, it 'cannot be construed as imposing on the authorities [as] an absolute duty to provide prisoners with rehabilitation (…) programmes and activities'.[34] Regardless of whether this moral right to neurorehabilitation exists, to which there is more discussion than I've outlined here, it is at least implicitly recognised that offenders ought not to be made or kept vulnerable or be denied the effective opportunity for rehabilitation.[8]

Conclusion

In this paper, I have argued that it is morally permissible to offer neurointerventions (such as TMS and tDCS) to offenders to reduce risk of reoffending. I have shown that, due to the fact that these neurointerventions are similar to medical interventions, minimising level of invasiveness is a necessary condition to fulfil when considering which techniques should be used, alongside maximising efficacy and minimising side effects. Specific to neurointerventions, I have addressed the normative concerns that there is a "threat" to narrative identity in using medical interventions to change behaviour. I have claimed that this is the very intention of the therapy and is not a "threat" but a goal which is not morally problematic. Finally, although vulnerability might put into question the voluntariness of offenders' consent to neurointerventions in the penal system, I have argued that offenders' vulnerability is the reason for which we should offer neurointerventions as this would be an effective way of reducing their pathogenic vulnerability and enables rehabilitation from imprisonment.

This is an important discussion because it is not unrealistic to believe that such non-invasive neurointerventions will be developed in the next few years. Furthermore, these arguments can hopefully be used to inform other discussions about prisoners engaging in clinical trials, for example.

1. Ryberg, J. Neurointerventions and crime prevention. An ethically inappropriate discussion? *Theory and Criticism of Social Regulation* **1**, 193–207 (2021).
2. Forsberg, L. Anti-libidinal Interventions and Human Rights. *Human Rights Law Review* **21**, 384–408 (2021).
3. Sergiou, C. S. *et al.* Transcranial Direct Current Stimulation Targeting the Ventromedial Prefrontal Cortex Reduces Reactive Aggression and Modulates Electrophysiological Responses in a Forensic Population. *Biol Psychiatry Cogn Neurosci Neuroimaging* **7**, 95–107 (2022).
4. Tesink, V., Douglas, T., Forsberg, L., Ligthart, S. & Meynen, G. Neurointerventions in Criminal Justice: On the Scope of the Moral Right to Bodily Integrity. *Neuroethics* **16**, 26 (2023).
5. Hansson, S. O. Implant ethics. *J Med Ethics* **31**, 519–525 (2005).
6. Baylis, F. Neuroethics and Identity. in *Handbook of Neuroethics* 367–372 (Springer Netherlands, Dordrecht, 2015). doi:10.1007/978-94-007-4707-4_9.
7. McGregor, J. Undue Influence as Coercive Offers in Clinical Trials. in *Coercion and the State* 45–59 (Springer Netherlands, Dordrecht, 2008). doi:10.1007/978-1-4020-6879-9_4.
8. Ligthart, S., Dore-Horgan, E. & Meynen, G. The various faces of vulnerability: offering neurointerventions to criminal offenders. *J Law Biosci* **10**, (2023).
9. Wiegand, T. J. Captive subjects: Pharmaceutical testing and prisoners. *Journal of Medical Toxicology* **3**, 37–39 (2007).
10. Ryberg, J. *Neurointerventions, Crime, and Punishment*. (Oxford University Press, 2019). doi:10.1093/oso/9780190846428.001.0001.
11. De Marco, G., Simons, J., Forsberg, L. & Douglas, T. What makes a medical intervention invasive? *J Med Ethics* **50**, 226–233 (2024).
12. Ciechanski, P. & Kirton, A. Transcranial Direct-Current Stimulation (tDCS). in *Pediatric Brain Stimulation* 85–115 (Elsevier, 2016). doi:10.1016/B978-0-12-802001-2.00005-9.
13. Chail, A., Saini, R., Bhat, P., Srivastava, K. & Chauhan, V. Transcranial magnetic stimulation: A review of its evolution and current applications. *Ind Psychiatry J* **27**, 172 (2018).
14. Larson, P. S. Deep Brain Stimulation for Movement Disorders. *Neurotherapeutics* **11**, 465–474 (2014).
15. Bluhm, R., Cortright, M., Achtyes, E. D. & Cabrera, L. Y. "They Are Invasive in Different Ways.": Stakeholders' Perceptions of the Invasiveness of Psychiatric Electroceutical Interventions. *AJOB Neurosci* **14**, 1–12 (2023).
16. Beaudry, G., Yu, R., Perry, A. E. & Fazel, S. Effectiveness of psychological interventions in prison to reduce recidivism: a systematic review and meta-analysis of randomised controlled trials. *Lancet Psychiatry* **8**, 759–773 (2021).
17. Schechtman, M. Philosophical Reflections on Narrative and Deep Brain Stimulation. *J Clin Ethics* **21**, 133–139 (2010).
18. Gilbert, F., Viaña, J. N. M. & Ineichen, C. Deflating the "DBS causes personality changes" bubble. *Neuroethics* **14**, 1–17 (2021).
19. Schüpbach, M. *et al.* Neurosurgery in Parkinson disease. *Neurology* **66**, 1811–1816 (2006).
20. Witt, K., Kuhn, J., Timmermann, L., Zurowski, M. & Woopen, C. Deep Brain Stimulation and the Search for Identity. *Neuroethics* **6**, 499–511 (2013).
21. Schechtman, M. The constitution of selves. in (1996).
22. Pugh, J. Clarifying the Normative Significance of 'Personality Changes' Following Deep Brain Stimulation. *Sci Eng Ethics* **26**, 1655–1680 (2020).
23. Agid, Y. *et al.* Neurosurgery in Parkinson's disease: the doctor is happy, the patient less so? in *Parkinson's Disease and Related Disorders* 409–414 (Springer Vienna, Vienna, 2006). doi:10.1007/978-3-211-45295-0_61.
24. Houeto, J. L. Behavioural disorders, Parkinson's disease and subthalamic stimulation. *J Neurol Neurosurg Psychiatry* **72**, 701–707 (2002).
25. Gilbert, F. The burden of normality: from 'chronically ill' to 'symptom free'. New ethical challenges for deep brain stimulation postoperative treatment. *J Med Ethics* **38**, 408–412 (2012).
26. Bluhm, R., Cabrera, L. & McKenzie, R. What we (Should) Talk about when we Talk about Deep Brain Stimulation and Personal Identity. *Neuroethics* **13**, 289–301 (2020).

27. Hariz, G., Limousin, P., Tisch, S., Jahanshahi, M. & Fjellman-Wiklund, A. Patients' perceptions of life shift after deep brain stimulation for primary dystonia—A qualitative study. *Movement Disorders* **26**, 2101–2106 (2011).
28. Suman, O. *et al.* The Effectiveness of Rehabilitations in Prisons and the Criminal Justice System. *The International Journal of Indian Psychology* **11**, (2023).
29. Eyal, N. Informed Consent. *The Stanford Encyclopedia of Philosophy* (2019).
30. Fazel, S., Hayes, A. J., Bartellas, K., Clerici, M. & Trestman, R. Mental health of prisoners: prevalence, adverse outcomes, and interventions. *Lancet Psychiatry* **3**, 871–881 (2016).
31. McGregor, J. Undue Influence as Coercive Offers in Clinical Trials. in *Coercion and the State* 45–59 (Springer Netherlands, Dordrecht, 2008). doi:10.1007/978-1-4020-6879-9_4.
32. Choy, O., Focquaert, F. & Raine, A. Benign Biological Interventions to Reduce Offending. *Neuroethics* **13**, 29–41 (2020).
33. Blitz, M. J. Extended Reality, Mental Liberty, and State Power in Forensic Settings. *AJOB Neurosci* **13**, 173–176 (2022).
34. 'Harakchiev and Tolumov v. Bulgaria' (2014) European Court of Human Rights, case 176