A Note on Psychological Continuity Theories of Identity and Neurointerventions

Holmen, Sebastian Jon

Published in:
Journal of Medical Ethics

DOI:
10.1136/medethics-2021-107492

Publication date:
2022

Document Version
Peer reviewed version

Citation for published version (APA):
Holmen, S. J. (2022). A Note on Psychological Continuity Theories of Identity and Neurointerventions. *Journal of Medical Ethics*, *48*(10), 742-745. https://doi.org/10.1136/medethics-2021-107492

A NOTE ON PSYCHOLOGICAL CONTINUITY THEORIES OF IDENTITY AND NEUROINTERVENTIONS

Abstract

An important concern sometimes voiced in the neuroethical literature is that swift and radical changes to the parts of a person's mental life essential for sustaining his/her numerical identity can result in the person ceasing to exist – in other words, that these changes may disrupt psychological continuity. Taking neurointerventions used for rehabilitative purposes as a point of departure, this short paper argues that the same radical alterations of criminal offenders' psychological features which under certain conditions would result in a disruption of numerical identity (and, thus, the killing of the offender) can be achieved *without* these having any effect on numerical identity. Thus, someone interested in making radical alterations to offenders' psychology can avoid the charge that this would kill the offenders, while still achieving a radical transformation of them. The paper suggests that this possibility makes the question of what kinds of qualitive alterations to offenders' identity are morally permissible (more?) pressing, but then briefly highlights some challenges for arguments against making radical qualitative identity alterations to offenders.

1. Introduction

As our knowledge of the workings of human brain expands, so does our ability to enhance or alter our own or others' affective, cognitive or motivational states by neurotechnological means. In the context of criminal justice, the possibility of behaviour modification by means of neurointerventions[i] has led to debate among ethicists and law scholars concerning the ethics of

[i] I here understand neurointerventions as interventions that achieve their effect(s) by working directly on the recipient's brain.

employing them as a means to reduce reconviction rates among some groups of offenders.[1–6] An important moral issue regarding the use of such neurointerventions in general revolves around their effect on the identity of the person who uses it or on whom it is used.[7,8] For example, much ink has been spilled in the philosophical debate concerning the potential impact of Deep Brain Stimulation on a subject's identity.[9] And a similar concern in the context of criminal justice specifically has been highlighted by Nicole Vincent, who points out that ethical problems with some forms of neurointervention used on criminal offenders might arise because they could "sever the link between their former and their latter self".[10] On a general level, the neuroethical identity debates concerning identity-affecting neurointerventions can be divided into roughly two spheres. The first starts from different approaches to answering what Marya Schectmann has termed *the characterization question*, a question that concerns "which beliefs, values, desires and other psychological features make someone the person she is".[11] With this approach, the relevant question is often framed in terms of whether the effect of a neurointervention is a threat to the narrative(s) that individuals construct about themselves and why this is morally problematic, when it is.[ii] The other part of this debate, the part that this paper will focus on, concerns the potential impact that neurointerventions might have on an individual's numerical identity.

While most of neurointerventions' observed effects on identity are plausibly best understood as concerning narrative identity,[iii] scholars have suggested that in some rare cases certain technologies can cause swift and radical changes to parts of a person's mental life essential for sustaining his numerical identity.[12] In addition, Parker Crutchfield has recently argued that pervasive enough changes to a person's moral traits can result in this person being destroyed. As he

[ii] For an excellent overview of different approaches to answering the characterization question in relation to Deep Brain Stimulation's impact on identity, see [9].

[iii] This plausibly includes most (if not all) of the effects of neurointerventions which are currently used or have been suggested could be used for crime-prevention such as, for instance, psychopharmaceuticals with libido or aggression hampering effects. Thus, the suggestion is not that such treatments in isolation would affect numerical identity. I shall say a little more about this in section 2.

puts it, this shows that under certain conditions "moral enhancement can kill".[13] As I understand them, theorists raising such concerns explicitly or implicitly adopt a psychological continuity view of numerical identity. At least, only such theories would seem to suggest that using a neurointervention on someone could cause him/her to become a new numerically distinct entity. Other theories, most notably biological accounts of numerical identity, would, for instance, not reach this conclusion as they equate the end of numerical existence with the death of the human organism.[7] At any rate, I shall assume that concerns regarding a neurointervention's potential impact on a person's numerical identity concerns its impact on psychological continuity.

In this short paper, I will argue, taking neurointerventions used for rehabilitative purposes as a point of departure, that concerns regarding the threat that radical psychological changes induced by neurointerventions may pose to numerical identity can be completely avoided by administering such neurointerventions gradually. More precisely, I will argue that the same radical alterations of an offender's psychological features which under certain conditions would result in a disruption of numerical identity (and, thus, the killing of the offender), can be achieved *without* it having any effect on numerical identity. Thus, someone interested in making radical alterations of an offender's psychology can avoid the charge that this would kill the offender, while still achieving a radical transformation of that offender. And this, I will suggest, makes the question of what kinds of qualitive alterations to offenders' identity are morally permissible (more?) pressing, although I will briefly point to some challenges for arguments against making radical qualitative identity alterations of criminal offenders.

The paper proceeds as follows. In the next section, Section 2, I will set out more clearly the concept of psychological continuity, and explain how it plausibly could be disrupted by neurointerventions. I will also point to some empirical data that seems to indicates that we may have such interventions available in the future. Section 3 describes a scheme for gradually altering

an offender's qualitive identity and shows how this scheme, while attaining exactly the same alterations of an offender's psychological features, *is not* disruptive of numerical identity according to psychological continuity theories. In Section 4, I draw some conclusions from this observation and point to areas where more work is needed.

2. Disruptions of psychological continuity

To ground the idea that the alteration of an offender's identity could be achieved by neurotechnological alterations of his/her psychological features, more first needs to be said about the nature of the psychological continuity of a person. Roughly put, a psychological continuity view stipulates that for a person to be numerically identical to him-/herself at different times, there must be a relevant psychological relation or connection between the person at those times.[iv] A complication is that there are several suggestions regarding what elements ensure psychological continuity, and we cannot review them all here in detail. However, in the influential version of the account developed by Derek Parfit, one explicitly adhered to by at least some theorists expressing concerns about the potential impact of neurointerventions on numerical identity,[12] it is suggested that psychological continuity "is the holding of overlapping chains of *strong* connectedness".[14] Psychological connectedness, in turn, consists of particular direct psychological connections, such as memories, forming intentions and acting on them, beliefs and desires, as well as other persistent psychological features.[14][v] According to Parfit, there are enough direct psychological connections for strong connectedness to obtain if "the number of direct connections, over any day, is *at least*

[iv] More precisely, versions of the psychological continuity view of numerical identity often hold that, besides the presence of psychological continuity, the continuity must also take a non-branching form and have the right kind of cause.[14,30] However, since the alteration of numerical identity presumably involves severing the psychological links between the pre-intervention and post-intervention offender, a discussion of these two latter conditions does not seem relevant for the present discussion.

[v] More precisely, in order to not presuppose identity, Parfit formulates these psychological relations relied on by the account in quasi-terms.[14] Taking memories as an example, a person has a quasi-memory of an event if (1) the person seems to remember the event, (2) someone did experience the event, and (3) the (apparent) memory is causally dependent in the right way on the experience.

half the number that hold, over every day, in the lives if nearly every actual person".[14] On this account, then, causing someone (or oneself) to become a numerically distinct entity involves replacing or altering at least half of the direct connections (e.g., memories, desires, beliefs) that the person would otherwise share with his/her former self. More importantly, because psychological continuity is constituted by the presence of (enough) psychological connections, the account implies that (enough) alterations of a person's direct connections can lead him/her to become psychologically continuous with someone else (i.e., be a numerically distinct person). Or, as Parfit puts it:

> […] psychological changes matter. Indeed, on one view [i.e., the psychological continuity view] certain kinds of qualitative change destroy numerical identity. If certain things happen to me, the truth might not be that I become a very different person. The truth might be that I cease to exist – that the resulting person is someone else.[14]

It should now be clear what theorists concerned with the impact that neurointerventions may have on psychological continuity are more precisely concerned about: that neurointerventions could server the link between who a person is at T^1 and T^2 by reducing the direct connections that sustain numerical identity between these persons to a sufficient degree.

A relevant question is, of course, whether we are currently able or will likely in the future be able to alter the psychological features that ensure continuity. If such radical changes of the psychological features that ensure the continuity of a person are not likely to be possible, a concern regarding a neurointervention's impact on numerical identity is nothing but a hypothetical problem thought out by clever philosophers. However, while we cannot scrutinize here the

empirical evidence concerning every possible persistent psychological feature involved in ensuring psychological continuity, studies indicate that at least three such features which are arguably central for the psychological continuity of a person can be altered by neurotechnological means: his/her desires, beliefs and memories. To be clear, the suggestion is not that isolated alterations of these or other psychological features will suffice to destroy numerical identity – in order to do so, such changes would, as we have just seen, need to be pervasive and would most likely need to include radical changes to many different kinds of direct connections involved in ensuring psychological continuity. The point here is simply to illustrate that at least some important sources of continuity seem to already be malleable by neurotechnological means and that other sources may soon be.

Starting with preferences and desires, some neurological interventions already available seems capable of altering at least some of them. For example, studies suggest that a person's preferences for reciprocity and fairness can be manipulated by pharmaceutical means,[15] and anti-libidinal agents seems capable of reducing sexual desires in their recipients.[16] Furthermore, the strong desires for alcohol[17] and certain drugs[18] involved in addictive behaviour have been found to be reduceable by pharmaceutical means. Thus, while we may not presently be capable of altering any and all preferences or desires by direct means, these examples of desire-moderation demonstrate that it is not absurd to suggest that it may become scientifically possible to do so.

The second psychological feature arguably central to the psychological continuity of a person is his/her beliefs. It has been speculated that neurointerventions may indeed in the future give us the power to directly change what a person believes[19] or, at least, to alter the process of belief-formation.[20] These speculations are supported by at least one empirical study. In it, Colin Holbrook and his colleagues found that by downregulating activity in the posterior medial frontal cortex using Transcranial Magnetic Stimulation they were able to alter their subjects' political and

religious beliefs.[21] While these results need to be corroborated by further studies, they do indicate that alterations of beliefs by neurotechnological means may be within our scientific capabilities in the medium term.

We turn now to the third and perhaps most important psychological feature of psychological continuity: memories. While these remain speculative and mostly based on animal models, there are some indications that neurotechnologically induced manipulation of a person's memories may be possible in the future.[22] For example, some studies indicate that it may be possible to modify or even erase some memories by infusing a subject with a protein synthesis inhibitor, e.g. propranolol, that prevents memory reconsolidation.[23,24] In addition, neuromodulation techniques such as Deep Brain Stimulation[25,26] and optogenetics[27] have also been suggested as potential tools to modulate or erase some memories. And researchers have in at least one study also been successful in implanting memories in sleeping mice by neurotechnological means.[28]

All in all, then, while it currently remains mostly at the level of theory, these studies suggest that it may become possible to employ neurointerventions to make alterations to central psychological features that ensures a person's psychological continuity. This indicates, in my view, that the concern raised by some theorists regarding neurointerventions' potential impact on numerical identity should be taken seriously. Now, suppose that these or similar technologies are indeed developed. Suppose further, that the state found that in some cases making radical alterations of an offenders' memories, desires, beliefs or other psychological features central for preserving psychological continuity would be the only way to prevent the offender from committing serious crimes in the future.[vi] An obvious concern regarding such a proposal that a proponent of a

[vi] I will here leave it an open questions exactly what psychological features the state would need to alter in order to destroy numerical identity, and what group(s) of offenders would be suitable candidates for such radical interventions. I will do so, not because these questions are not important, but because the modest point I will attempt to make in the

psychological continuity theory of identity could raise, is that employing such a scheme would amount to killing the offender. There is, however, a way that this concern can be avoided by the state while still achieving the same radical transformation of the offender, or so I will now argue.

3. Sweeping and gradual identity-altering schemes

On the basis of what was said in the previous section, we can construct a simple illustration of how the alteration of an offender's numerical identity could work by introducing certain qualitative changes to the offender's psychological features. Suppose that at T^1 an offender has psychological connections q, y, x and z connecting him/her to his/her former self (T^0), and that these connections each account for a quarter of the total number of connections. The offender is now administered a *sweeping intervention* (*sweeping*) so that, at T^2, these connections have all been replaced with connections q*, y*, x* and z*, resulting in him/her no longer being psychologically continuous with the person at T^1. The offender's numerical identity would have been disrupted since strong connectedness no longer obtains between T^1 and T^2. However, as already indicated, such a radical alteration of an offender's psychological features need not affect his/her numerical identity.

To see why, consider a scheme of *gradual interventions* (*gradual*). Suppose that at T^1 an offender has psychological connections q, y, x and z connecting him/her to his former self (T^0). The offender is now administered a scheme of gradual interventions so that, at T^2, connection q has been replaced with connection q*. Another intervention is then, at T^3, administered to him/her that replaces connection y with connection y*. At T^4 another intervention is administered that replaces connection x with connection x*, and at T^5 the process is repeated and connection z is replaced by connection z*. Now, as is clear, the offender at T^5 in *gradual* and the offender at T^2 in *sweeping* are

coming section seems to stand regardless of what a scheme of radical psychological alterations of offenders more precisely involves in terms of specific psychological alterations and target group(s).

left with the same degree of changes to their psychological makeup at the end of the scheme – in both cases, psychological features q, y, x and z have been replaced by q*, y*, x* and z*. However, while *sweeping* results in the destruction of the T^1 offender's numerical identity, this is not the case in *gradual*: since we are supposing that each of the psychological connections (i.e., q, y, x and z) each account for a quarter of the total number of direct connections, altering one such connection at each time-slice is not sufficient for breaking the chain of overlapping connectedness between time-slices. And, so, while the offender in *gradual* experiences exactly as radical an alteration of his/her psychological features as in *sweeping*, the former scheme has no effect on his/her numerical identity. In the next section, I will draw some conclusions from this observation.

4. Concluding remarks

In my view, what has been said so far points to two main conclusions. First, insofar as we are concerned that a rehabilitative scheme involving the administration of neurointerventions may have an effect on an offender's numerical identity, i.e., psychological continuity, we should ensure that the sum of psychological features changed in any given treatment session does not exceed the level needed to sustain strong connectedness between the offender in different time-slices.[vii] As we have seen above, this can, at least in principle, be done without it having any impact on the effectiveness of the rehabilitative scheme – the offender will still be a radically different person after *gradual*.[viii] However, and second, given that *gradual* ultimately involves psychological alterations as radical as

[vii] This is not to suggest that doing so would be an easy task. It is, for example, currently not even clear how many direct connections are precisely needed to sustain strong connectedness. But, while there is no logical necessity in this, it seems plausible to suppose that *if* we develop technologies that could indeed be used for making someone into a numerical distinct person, we also have the capacity to develop ways to control the pervasiveness of the impact of these technologies.

[viii] But it is worth noting, that even if numerical identity is preserved in *gradual* there may be moral costs to employing this scheme rather than *sweeping,* costs that might make it preferable to employ the latter scheme, all things considered. If, for example, an offender enrolled in a scheme like *gradual* uses the time between interventions to seriously harm his follow inmates or the prison staff, then it is at least debatable whether the value of preserving numerical identity outweighs preventing such harm.

sweeping, we should ask whether such radical qualitative changes to an offender's identity are themselves morally dubious – are offenders, for example, able to successfully incorporate them into their self-narratives? I will leave a thorough investigation of this question for future work, but let me end by briefly highlighting two challenges that an opponent of employing a scheme such as *gradual* in the criminal justice system would have to meet.

First, supposing that the involvement in a scheme such as *gradual* was offered to, or even requested by, the offender him/herself, it is difficult to see why such a scheme should be considered more ethically concerning than other steps that offenders may take to alter their present self which are usually considered ethically unproblematic or even laudable (e.g., attending an anger-management class to curb aggressive behaviour or taking anti-depressants to combat depression). Someone opposing the voluntary use of a scheme such as *gradual* for crime prevention would have to demonstrate there is a morally relevant difference between these changes of psychological features.

Second, to deny that a scheme such as gradual should be used in the criminal justice system due to its impact on qualitative identity would seem to imply that at least one traditional form of punishment (i.e., incarceration, which is often argued, at least sometimes, to be an appropriate response to wrongdoing), as well as state-mandated rehabilitative measures, may be morally problematic. This may be so in regards to the former, because studies have shown that this form of punishment can cause alterations to the offender's (qualitative) identity.[2,29] Furthermore, state-mandated therapy, such as anger-management class, might plausibly also result in changes to an offender's identity (indeed, such rehabilitative measures are likely mandated in a pursuit to achieve exactly such changes in the offender). And notice that a scheme such as *gradual* could, at least in principle, be constructed so that, considered in isolation, the impact of each alteration would be no more pervasive than the alterations of an offender's psychological connections that might

result from incarceration or rehabilitative measures. Given this, arguments against mandating a scheme such as *gradual* to offenders will also be faced with the difficult task of avoiding being over-inclusive.

References

1 Douglas T. Criminal Rehabilitation Through Medical Intervention: Moral Liability and the Right to Bodily Integrity. *J Ethics* 2014;**18**:101–22. doi:10.1007/s10892-014-9161-6

2 Ryberg J. *Neurointerventions, Crime, and Punishment: Ethical Considerations*. New York: Oxford University Press 2020.

3 Birks D., Douglas T. Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice. Oxford: Oxford University Press 2018.

4 Shaw E. Direct Brain Interventions and Responsibility Enhancement. *Crim Law Philos* 2014;**8**:1–20. doi:10.1007/s11572-012-9152-2

5 Bublitz J.C., Merkel R. Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. *Crim Law Philos* 2014;**8**:51–77. doi:10.1007/s11572-012-9172-y

6 Ryberg J. Punishment, Pharmacological Treatment, and Early Release. *Int J Appl Philos* 2012;**26**:231–44. doi:10.5840/ijap201226217

7 DeGrazia D. *Human Identity and Bioethics*. New York: Cambridge University Press 2005.

8 The President's Council on Bioethics. BEYOND THERAPY: BIOTECHNOLOGY AND THE PURSUIT OF HAPPINESS. Washington, D.C.: 2003.

9 Pugh J. Clarifying the Normative Significance of 'Personality Changes' Following Deep Brain Stimulation. *Sci Eng Ethics* Published online first: 2020. doi:10.1007/s11948-020-00207-3

10 Vincent N.A. Restoring Responsibility: Promoting Justice, Therapy and Reform Through Direct Brain Interventions. *Crim Law Philos* 2014;**8**:21–42. doi:10.1007/s11572-012-9156-y

11 Schechtman M. *The Constitution of Selves*. Ithaca: Cornell University Press 1996.

12 Klaming L., Haselager P. Did my brain implant make me do it? Questions raised by dbs regarding psychological continuity, responsibility for action and mental competence. *Neuroethics* 2013;**6**:527–39. doi:10.1007/s12152-010-9093-1

13 Crutchfield P. Moral enhancement can kill. *J Med Philos (United Kingdom)* 2018;**43**:568–84. doi:10.1093/jmp/jhy020

14 Parfit D. *Reasons And Persons*. Oxford: Oxford University Press 1984.

15 Siegel J.Z., Crockett M.J. How serotonin shapes moral judgment and behavior. *Ann N Y Acad Sci* 2013;**1299**:42–51. doi:10.1111/nyas.12229

16 Lösel F., Schmucker M. The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *J Exp Criminol* 2005;**1**:117–46. doi:10.1007/s11292-004-6466-7

17 Rösner S., Hackl-Herrwerth A., Leucht S., *et al.* Opioid antagonists for alcohol dependence. *Cochrane Database Syst Rev* Published Online First: 2010. doi:10.1002/14651858.CD001867.pub3.

18 Mattick R., Breen C., Kimber J., *et al.* Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. *Cochrane Database Syst Rev* Published Online First: 2009. doi:10.1002/14651858.CD002209.pub2.

19 Clayton M., Moles A. Neurointerventions, Morality and Children. In: Birks D., Douglas T., eds. *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice*. Oxford: Oxford University Press 2018. 235–51.

20 Persson I., Savulescu J. The Perils of Cognitive Enhancement and the Urgent Imperative to

Enhance the Moral Character of Humanity. *J Appl Philos* 2008;**25**.internal-pdf://143.129.158.127/The Perils of Cognitive Enhancement and the Ur.pdf

21 Holbrook C., Izuma K., Deblieck C., *et al.* Neuromodulation of group prejudice and religious belief. *Soc Cogn Affect Neurosci* 2016;**11**:387–94. doi:10.1093/scan/nsv107

22 Glannon W. *The Neuroethics of Memory: From Total Recall to Oblivion*. Cambridge: Cambridge University Press 2019.

23 Soeter M., Kindt M. An Abrupt Transformation of Phobic Behavior After a Post-Retrieval Amnesic Agent. *Biol Psychiatry* 2015;:880–6. doi:https://doi.org/10.1016/j.biopsych.2015.04.006

24 Parsons R.G., Ressler K.J. Implications of memory modulation for post-traumatic stress and fear disorders. *Nat Neurosci* 2013;**16**:146–53. doi:10.1038/nn.3296

25 Merkow M.B., Burke J.F., Ramayya A.G., *et al.* Stimulation of the human medial temporal lobe between learning and recall selectively enhances forgetting. *Brain Stimul* 2017;**10**:645–50. doi:10.1016/j.brs.2016.12.011

26 Fell J., Staresina B.P., Lam A.T.A. Do, *et al.* Memory Modulation by Weak Synchronous Deep Brain Stimulation: A Pilot Study. *Brain Stimul* 2013;**6**:270–3. doi:https://doi.org/10.1016/j.brs.2012.08.001

27 Lacagnina A.F., Brockway E.T., Crovetti C.R., *et al.* Distinct hippocampal engrams control extinction and relapse of fear memory. *Nat Neurosci* 2019;**22**:753–61. doi:10.1038/s41593-019-0361-z

28 De Lavilléon G., Lacroix M.M., Rondi-Reig L., *et al.* Explicit memory creation during sleep demonstrates a causal role of place cells in navigation. *Nat Neurosci* 2015;**18**:493–5. doi:10.1038/nn.3970

29 Haney C. The Psychological Impact of Incarceration: Implications for Post-Prison

Adjustment. *US Dep Heal Hum Serv* Published online first:

2002.http://img2.timg.co.il/CommunaFiles/19852476.pdf

30 Shoemaker S. Personal identity: A Materialist Account. In: Shoemaker S., Swinburn R., eds. *Personal Identity*. Oxford: Basil Blackwell 1984. 67–132.