Direct Brain Interventions, Criminal Rehabilitation and The Value of Moral Reasoning

skrevet af Sebastian Holmen, RUC · udgivet februar 2018

Abstract
In this paper, I critically discuss whether G. Owen Schaefer's argument from the inherent value of authentic moral reasoning against the use of state-mandated value-modulating DBIs, (i.e. direct brain interventions that aim to instil certain conceptions of good and bad in an individual without valid consent), succeed in establishing the moral impermissibility of these DBIs as part of criminal offender rehabilitation. For several reasons, both principal and empirical, I argue it does not.

1. Introduction and state of the art
Novel insights into the neurological workings of the human brain suggest that it may be possible in the future to influence the behaviour of individuals by direct means¹ when, for instance, “that behaviour is medically unremarkable but socially undesirable” (Farah 2002: 1125). The direct means that have received the majority of attention in the literature include the use of Selective Serotonin Reuptake Inhibitors (SSRIs) to reduce aggressive behaviour (Farah 2002; Douglas 2008), nasal spray to

¹ Whether an intervention is categorized as working directly or indirectly on a person's behaviour is often said to depend on whether the intervention circumvents the subject's internal processes; see, e.g., (Bublitz and Merkel 2012) and (Raus et al. 2014). For the remainder of this paper, I will assume that there is a morally relevant difference between these two types of interventions, but it should be noted that this view is not uncontroversial; see, e.g., Greely (2008: 1133f).
elevate the subject’s level of the hormone and neurotransmitter oxytocin, which shows signs of being able to promote in-group cooperation and altruistic behaviour (Persson and Savulescu 2012), and the beta-blocker Propanolol to reduce implicit racial bias (Terbeck 2016; Terbeck et al. 2012). Other non-pharmaceutical techniques that show promise in aiding anti-social behavioural modulation are deep brain stimulation (Lu, Wang, and Kosten 2009), transcranial magnetic stimulation (Young et al. 2010) and neuro-feedback (Sitaram, Caria, and Birbaumer 2009). In addition to raising interesting questions regarding whether these means could or should help to solve some of the pressing problems facing humanity (Persson and Savulescu 2012, 2008; Douglas 2008), these Direct Brain Interventions (DBIs) are of special interest to anyone with a narrow concern regarding criminal offender rehabilitation and reform—that is, reducing the likelihood of an offender recidivating. This interest is not surprising given that rates of recidivism have been reported to be as high as 50 percent in the UK (Ministry of Justice 2012) and as high as 70 percent in some parts of the US (Durose, Cooper, and Snyder 2014). If rehabilitation and reform are among the legitimate aims of state punishment, then new ways to accomplish this seem to be in high demand.

The debate regarding the moral permissibility or desirability of these interventions revolves around several distinct and important neuroethical questions. Some argue that even if DBIs are deemed safe and effective, they should only be administered after obtaining valid consent (Bomann-Larsen 2013; Craig 2016; Bublitz and Merkel 2012), whereas others argue

---

2 But see, e.g., Wiseman (2016).

3 It might, of course, be possible to reduce rates of recidivism in ways other than rehabilitation. For instance, we might instead incapacitate offenders by incarcerating them indefinitely. This holds true for some DBIs as well, such as ice pick lobotomies, which would likely leave the offender unable to commit further crimes. Such proposals would, however, likely (and rightly so) be met with more serious ethical objections than a proposal focusing on rehabilitation.

4 Throughout this paper, I will assume that rehabilitation and reform represent at least one of the legitimate aims of state punishment, but see, e.g., Hirsch, Ashworth, and Roberts (2009) for a review of some possible objections.
that it might sometimes be permissible to mandate DBIs as part of a criminal sentence (Douglas 2014; Ryberg and Petersen 2013; Petersen and Kragh 2017; Pugh and Douglas 2016). In addition, some have focused on the target of DBIs that might be permissibly offered or mandated. Some have argued that DBIs should only be used to better or remediate an agent's motivation to do what he himself believes to be good (Douglas 2008; Schaefer 2015), whereas others argue that it might be permissible to directly modulate or otherwise directly affect an agent's insight into goodness or badness (DeGrazia 2014). Although both types of DBIs might lead to behavioural improvement and might aid in offender rehabilitation, I will take a special interest in the latter types of DBIs and refer to them as value-modulating DBIs. I limit my scope to these interventions because, as will be clear in what follows, the argument directed against value-modulating DBIs whose evaluation is the aim of this paper seems to be directed at these types of interventions. This makes sense after all, since it is unclear how an improvement of the offender's motivation to do what he himself believes to be good would have a negative effect on the authenticity of an offenders reasoning process. It should be noted, that I do not intend to argue that safe and effective techniques to directly modulate an individual's values yet exist. However, this might be because research into these possibilities is still in its infancy. We would do well to consider sooner rather than later how we ought to respond to the moral challenges that might arise if such techniques are discovered. As Greely (2008) notes, there is a real danger that if we do not, we might be caught off guard and realize too late “that neuroscience and the criminal justice system have combined to produce the next great prefrontal lobotomy [...]” (Greely 2008).

Objections to the use of value-modulating DBIs are diverse and have been substantiated with reference to the adverse effects of these types of DBIs might, if not received on a voluntary basis, have on the offender's
freedom (Harris 2011; Simkulet 2016) and autonomy (Bublitz and Merkel 2009). Concerns have also been raised in regards to the ascription of moral and legal responsibility to the post-intervened individual (Vincent 2012), as well as the adverse effect their use could have on the relationship between the offender and the state (Shaw 2012; for a critique see Holmen 2017). However, this paper takes as its points of interest another objection to employing these value-modulating DBIs. The objection that will be critically discussed is the view that there is something intrinsically valuable in individuals using their own ability to reason about moral issues. This view presents a challenge to a proposal for state-mandated value-modulating DBIs as part of a rehabilitative scheme as the offenders might then lose something of great moral value if we were to instil conceptions of good and bad directly in them (i.e., circumventing their own reasoning process). This paper proceeds as follows. In section 2, I present an argument regarding the inherent value of authentic contemplation and deliberation on moral issues recently presented by G. Owen Schaefer (2015). I argue that Schaefer does not do enough to show that the use of value-modulating DBIs differs in any morally relevant respect from the unproblematic (to most) inculcation of values already occurring in other social institutions. Second, I argue that the scope of Schaefer's argument might be considerably limited given that the social-intuitionist model of moral judgement is empirically sound. Third, I argue that insofar as authenticity involves not only producing one's own ideas about how one should live but also expressing these ideas by, for example, acting on them, a value-modulating DBI that would keep the offender from future incarceration (which arguably limits an agent’s ability to act on his own ideas) might leave him more authentic post-intervention. Lastly, I argue that even if this is not true (i.e., even if the offender is left less authentic post-intervention), this might not suffice to establish the moral impermissibility of these DBIs since other values, such as the well-being of
individuals, should also be considered. Section 3 summarizes the findings and concludes.

2. The inherent value in moral reasoning

In a recent article, Schaefer (2015) proposes several arguments in support of the view that the state should not mandate the use of value-modulating DBIs. Specifically, in Schaefer view, the state should never employ interventions (or direct moral enhancements, as he dubs these interventions) that “aim at bringing about particular ideas, motives or behaviours” (ibid.: 261). Most of Schaefer's arguments in support of this view pertain to the adverse effects of the widespread societal use of these DBIs, e.g. that such use would thwart the dissent necessary for moral progress since all individuals would be morally homogenous. They therefore seem to pose no immediate challenge to a proposal for a rehabilitative scheme in which only a minority of a population (i.e., criminal offenders) would be affected. Schaefer does, however, present a concern that does not necessitate the full-blown societal use of DBIs that modulate values to be morally problematic and that therefore must be addressed by a proponent of state-mandated offender rehabilitation by means of DBI. This concern stems from Schaefer's conviction of the inherent value in moral reasoning i.e., the practical reasoning by collectives or individuals or, as it will be framed here, the act of contemplating and deliberating “about what, morally, one ought to do” (Richardson 2014). The idea that moral reasoning should be held in the highest esteem has, as Schaefer notes, a long history in moral philosophy dating (at least) back to Socrates, who in Apology called the daily reflection on virtue “the greatest good to man” (Plato 2006: 133). Although Schaefer explicitly refuses to adopt this somewhat extreme view, he maintains that the core idea, that contemplation and deliberation on or
about moral issues should be protected and promoted for its own sake, is sound. In his words,

It is quite plausible to think that there is value in the process itself of deliberating over a moral proposition, both within one's own mind and in discussion with others [...] [T]here is something intrinsically good about such reasoning, something worth promoting and protecting. (Schaefer 2015: 268)

From this assumed intrinsic value of moral reasoning, Schaefer concludes that state-mandated interventions\(^5\) are “problematic to the extent that one's ideas can no longer be said to be the product of one's own reasoning” (ibid.: 269). In the following sentence, he contends that this is exactly what employing value-modulating DBIs would do since such interventions make a person's “own reasoning process more or less obsolete; even if one goes through the motions of reasoning, the enhancers have 'rigged the game,' so to speak, to ensure the desired outcome” (ibid.: 269). Although Schaefer does not use the term himself, his concern seems to involve the possible disregard of the *authenticity* of the reasoning process and, by extension, the moral ideas that are a product of this process. The term authenticity can be used in different ways. However, as it will initially be understood here (more will be said about the notion below), something (or someone) is authentic if and only if it is what it claims to be in origin\(^6\). Reframing Schaefer's argument in terms of this minimal conception of authenticity is useful, as it sheds light on why, in Schaefer's view, it is morally

---

\(^5\)Schaefer in fact argues that it is problematic even if an individual herself asks for such an intervention. That is, an individual would lose something immensely valuable if he were to choose to undergo a value-modulating DBI that would either circumvent his own reasoning or direct his own reasoning towards a specific conclusion. I shall not engage here in a comprehensive discussion of this view as I tend to think that a successful refutation of Schaefer's argument against *mandatory* DBIs will suffice to show that requesting such interventions should not be objected to with reference to the need to protect the authenticity of the agents moral reasoning.

\(^6\)See also Varga and Guignon (2014) for a similar definition.
problematic if one's moral reasoning process and the ideas of how one should live one's life does not remain distinctly one's own. This is so, because the moral reasoning process would then not be what it claims to be in origin, namely, the individual's own. An example will perhaps help elucidate this further. Suppose we were to inculcate a particular moral idea directly in an offender (e.g., that it is wrong to commit violent crimes) or "rig" the offender's reasoning process to cause the offender to reach this conclusion. This would be unacceptable in Schaefer's view as this idea cannot be said to be a product of the offender's authentic reasoning process but rather is a product of an external source's reasoning process or the product of a reasoning process that has been directed towards the idea the external source finds best. Hence, the offender would have lost something of intrinsic moral value, namely, his own contemplation on or deliberation about this moral issue and its product. Considering the prominent place that reasoning has held (and, indeed, still holds) in moral philosophy for several millennia as well as the fact that authenticity is, as Neil Levy (2007: 73) notes, “one of the definitive values of modernity”, Schaefer's argument has some intuitive appeal. However, as will be argued below, there is reason to doubt that Schaefer's argument succeeds in ruling out the use of value-modulating DBIs as part of a rehabilitative scheme in toto.

The first challenge one might raise to Schaefer's argument is empirical in nature, as it is arguably unclear to what extent our moral ideas or judgements can be said to be a product of reasoning. For instance, in his widely cited article on the social intuitionist model of moral judgement, the moral psychologist Jonathan Haidt (2001) argues that there is some support for the hypothesis that “moral reasoning does not cause moral judgment; rather, moral reasoning is usually a post hoc construction,
generated after a judgment has been reached” (p. 814). What Haidt is suggesting is that most of our moral judgements are the result of quick automatic evaluations, i.e. intuitions, which we then (sometimes) subsequently underpin with reference to abstract principles. As he has so eloquently put it elsewhere “Intuitions comes first, strategic reasoning second” (Haidt 2013: 1). More work is required to establish the validity of this claim, but if this provocative hypothesis is correct, then it seems that Schaefer’s concern might be unwarranted in most cases as this would show that most of our moral ideas are not the product of moral reasoning. Therefore, one would likely not reduce the amount reasoning produced by an agent if one were to directly modulate his values. However, even in the social intuitionist model of moral judgement in which the role of reasoning is predominantly to provide post hoc justifications for intuitions, it is recognized that agents, on rare occasions, do reach moral judgements by reason alone. Therefore, even if the social intuitionist model is empirically sound, Schaefer could argue that his concern (i.e., that value-modulating DBIs would disregard the inherent value in engaging in an authentic reasoning process) still applies, even though the number of cases to which it pertains is considerably smaller than he anticipated.

Second, one might question whether the mandatory inculcation of certain values by means of DBIs is distinguishable from the inculcation of values that already occurs in some social institutions which I suspect that most people find morally unproblematic or even desirable. Most educational institutions, for instance, have the inculcation of certain values as one of their main objectives. This is perhaps most obvious in public schools where children, sometimes without understanding the reasons, are instructed to be e.g. civil and kind (DeGrazia 2014). More work is required if one wishes to substantiate the assertion that these two instances of value inculcation are different in any morally relevant respect. Alternatively,

7 See also Haidt (2013).
Schaefer could of course “bite the bullet” and assert that inculcation of values in public school (and other social institutions) is morally wrong for the same reasons that he argues that inculcation of values in individuals by means of DBIs is wrong i.e. it circumvents or rig intrinsically valuable authentic reasoning processes. To be sure, this would be a consistent argument. However, it would seem to be faced with (at least some of) the challenges presented below, because it ultimately relies on the same premise. i.e. that one’s reasoning process and the ideas of how one should live one's life ought to remain distinctly one’s own, being plausible. Another problem with this argument, is that it seems to go against what I take to be a common conception of the important role that our social institutions should play in the moral education of our children. That is, are we truly prepared to argue that even if the values or virtues that are inculcated by teachers and others are acceptable from all reasonable views of morality (which I suspect e.g. civility, kindness and not unnecessarily harming sentient beings, would be), it would still be wrong to instil them? I for one find this to be a highly implausible view.

As a third challenge, one might note that, rather surprisingly, a possible implication of protecting and promoting authentic reasoning processes and ideas would seem to be that some offenders could ultimately come to be less authentic, all things considered. This might be so because one might argue that the value of authentic moral reasoning comes not just from arriving at one's own ideas about how one should live, but also from acting out these ideas. This at least seems to be the view pressed by advocates for the moral significance of authenticity. For instance, Bernard Williams, has once described authenticity as “the idea that some things are in some sense really you, or express what you are, and others aren’t” (Williams as quoted in Guignon 2004: viii). Insofar as the term

---

8 I thank two anonymous reviewers for pressing me on this point.
9 The idea that the values or virtues that are inculcated in our public schools should be acceptable from all reasonable views of morality has been suggested by DeGrazia (2014).
“expressing” can be said to include acts, then it seems that part of what matters in being authentic, is to act in accordance with one's own values. So how do I propose that protecting or promoting authentic moral reasoning might lead to individuals being less authentic? Suppose that John received a long-term prison sentence because he after employing is moral reasoning had concluded, that he should attack Bob. The thing to notice is that during his incarcerated for his wrongdoing, John is unable to act on some of the ideas that express who he is, such as, for instance, attending political rallies, playing sports or be present in the lives of his children. In this case, it seems plausible to argue that the John's own moral reasoning process (which let to him being incarcerated) might, in fact, have rendered him less authentic, all things considered. If so, then it is possible that administering a value-modulating DBI that would instil directly in him (or “rig” his reasoning process to have him conclude) that he should not commit violent crimes might leave him more authentic than he would otherwise be if he recidivated after being released. Schaefer could perhaps attempt to sidestep this challenge by stressing that he is concerned solely with the value of authentic moral reasoning and not with the broader notion of the moral value of persons being authentic. That is, he could argue that even if a person's own moral reasoning does not contribute to the person being authentic (or, as suggested above, could be detrimental to it), it still has value in its own right. However, in my view, this would drain Schaefer's argument for much of its intuitive appeal as well as explanatory power, which both arguably stems from the role that authentic moral reasoning plays in individual's ability to live authentic lives.

Moreover, even if one agrees with Schaefer's assertion that having one's own reasoning about matters of morality has value in itself, I believe very few (if any) would hold no other value(s) as also having this status – in fact, I doubt that even Schaefer himself will argue as much. Such a
monist theory of the good would, at least in a consequentialist framework, seem to imply that all our actions should be evaluated based on how much they contribute to the total amount of authentic contemplation and deliberation. Imagine a society in which everyone adopts this view of morality; little would occur but contemplation and the deliberation of the matter of morality. However, what is perhaps more puzzling is that one would seem to be blameworthy for spending time on other activities that most consider good. For example, it would be morally wrong to devote one's time to charitable work since this would likely take time away from one's contemplation and deliberation. Therefore, contemplation and deliberation are surely not all that matters and I suspect that most would agree that, other things being equal, individuals' well-being has intrinsic value as well (regardless of what one might take to constitute well-being). If so, we would have to know how to trade-off between these two values in a situation in which both values could not be adequately pursued. This seems to be the exact situation we face when attempting to decide whether an offender should receive a mandatory DBI. This is so because an intervention whose aim is to reduce the chance of the offender recidivating (i.e., to rehabilitate her) would, if successful, arguably leave the offender as well as possible future victims of her crime better off in terms of well-being. However, to achieve this increase in well-being, we might have to circumvent the offender’s reasoning process, which would mean a decrease of this other moral good. But if we refrain from administering these DBIs, thereby maintaining the offender’s current level of reasoning, we would likely be causing the offender and his future victims to experience a substantial decrease in terms of well-being. One way to solve this dilemma would be to subtract the decrease in one value from the increase in the other. If the result were, all things considered, positive, the action would be morally justified. I will not attempt to provide such an all-things-

---

10 It could, of course, be argued that these values are incommensurable and that it follows...
considered account, but it seems unlikely that such an account would always be in favour of protecting or promoting the authenticity of offenders' contemplation and deliberation.

In sum, the argument from the inherent value of authentic moral reasoning as an objection against mandatory value-modulating DBIs seems to face possible empirical and, certainly, principled problems insofar as it is thought to exclude the use of these DBIs as part of an attempt to rehabilitate certain criminal offenders. However, as a final remark it should be noted that it cannot be inferred from what has been said in this paper, that we should employ value-modulating DBIs as part of a rehabilitative scheme. This is so, because one could argue that one of the concerns mentioned in the introduction, e.g. concerns for the offender's autonomy or ascription of responsibility to the post-intervened individual, would suffice to render their use morally impermissible.

3. Conclusion
I have argued that the argument from the inherent value of authentic contemplation and deliberation recently defended by G. Owen Schaefer does not succeed in ruling out all value-modulating DBIs as part of offender rehabilitation. I argued this point by first noting that more work is needed to show that the state-mandated use of value-modulating DBIs is different from the (to most people) unproblematic inculcation of values in public schools in any relevant respect. Second, I noted that if the social intuitionist model about the nature of moral judgement is empirically sound, then Schaefer's concern might only be relevant in a considerably fewer number of cases than he seems to believe. Third, I argued that insofar as authenticity also involves acting on one's own ideas about how

that no trade-off between them is possible. However, notice that even if this is true, it does not unequivocally support the conclusion that we should not administer value-modulating DBIs as part of rehabilitative schemes as both administering and not administering them would then be supported by good reasons.
one should live, a value-modulating DBI that would ensure that the offender does not recidivate might leave the offender in a better position to act on these ideas. Lastly, I argued that even if this is not the case, it might not suffice to show that these DBIs should not be used as factors such as individuals’ level of well-being should also be considered.
References


Terbeck, S., 2016. *The Social Neuroscience of Intergroup Relations - Prejudice, can we cure it?* Switzerland: Springer International Publishing.


