Those who forget the past: An ethical challenge from the history of treating deviance[1]

Emily McTernan

Treating those who commit crimes or behave in ways deemed socially undesirable with medical interventions targeting the brain, or 'neurointerventions', comes with a history. That history is one full of appalling cases and this chapter's purpose is to argue that this should affect our assessment of whether it is ethical to try again. Among many other examples, there is the chemical castration of men convicted of consensual same-sex sexual acts, such as Alan Turing who committed suicide two years later. There is the use of electric shocks to treat those deemed 'criminally insane', aggressive, or in general 'non-compliant'. In its early incarnation, shock treatment came with severe side effects, including broken bones, spinal injuries, and memory loss.[2] Further, shock treatment was sometimes deliberately misused, as in the case of the 'Georgia power cocktail', which was used to punish the uncooperative in Milledgeville State Hospital, Georgia.[3]

As a final example of perhaps the most infamous treatment in the history of neurointerventions, there is the use of lobotomies in an attempt to achieve what was deemed socially desirable behaviour. On one estimate, some 60,000 lobotomies were carried out in the US and

---

1 With thanks to the participants at the workshop on criminal neurointerventions, University of Oxford, for their helpful feedback. I am also very grateful to David Birks, Tom Douglas, Chris Nathan, Chris Mills, and two anonymous referees for written comments.

[2] Many other surgeries and treatments with severe side effects were also deployed. For examples see, for instance, Elliot S. Valenstein, *Great and desperate cures: The* rise *and decline of psychosurgery and other radical treatments for mental illness* (New York: Basic Books, 1986); Andrew Scull, *Madhouse: A tragic tale of megalomania and modern medicine* (Yale University Press, 2007).

[3] For the example, see Edward Shorter and David Healy, *Shock Therapy: A history of electroconvulsive treatment in mental illness* (New Brunswick: Rutgers University Press, 2013), esp. pp. 93-94. Electroconvulsive therapy remains in use and continues to be controversial.

Europe between 1936 and 1956.[4] Famous cases include the lobotomies carried out on Howard Dully, who was lobotomised at age 12 for his 'defiant' personality, and on Rosemary Kennedy, who was left incapacitated and institutionalised.[5] One route to grasp the wrongs done is offered by Herman Snow's review in 1949 of twenty-seven, mostly female, patients who were lobotomised.[6] Snow describes the behaviours that motivated the lobotomies and what he deemed the positive outcomes of treatment. In one case, Snow describes a woman who was before treatment, 'uncooperative, distractible, irrelevant, untidy.[7] Those patients for whom the treatment was deemed to work were variously made 'neat', 'tidy' and compliant, becoming 'well dressed', with 'better eating manners'.[8] For the pursuit of such socially desirable behaviours, invasive and radical surgery, which often compromised intellectual abilities such as abstract reasoning or the capacity to make future plans, was deemed appropriate.

Cases like the above may account for a certain hesitation over embracing the modern incarnation of such interventions.[9] A variety of suggestions of potential contemporary neurointerventions have been made, including treating those who are aggressive by lowering their

---

[4] Statistic cited in Miguel A. Faria Jr, 'Violence, mental illness, and the brain–A brief history of psychosurgery: Part 1–From trephination to lobotomy' *Surgical Neurology International* 4 (2013), p.49.

[5] Howard Dully & Charles Fleming, *My lobotomy: A Memoir*, reprint (New York : Broadway Books, 2008). Popular culture offers further infamous cases, such as in Ken Kesley's *One Flew Over the Cuckoo's Nest* (New York: Viking Press, 1962).

[6] Herman B. Snow, 'A review of 27 prefrontal lobotomy patients' *Psychiatric Quarterly* 23.1 (1949), pp. 26-34.

[7] *Ibid.*, p. 27. While aggression and violence were used to motivate the treatment, what is especially problematic is that the treatment is also justified by appeal to the presence of these other features such as being untidy.

[8] *Ibid.*, pp. 31-32.

[9] Some go further, like the antipsychiatry movement. For a brief survey of this movement, see Marcelo T. Berlim, Marcelo PA Fleck, and Edward Shorter, 'Notes on antipsychiatry' *European archives of psychiatry and clinical neuroscience* 253.2 (2003), pp. 61-67

testosterone or providing antidepressants; chemically castrating sex offenders; drugging offenders to lessen impulsiveness or to encourage them to conform to society's rules; or engaging in forms of brain stimulation.[10] This chapter examines what force, if any, the appalling nature of the history of using neurointerventions to respond to socially undesirable behaviour has when we assess these contemporary proposals. Does some course of action having a bad history provide reason to object to it and, if so, under what conditions?

My purpose is not to offer a conclusive case against neurointerventions, but rather to defend the relevance of the history of this approach to tackling socially undesirable behaviour, when addressing its ethical prospects. Recent work on treating crime through neurointerventions at times fails to mention that history, or mentions it only in passing.[11] Further, proponents of one popular variety of applied ethics are likely to hold that such historical cases are an irrelevance, preferring abstract imaginary examples taken not to distract us from core ethical issues such as whether these treatments require consent.

In contrast, this chapter argues that to ignore the history is to miss an important dimension of the ethical debate and, hence, that we have reason to engage with real, historical cases as well as imaginary ones. History is relevant here not out of a concern about compensat-

---

[10] On treating aggression, see for instance Jesper Ryberg, 'Punishment, pharmacological treatment, and early release' *International Journal of Applied Philosophy* 26.2 (2012), pp. 231-244. On aggression and impulsiveness, among other traits, see Thomas Douglas, 'Criminal rehabilitation through medical intervention: Moral liability and the right to bodily integrity' *Journal of Ethics* 18.2 (2014), pp. 101-122. On castration, see, for instance, Meyer III, Walter J., and Collier M. Cole. "Physical and chemical castration of sex offenders: A review." *Journal of Offender Rehabilitation* 25.3-4 (1997), pp. 1-18; Don Grubin and Anthony Beech, 'Chemical castration for sex offenders' *BMJ* 340 (2010). On impulsiveness and law abidingness see Elizabeth Shaw, 'Direct brain interventions and responsibility enhancement', *Criminal Law and Philosophy* 8.1 (2014) pp. 1-20, esp. pp. 14-15.

[11] For examples, see Douglas 'Criminal rehabilitation through medical intervention: Moral liability and the right to bodily integrity'; Shaw, 'Direct brain interventions and responsibility enhancement'; for a mention in passing, Ryberg 'Punishment, pharmacological treatment, and early release', at p. 242.

ing victims of past injustice but, instead, as a source of epistemic clues in our assessment of the ethical prospects of neurointerventions. In what follows, I will argue that the history provides a defeasible reason or presumption against treating those who commit crimes or act in socially undesirable ways through neurointerventions.

Section 1 formulates a pessimistic induction from the bad history of neurointerventions in treating socially undesirable behaviour to their current incarnations. Section 2 then outlines the role such a pessimistic induction should play in our ethical thinking. The remainder of the chapter addresses the prospects for the defender of neurointerventions as a way to treat socially undesirable behaviour, in light of the pessimistic induction. In section 3, I argue that two apparent defences of modern neurointerventions are inadequate. In concluding, I consider what we can learn from the broader context in which the pessimistic induction places neurointerventions, as amongst other medical interventions that aim at the social good.

1. **A pessimistic induction**

First then, how and when does a bad history matter, in the sense of providing some grounds to object to a revival of a particular policy or course of action? The simplest answer would be the following inductive argument. Having been mistaken in the past, we have reason to think that we would continue to get things wrong by following a relevantly similar approach now. So, lobotomies were at one point regarded as a cutting edge technology and much praised, but now many regard them as a serious scientific mistake.[12] Thus, we should be wary of altering the brains of criminals, for fear of making the same kind of serious mistake.

Yet all of medicine, and science more generally, is open to the challenge that it has been

---

[12] For one discussion of the change in attitudes towards lobotomies, in the context of deep brain stimulation as a new proposed therapy, see Jennell Johnson, 'A dark history: memories of lobotomy in the new era of psychosurgery' *Medicine Studies* 1.4 (2009), pp. 367-378, pp. 367-8. As evidence of the approval of lobotomies as a form of treatment, consider that a Nobel Prize was awarded to Egas Moniz in 1949 for his work in the area.

mistaken before, so why not think that it is so now? One classic challenge to scientific realism is a formulation of something like this thought: the pessimistic meta-induction. To give a rough formulation, our past scientific theories turned out to be false despite seeming successful at the time and so, by induction, we have no reason to think that our current scientific theories are true despite their successes.[13] Such caution based on previous mistakes in science is not entirely irrelevant; for instance, it is worth being cautious before assuming that we currently have the right account of how the brain works. Putting to one side this general challenge to scientific realism, however, is there any objection targeting the use of neurointerventions to treat socially undesirable behaviour in particular, which does not generalise to all and any other courses of action based on scientific knowledge?

Consider more precisely how we have been wrong about neurointerventions when they were playing this role in the past: they were ethically appalling.[14] What goes wrong is not only a scientific mistake but also a moral one. Evidently, part of what is wrong with lobotomies and the like is their scientific shortcomings. We think that lobotomies were an insufficiently advanced, overly blunt way to intervene in the brain, or that electric shock therapies were delivered with insufficient care. But, then, the lobotomised woman would not engage in the same kinds of behaviour afterwards as she did before. In one sense, the treatment was effective by the lights of what the doctors intended. Another part of what went wrong in the historical cases though, and the part that this chapter focuses on, is that people were wronged by these treatments.

---

[13] For one formulation of this see L. Laudan, 'A Confutation of Convergent Realism', *Philosophy of Science*, 48 (1981), pp. 19–48.

[14] It might be objected that formulating an ethical pessimistic induction adopts the wrong attitude to history. One might insist that we cannot judge history this way, that there would be something anachronistic in imposing our moral code to judge those in the past and their behaviour. To respond, I remain neutral on how and if we can hold those in the past accountable. I address our ethical decision making now, and ask how the history might be relevant. From here at least these past actions are morally appalling.

First, those who suffered treatment were harmed and/or put at risk of harm; for example, losing central intellectual capacities in the case of lobotomies or suffering broken bones in the early versions of electric shock treatments. These consequences can be viewed as unintended side effects of an undeveloped science, but they are also a way in which the treatments were ethically problematic: people were harmed or put at risk of harm. Whenever we try out new treatments, or even use known treatments, we risk doing harm through side effects, foreseen or unforeseen, and that should affect our moral assessments of that course of action.[15] Second, some of the treatments had objectionable goals. In some cases, those treated were not the appropriate subjects of such interventions, including defiant children or men engaging in consensual same-sex relations. So too, in some cases the interventions had inappropriate ends, such as aiming at compliance or patients becoming 'neat' and 'tidy'. Third, some of the interventions involved means that were in excess of their intended ends: the treatments were disproportionate in effects, or in their degree of invasiveness, for the outcomes at which the doctors aimed (i.e. the intended benefits of the treatment). A full account of the wrongness of past neurointerventions likely appeals to both the scientific shortcomings and such a cluster of ethical wrongdoings. In contrast, not all flawed science wrongs people.

So, in the case of neurointerventions the mistake is not only scientific but also moral. To restate my proposed pessimistic induction then, if a course of action has been historically ethically terrible we have reason to doubt that it is now the correct course of action to take.[16] For this chapter's purposes, I understand 'course of action' broadly and inclusively, as a description of the choice to Φ, as well as in fact carrying out Φ, and the manner in which it is carried out. Previously, it has been unethical to engage in neurointerventions to target socially undesirable behaviour. So we should be wary of introducing

---

[15] With thanks to an anonymous referee for suggesting this clarification. The discussion of side effects is continued later in the chapter.

[16] A stronger formulation would be to claim that we have no reason to think it the right course of action to take, given historical mistakes.

such neurointerventions into our criminal justice system now: we have a defeasible reason or presumption against engaging in these neurointerventions.[17] Indeed, a similar form of this pessimistic induction regarding potential scientific advances also appears in popular thinking about eugenics. After the twentieth century, some think that eugenics is no longer an acceptable area of scientific inquiry, given how wrong it went not only under the Nazis, but also in the US and Europe where, for instance, some faced compulsory sterilisation.[18]

Before considering the implications of this pessimistic induction, I first address three issues with its formulation, the first two regarding its extension and the last, its pessimism. First, then, one might ask whether the pessimistic induction can extend to any courses of action with a bad history, whether or not these are scientifically informed. In a similar vein, the reader may wonder why scientific errors have been included alongside the moral ones above, as in the case of side effects. What work is done by invoking the scientific means? The chapter focuses only on cases where certain scientifically informed means have a history of ethical wrongdoing and it is beyond its scope to fully explore the differ-

---

[17] Mine is a historically based principle, but there may be some relation to the non-historical precautionary principle, where, roughly, in the face of scientific uncertainty, we should take precautions where harm may be caused. For a discussion see Stephen M. Gardiner, 'A core precautionary principle' *Journal of Political Philosophy* 14, 1 (2006), pp. 33-60. Alternatively, one might consider discussions of how scientists should deal with the risk that their findings are wrong when they communicate with the public, e.g. Stephen John, 'Inductive risk and the contexts of communication', *Synthese,* forthcoming. But note, my focus is on policy making or what we do as a society, and not on how scientists should alter their communications or findings given the risks in getting things wrong.

[18] For details, see Elof Axel Carlson, *The unfit: A history of a bad idea* (New York: Cold Spring Harbor Laboratory Press, 2001); Michael G. Silver, 'Eugenics and compulsory sterilization laws: Providing redress for the victims of a shameful era in United States history' *George Washington Law Review* 72 (2004), pp. 862-892. Allan Buchanan describes the 'ethical autopsy' that has occurred after eugenics in his 'Institutions, beliefs and ethics: Eugenics as a case study', *Journal of Political Philosophy* 15, 1 (2007), pp. 22-45, where he proposes an alternative to the conventional account.

ences that come from extending the pessimistic induction beyond such cases. However, there is reason to think that the involvement of science makes some, perhaps small, difference to the strength of a pessimistic induction. In general, we face uncertainty when making ethical choices concerning the full consequences of our actions. Still, our degree of uncertainty seems to be perhaps increased by involving scientific means. Many times, scientific findings are later revised. In particular, medical and psychiatric interventions often turn out to have unanticipated side effects, some of which end up being severe. And neurointerventions are for the most part new(ish) treatments or, at the least, often not tested in the population at hand, such as trying to treat aggression rather than depression with anti-depressants. The use of scientific means thus alters the ethical assessment.

As the second issue with extension, it might be objected that the reformulated version of a pessimistic induction may overgeneralise, as did the purely scientific version with which this section began. In particular, the history of medicine is one of appalling things being done and of direct harms to persons, although, admittedly, often with good intentions. These include, among many others, performing leg amputations without general anaesthetic; experimenting on vulnerable populations, like children with disabilities; and a state of affairs where it was safer to give birth on the street on the way to the hospital than in the hospital.[19] Not all of these may be convincing as instances of wrongdoing. For instance, amputation without anaesthetic may have been the best among a set of bad options. All the same, one might think there is enough wrongdoing to run a general pessimistic induction for medicine. Yet, surely we do not want to say that practicising medicine is ethically suspect?

---

[19] On childbirth, for a summary and statistics, see, for instance, Caroline M. De Costa 'The contagiousness of childbed fever: A short history of puerperal sepsis and its treatment' *Medical Journal of Australia* 177.11/12 (2002), pp. 668-672. On amputation and surgery see, for instance, Roy Porter, *Blood and guts: A short history of medicine* (New York City: WW Norton & Company, 2004). On experiments on children with disabilities – defending it, and from one of the people involved in the study that infected children aged 3-10 with hepatitis - see Saul Krugman, 'The Willowbrook hepatitis studies revisited: ethical aspects' *Review of Infectious Diseases* 8.1 (1986), pp. 157-162.

The reply to this is two-fold. First, giving a more precise account of the wrongdoing of neurointerventions being used to treat socially undesirable behaviour grounds a stronger pessimistic induction than that which could be formulated for medicine in general. In many of the medical cases, the intention was to help the person treated and harm is done accidentally. In contrast, for these neurointerventions, there is not only harm done to individuals, but also the harm is motivated by a desire to fit individuals into society or to make society better in ways that were objectionable. Further, this particular variety of wrongdoing is shared by a set of central cases where a pessimistic ethical induction is attractive and compelling. It is true of the worst elements of the history of psychiatry, where we 'treat' to fit people to social norms or ideals of socially desirable behaviour, and so remove from society those whose behaviour does not fit: the rebellious or noncompliant woman, for instance.[20] It is true of the history of eugenics, where attempts were made to make a 'better' society or 'better' people, and resulted in, for instance, forcible sterilisation of those with disabilities.[21] And it is true of treating criminals and the socially disobedient with neurointerventions.

As such, for the set of cases just detailed, there can be a strong formulation of an ethical pessimistic induction. Aiming at the social good through medical interventions has been deeply unethical in the past. This gives us a reason for thinking that it will continue to be so, and so a (defeasible) reason why we should not use medical interventions to aim at social control or a 'better society'. Some might think that the history of neurointerventions to treat socially undesirable behaviour is not on a par with the excesses of psychiatry or eugenics in the twentieth century, in its degree of wrongdoing. However, if this form of the ethical pessimistic induction holds, insofar as neurointerventions for socially undesirable behaviour are relevantly

---

[20] On the history of women in psychiatric care see, for instance, Elaine Showalter, *The female malady: Women, madness, and English culture, 1830-1980* (New York CityL Pantheon, 1985). For a counterargument, see Joan Busfield 'The female malady? Men, women and madness in nineteenth century Britain' *Sociology* 28, 1 (1994), pp. 259-277.

[21] See footnote 17 for references.

similar to this set of wrongful actions, then that broader history of greater wrongdoing strengthens the case against neurointerventions. Further, there is nothing ad hoc about grouping neurointerventions for socially undesirable behaviour with eugenics and the worst elements of psychiatry. These are cases that share both method and aim: of treating people, via medical interventions, for the sake of the social good.

As the second aspect of the reply to the extension objection, there may not be, in fact, any *over*generalisation if we also formulate a pessimistic induction for medicine, albeit a weaker one. Consider the various restrictions on the behaviour and power of doctors, which we might consider a response to a bad history. For instance, many countries insist on informed consent from patients, and have a stringent approval system for new drugs and ethical approval systems for experiments. I return to these in more detail later in the chapter. But they appear to be answers to a pessimistic induction, providing ways to block its force by removing or mitigating the causes of previous wrongdoing. In short, perhaps medicine also faces a pessimistic induction, but the continued practice of medicine is defensible owing to these new protections for patients.

The third concern about the formulation of the pessimistic induction arises from the above discussion of medicine. It might be argued that at least in the case of medicine, and perhaps in the case of neurointerventions for socially undesirable behaviour, I could have formulated an optimistic induction rather than pessimistic one. Let us start with the stronger of the two cases. In general, one might argue, medicine is a success. A great many people have been helped rather than harmed. So, does it follow that having done good in the past with medicine, we have reason to think that medical interventions are a way to continue to do good, or some presumption in favour of medical intervention? Wouldn't the good effects of medicine outweigh the bad?

Further, perhaps one could construct a similar argument for neurointerventions as a response to socially undesirable behaviour. One could argue that psychiatry's brain interventions have, on average, done good and place neurointerventions for undesirable behaviour

amongst these – although I suspect that this claim of doing good on average would be tricky to defend, given the historical facts. Alternatively, one could seek to re-situate neurointerventions as one among many medical treatments. As such, perhaps we can defend pursuing change through this particular medical intervention strategy, given the broader successes of medicine.

To reply, first, an optimistic induction where the wrongdoing is simply cancelled out by the good done assumes a simplistic consequentialist view, where all we are concerned about is overall benefit, aggregating across individuals. Instead, one might hold that to respect the separateness of persons we should not aggregate in that way, or deny that rights violations can be justified on the grounds of overall benefit. In addition, such an optimistic induction looks likely to be unhelpful in guiding our current practices. It won't show us how to make things better. In contrast, a pessimistic induction that, say, casts doubt on the advisability of letting doctors alone determine what is best for patients can be a useful guide to how to improve practice. We have reasons for caution and reason to create safeguards where medicine has gone wrong before. In the next two sections, I examine the role that the pessimistic induction should play in our ethical reasoning, demonstrating how the bad history of medicine could be – and to some extent – is answered.

The second reply is that we have good reason to place neurointerventions in a category with interventions like eugenics and 'bad' psychiatry in order to form a pessimistic induction, and not 'good' psychiatry or medicine in general in order to form an optimistic induction. Namely, given their similarity in aims and methods, there is nothing ad hoc about grouping neurointerventions with eugenics and bad psychiatry: all these aim at making 'better' people, with ideals about how people ought to be or behave motivated by appeals to the good of society. Of course, how to divide up courses of action into groups will always be open to debate. There is unlikely to be a single right way. Nonetheless, a history of wrongdoing is not answered or undermined by grouping some course of action with some other benign courses of action that are different in the salient respects - the respects that make a difference, morally speaking.

2. **Against starting from nowhere**

In the context of ethics, however, one might doubt that the inductive move detailed above has force. Why not echo Ryberg, and say about neurointerventions that 'obviously former cases of use or misuse of such interventions do not imply that treatment cannot be properly applied'?[22] This section outlines the role of a pessimistic ethical induction in our moral reasoning. A bad history does not show a treatment cannot be properly applied but, rather, means that performing such treatments both requires a particular kind of justification and opens the agents involved to moral criticism.

So, suppose that I intend to do good by Φ-ing, but it goes wrong and in retrospect we judge that Φ-ing was appalling: that, say, it did someone harm and was morally reprehensible. I then try to do good in the same way, again Φ-ing. As a rough attempt to capture the intuitive ethical principle in play, I am plausibly open to moral criticism for Φ-ing again, unless I have good grounds for holding that the two cases where I Φ-ed were relevantly dissimilar. So, take a surgeon who thinks that they have come up with a new way to perform some procedure. The surgeon tries it out, with terrible results. Then, the surgeon tries it a second time. The first might be forgivable, but the second opens the surgeon to moral criticism, barring some adequate justification of repeating the mistake.[23]

To spell out the moral idea here further, first, one might observe that an agent who pays no attention to her past mistakes does not act well, morally speaking. Or, for those who resist relating the permissibility of acts to an agent's character, this agent is a poor moral agent. She fails to possess the kind of practical intelligence that good moral agents possess. In particular, we might accuse her of a kind of ethical hubris in her disregard for what has gone wrong in the past. That ethical hubris wrongs those who are affected by her acts: the agent fails to take due care and seems guilty of a kind of negligence. She imposes a known risk on

---

[22] Ryberg 'Punishment, pharmacological treatment, and early release', at p.242.

[23] With thanks to an anonymous reviewer for this example.

another, without sufficient justification. The agent who ignores the past also suffers from an epistemic shortcoming. Someone who fails to learn from her mistakes, or to take them as a reason to reconsider, is unlikely to succeed in achieving her aims. She seems to fail to use inductive reasoning, which is a fairly serious shortcoming in coming to know about the world.

Second, when I Φ again, then there seems to be an extra justification required for my actions, one which was not present the first time. To illustrate this, consider the person who has Φ inflicted on her for the second time and her grounds for complaint. She can reasonably ask, why didn't you know better, given the history of attempts? I have to answer why Φ was different this time, why I wasn't really Φ-ing, or why Φ-ing is now defensible, such as by showing that Φ-ing was better than the alternatives, including inaction. If I fail to answer that reasonable demand for justification, then I am open to moral criticism – and not only for failing, but for failing again.[24]

Returning to neurointerventions to treat socially undesirable behaviour and their history, then, one can demand an account of why the modern incarnation of such neurointerventions is unlike those cases where we lobotomised non-conforming women, chemically castrated men accused of same-sex relations, and gave electric shock treatment to the 'non-compliant'. Without that, one is vulnerable to charge of ethical hubris and failing to justify one's action. This restates the ethical pessimistic induction: we have a case for thinking that neurointerventions as a means to tackle socially undesirable behaviour would be an ethical mistake again, since they were in the past, and now we have to answer that case. To make the case that neurointerventions now are different to neurointerventions then, the details of the history would become relevant to determine whether the two sets of cases are similar enough for the principle above to apply: that both are 'Φ-ing'.

To illustrate the role of an ethical pessimistic induction and how to answer it, consider

---

[24] Various moral approaches might spell this out differently. This is admittedly sketchy but partly deliberately. Even a consequentialist should worry about hubris in the ethical sphere, with regards to the likelihood an agent will succeed in doing good.

again the case of medicine. I suggested that medicine, too, might be the subject of an ethical pessimistic induction, albeit a weaker one. Consider, then, a doctor who treats a patient with some medical intervention. There must be, in principle, some way to justify doing so against a history of medical mistakes. Often, however, such a justification is available, for two reasons. First, there are often grounds for arguing that a medical intervention is likely to work, or that on balance, its good effects outweigh the bad, given careful experiments and trials of the treatment proposed. That will meet the demands of justification stemming from one form of the ethical pessimism for medicine regarding the harm to individuals of previous treatments, and so make it unlike previous medical interventions that lacked such evidence. Note, this is not going to be a fruitful line for the proponent of neurointerventions for socially undesirable behaviour to take, since inadvertent harm is not the place from which that pessimistic induction draws its force.

Of course, there are other aspects to the wrongdoing in medicine, and there are other responses to give to these. So, in light of the wrongfulness of the history aside from causing accidental harm, we have sought to limit how doctors can behave. We try to reign in their ethical hubris by ensuring patient rights are respected, requiring informed consent rather than doctors doing what they think best, restricting who can be experimented on, having checks on doctors' performance, and so on. Again, then, the case can be made that doctors do not engage in the same kind of intervention as found in the most ethically problematic elements of medical history. These changes to the medical profession can be construed as an acknowledgement of,

and answer to, that bad history.25[26]

At this point, the proponents of neurointerventions for treating socially undesirable behaviour may hope that they too have some easy way to defend themselves against an ethical pessimistic induction akin to the defences that are available for medicine. However, the rest of this chapter will be devoted to suggesting that the case is not an easy one to make, although I do not aim to show that all things considered we should not engage in neurointerventions to treat crime. The history of neurointerventions is only one dimension to the necessary ethical debate, providing a defeasible reason or presumption against opting for this means again.

3. The modern incarnation of neurointerventions

I consider two easy ways of defending neurointerventions for treating socially undesirable behaviour from a historically based pessimism that spring to mind: one, that different people are targeted, the other that science is better now. First, then, one might argue that was wrong in the cases of bad neurointerventions outlined in the introduction is that they targeted for the most part those who should not have been: rebellious women, gay men, and so on. Yet, that response does not capture all the instances of treatment that ground the claim that the history of neurointerventions is one of appalling cases. For example, consider the treatment of the 'criminally insane'. It is not the case that the only wrong done was targeting the wrong people. Rather the

---

25 For example, in the UK, the General Medical Council, on informed consent and patient rights, available at http://www.gmc-uk.org/guidance/ethical_guidance/consent_guidance_index.asp; and NHS guidelines, available at http://www.nhs.uk/conditions/consent-to-treatment/pages/introduction.aspx. On medical experiments see for example the World Health Organisation's guidelines, in 'Standards and operational guidance for ethics review of health-related research with human participants', WHO Press, 2011. Available at http://www.who.int/ethics/research/en/.

[26] As a clear instance of this, consider the Nuremberg Code, regarding experiments on human subjects, which was a direct response to the appalling experiments done in Nazi concentration camps. For a discussion, see Leaning, Jennifer, 'War crimes and medical science', *BMJ (*1996), p. 313.

wrong also lies in what was done to those people for the sake of ensuring socially desirable behaviour. Lobotomies, even on those prone to aggression who sometimes had episodes of violent behaviour, are a step too far, for instance, given the possible effects of the treatment in severely diminishing a person's capacities.

Further, simply targeting different people fails to address directly the entirety of the ethical pessimistic induction. That induction raises doubt concerning deploying a particular method for achieving a certain kind of end: using medical or scientific interventions to address socially undesirable behaviour. We still have reason to doubt that we will succeed in picking a set of 'social undesirables' or in treating that group appropriately, in a proportionate way, morally speaking. Even the worst criminal can still demand that we justify imposing such an ethically risky form of treatment on him – given the terrible history – and we are still open to moral criticism for doing so without reasons that defeat the pessimistic induction.

As a second seemingly easy defence one might answer the pessimistic induction by insisting that we are now better at science, such that we know that modern neurointerventions for socially undesirable behaviour are likely to work or, at least, that their side effects will be less problematic. Certainly, suggestions of surgery on the brain are less favoured, replaced with suggestions of deploying drugs to alter brain chemistry instead. Perhaps modern science would be more light touch than past attempts. One might suggest that, as a result, to perform contemporary neurointerventions for socially undesirable behaviour is to pursue a different course of action to their historical incarnation.

However, it is unclear that modern science has improved sufficiently to block a pessimistic induction. The brain is not yet well understood, let alone understood well enough for scientists to successfully intervene in only light touch ways. Further, those proposed treatments within the philosophical literature that are even vaguely realistic come with side effects, which are often severe. So, antidepressants, for instance, can cause side effects including loss of sexual desire, fatigue, insomnia, weight gain, nausea, or anxiety.[27] Thus, we do not

[27] See, for illustration, http://www.nhs.uk/Conditions/Antidepressant-drugs/Pages/Side-effects.aspx.

seem to have minimally invasive and well-understood treatments to hand.

The above may express too 'luddite' an attitude for some who have greater faith in our scientific knowledge or in the possibility of progress. Alternatively, the applied ethicist who favours very hypothetical cases may respond by imagining some perfect case, where we really do understand the ways in which the drug will work, and the side effects really are limited. The evidence is as good in this hypothetical case, this applied ethicist continues, as our best researched science. Does that answer the pessimistic induction?

First, to reply, it is unclear that such hypothetical cases are the best starting point. Starting from such cases it may obscure some of what matters ethically. As one example, our medical treatments are very rarely without side effects, as inspecting the information leaflet in any drug packet will demonstrate. So, perhaps it is worth keeping in view the likely costs of any treatment. As another example, many likely neurointerventions for socially undesirable behaviour would be experiments, at least in turning what has been a treatment for a clinical population into one for a subclinical population or a population with a different issue, such as the suggestions of using antidepressants on the aggressive or providing Ritalin to the impulsive. Further, if we are in the business of treating 'criminals' these experiments would be being carried out on what can be considered a vulnerable population, of prisoners or potential prisoners. This is a population that is in a violent, coercive environment (or, in the case of those at the point of conviction, threatened with such), whose members are disproportionately likely to have passed through the care system and/or be otherwise from a disadvantaged background.[28] Note, here I am not arguing that prisoners should not be experimented on, all things consid-

---

[28] For a review of research about race, poverty and incarceration rates in the US, along with the effects of imprisonment on perpetuating inequality, see Sara Wakefield and Christopher Uggen, 'Incarceration and stratification' *Annual review of sociology* 36 (2010), pp. 387-406. On the care system and imprisonment in the UK, see for example E. Kennedy, 'Children and Young People in Custody 2012–13' (London: HM Inspectorate of Prisons and Youth Justice Board, 2013), who reports that a third of young offenders had been in the care system.

ered, but rather, that this is an ethical debate that needs to be had, and so a detail of the real world case to which we should attend. As such, it can be unhelpful to abstract away from all features of the likely actual situation in favour of one's imagined cases, so obscuring the central ethical issues at hand.

As a second response, to have less invasive and better understood treatments would not address all of the kinds of wrongdoing that ground a pessimistic induction. That pessimistic induction was grounded not only on accidental harm from treatments not turning out as expected, but also the fact that those treated were wronged. Suppose, for instance, that neurointerventions targeting criminals were as well understood as our best medical treatments. Still, a fundamental dissimilarity persists, over who benefits. Medicine generally aims to help the individual.[29] These neurointerventions used to treat criminals aim to benefit society. When we have deployed medical interventions for such purposes in the past, as in eugenics and psychiatry, we have gone wrong.

It might be objected that neurointerventions for those who engage in socially undesirable behaviour can be for the benefit of the person treated, not society, or not only for the benefit society.[30] On the face of it that would place such treatments on better footing with respect to avoiding the pessimistic induction. For some particular examples this line of response might work. Drug addiction treatment for who would like to cease to be addicts may be one such example. But it is largely implausible as a reconstruction of the reasons for such treatments for undesirable behaviour in general. So, to defend the reconstruction would be to claim that the individual treated always benefits from no longer being someone likely to act in socially undesirable ways and does so enough for this benefit to justify the risks of the treatment (assuming here, for the sake of argument only, that this is something the interventions could achieve). That claim, however, will be hard to defend. For a start, it takes an implausibly pater-

[29] There are some exceptions in public health, such as vaccines and quarantines, although the latter will often also benefit the individual and is very carefully restricted in many countries.

[30] With thanks to David Birks for pressing this objection.

nalistic stance in its claims to know about what is really good for an individual. Someone might prefer to remain as they were, or to be unaffected by the side effects of treatment. Further, to claim that the treatment is for the individual's own good since 'fixing' their 'criminality' would mean that they escaped further imprisonment is questionable. Given that the conviction rates for many crimes are very low, such a calculation may not work, especially for crimes where the individual gets some benefit from committing it. Finally, it may be objectionable, morally speaking, to define an individual's own good, and so justify treatment, in terms of what we as a society make true; for instance, that imprisonment is worse than the side effects of medical treatment owing to the nature of our jails.[31]

So, neither claiming that we now target the right people nor that we now have better science suffices to defeat the pessimistic induction. Yet what motivates these easy defences might be an underlying belief in progress. As we make the kinds of moral and scientific mistakes found in the history of neurointerventions for socially undesirable behaviour, one might think that we get more likely to be right the next time.[32] We would now be better at picking the right people to implement such interventions on, or at coming up with appropriate interventions proportionate to our ends. However, on what grounds could one defend such a Whiggish view of moral progress? There seems to be little evidence in our history that moral failings are seldom repeated. Genocides, to take one exemplar of a moral atrocity, are not one-off moral mistakes, but rather a repeating phenomenon.

Conclusion: The prospects for neurointerventions

So, proponents of neurointerventions to treat socially undesirable behaviours or tendencies that way should care about the history of such interventions, since that history grounds a presumption against taking this course of action. The history does so by motivating a pes-

---

[31] This may be related to Cohen's interpersonal test, e.g. see G. A. Cohen, *Rescuing Justice and Equality* (Harvard University Press, 2008).

[32] With thanks to Chris Nathan and Chris Mills for suggesting this objection.

simistic induction that requires answering. We need some account of why current neurointerventions as a response to socially undesirable behaviour are different to those of the past, in ethically salient ways. Furthermore, just performing these interventions on different people and with the use of modern science may not suffice to escape that induction.

There is, however, something further that we can draw from the pessimistic induction, which may provide a way for proponents of neurointerventions for undesirable behaviour to improve their chances, ethically speaking. That further something is the context in which neurointerventions are placed, both in comparison to medicine in general and as a practice amongst other ethically similar practices – i.e. other uses of medical interventions for the purposes of social control or the social good, such as eugenics and the worst elements of psychiatry. This context might provide us with some clues when making the case for neurointerventions against a terrible history. To conclude, then, I gesture towards the clues that might arise.

So, from the context, we might gain some sense of how to avoid ethical pitfalls. When do medical interventions, and especially psychiatric ones, avoid being unethical? One part of the answer is that they are sometimes ethical when they are genuinely for the benefit of the individual and we steer clear of assumptions about who is or is not 'fit' and about whose behaviour is socially undesirable. Further, a treatment should not be tested on a vulnerable population, its side effects should have been carefully tested, and its degree of invasiveness and risk of side effects should be proportional to its aims. Medics should treat only with consent and with regard to the patient's own interests.

There will be a set of other salient concerns to fill out here. But perhaps addressing concerns like the above would provide a route for proponents of neurointerventions in those accused or convicted of crimes to answer the pessimistic induction. They could draw on the ethical improvements of medicine and psychiatry, along with the reasons why eugenics and some part of psychiatry were abhorrent, to propose a limited but more ethically secure course of action involving neurointerventions. Taking this route would still make history relevant: it is the history that informs us about what to attend to ethically speaking.

However, the proponent of neurointerventions who takes such a route may end up with a nearly empty, or at least radically limited, set of proposals. Certainly, it is unclear that many (or any) of the suggested neurointerventions for criminals would exhibit all of these ethical improvements, with the exception of optional drug addiction treatment. Any neurointerventions that meet these ethical improvements will not be treating those convicted of crimes for the sake of altering undesirable behaviour, to benefit society, or to punish, nor would they be coercive or otherwise violate the requirements of informed consent. To propose neurointerventions that do reflect all the ethical improvements of modern medicine and psychiatry may be just to say that doctors should continue to treat those who suffer from physical or mental illnesses, who desire and consent to treatment, and only where the means are proportionate to the ends. In short, neurointerventions become permissible where they are part of normal medical practice, subject to the usual restrictions on that context. This continuation of the status quo is not, it seems, what the proponents of neurointerventions as a way to treat criminals had in mind.